

Reinforcement Learning Course: WiSe 2020/21

Marin Bukov

Faculty of Physics, Sofia University, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria

(Dated: December 16, 2020)

I. DEEP Q-NETWORK LEARNING

Many thanks to Jiahao Yao from UC Berkeley for helping me prepare this notebook!

A. Setup

You will need to make a copy of this notebook in your Google Drive before you can edit the homework files. You can do so with **File** → **Save a copy in Drive**.

```
In [ ]: #@title Jax with TPU

        #@markdown *(uncomment and run this block first)*

        #@markdown **CAVEAT:** this is currently slower than GPU, but this will
→supposedly
change very soon: see this [notebook](https://github.com/google/jax/blob/master/
→cloud_tp
u_colabs/JAX_NeurIPS_2020_demo.ipynb) and this [NeurIPS 2020
video](https://drive.google.com/file/d/1jKxefZT1xJDUxMman6qrQVed7vWIOMIn/view).

# # get the latest JAX and jaxlib
# !pip install --upgrade -q jax jaxlib

# # Colab runtime set to TPU accel
# import requests
# import os
# if 'TPU_DRIVER_MODE' not in globals():
#     url = 'http://' + os.environ['COLAB_TPU_ADDR'].split(':')[0] +
':8475/requestversion/tpu_driver_nightly'
#     resp = requests.post(url)
#     TPU_DRIVER_MODE = 1

# # TPU driver as backend for JAX
# from jax.config import config
# config.FLAGS.jax_xla_backend = "tpu_driver"
# config.FLAGS.jax_backend_target = "grpc://" + os.environ['COLAB_TPU_ADDR']
# print(config.FLAGS.jax_backend_target)

In [ ]: #@title mount your Google Drive
        #@markdown Your work will be stored in a folder called `DQN_atari` by default to
→prevent
Colab instance timeouts from deleting your edits.

import os
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
In [ ]: #@title Atari Environments
        #@markdown We will use the Gym Pacman environment.
```

```
env_name = 'pacman'
gym_name_map = {
    'pacman': 'MsPacman-v0',
}
gym_name = gym_name_map[env_name]
print('Gym Env:', gym_name)
```

Gym Env: MsPacman-v0

```
In [ ]: #@title set up mount symlink
        %cd /content
        DRIVE_PATH = '/content/gdrive/My\ Drive/DQN_atari2'
        DRIVE_PYTHON_PATH = DRIVE_PATH.replace('\\', '')
        if not os.path.exists(DRIVE_PYTHON_PATH):
            %mkdir $DRIVE_PATH
```

```
## the space in `My Drive` causes some issues,
## make a symlink to avoid this
SYM_PATH = '/content/DQN_atari'
if not os.path.exists(SYM_PATH):
    !ln -s $DRIVE_PATH $SYM_PATH
```

```
%cd $SYM_PATH
```

```
/content
/content/gdrive/My Drive/DQN_atari2
```

```
In [ ]: #@title apt install requirements
```

```
##@markdown Run each section with Shift+Enter
```

```
##@markdown Double-click on section headers to show code.
```

```
##@markdown If you see some ERRORS, they are caused by dependencies which are
preinstalled in Google Colab; you may ignore them.
```

```
!apt update
!apt install -y --no-install-recommends 
    build-essential \
    curl \
    git \
    gnupg2 \
    make \
    cmake \
    ffmpeg \
    swig \
    libz-dev \
    unzip \
    zlib1g-dev \
    libglfw3 \
    libglfw3-dev \
```

```

libxrandr2 \
libxinerama-dev \
libxi6 \
libxcursor-dev \
libgl1-mesa-dev \
libgl1-mesa-glx \
libglew-dev \
libosmesa6-dev \
lsb-release \
ack-grep \
patchelf \
wget \
xpra \
xserver-xorg-dev \
xvfb \
python-opengl \
ffmpeg > /dev/null 2>&1

```

```
!pip install opencv-python==3.4.0.12
```

```

Ign:1 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
InRelease
Ign:2 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64 InRelease
Hit:3 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Release
Hit:4 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64 Release
Get:5 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Hit:6 http://archive.ubuntu.com/ubuntu bionic InRelease
Get:7 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ InRelease [3,626 B]
Get:8 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic InRelease [15.9 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Get:12 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease [21.3 kB]
Get:13 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ Packages [40.7 kB]
Get:14 http://archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Get:15 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main Sources [1,697
kB]
Get:16 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages
[1,368 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages [2,136
kB]
Get:18 http://security.ubuntu.com/ubuntu bionic-security/multiverse amd64 Packages
[15.8 kB]
Get:19 http://security.ubuntu.com/ubuntu bionic-security/restricted amd64 Packages
[223 kB]
Get:20 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages [1,784
kB]
Get:21 http://archive.ubuntu.com/ubuntu bionic-updates/multiverse amd64 Packages [53.8
kB]
Get:22 http://archive.ubuntu.com/ubuntu bionic-updates/restricted amd64 Packages [266
kB]
Get:23 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages [2,241 kB]
Get:24 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main amd64 Packages
[869 kB]
Get:25 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic/main amd64 Packages
[46.5 kB]

```



```
(from -r requirements_colab.txt (line 3)) (2.3.8)
Requirement already satisfied: scipy in /usr/local/lib/python3.6/dist-packages (from
gym[atari]==0.17.2->-r requirements_colab.txt (line 1)) (1.4.1)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.6/dist-packages
(from gym[atari]==0.17.2->-r requirements_colab.txt (line 1)) (1.18.5)
Requirement already satisfied: pygame<=1.5.0,>=1.4.0 in /usr/local/lib/python3.6/dist-
packages (from gym[atari]==0.17.2->-r requirements_colab.txt (line 1)) (1.5.0)
Requirement already satisfied: cloudpickle<1.4.0,>=1.2.0 in
/usr/local/lib/python3.6/dist-packages (from gym[atari]==0.17.2->-r
requirements_colab.txt (line 1)) (1.3.0)
Requirement already satisfied: atari-py~=0.2.0; extra == "atari" in
/usr/local/lib/python3.6/dist-packages (from gym[atari]==0.17.2->-r
requirements_colab.txt (line 1)) (0.2.6)
Requirement already satisfied: Pillow; extra == "atari" in
/usr/local/lib/python3.6/dist-packages (from gym[atari]==0.17.2->-r
requirements_colab.txt (line 1)) (7.0.0)
Requirement already satisfied: opencv-python; extra == "atari" in
/usr/local/lib/python3.6/dist-packages (from gym[atari]==0.17.2->-r
requirements_colab.txt (line 1)) (3.4.0.12)
Requirement already satisfied: EasyProcess in /usr/local/lib/python3.6/dist-packages
(from pyvirtualdisplay==1.3.2->-r requirements_colab.txt (line 2)) (0.3)
Requirement already satisfied: future in /usr/local/lib/python3.6/dist-packages (from
pygame<=1.5.0,>=1.4.0->gym[atari]==0.17.2->-r requirements_colab.txt (line 1))
(0.16.0)
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (from
atari-py~=0.2.0; extra == "atari"->gym[atari]==0.17.2->-r requirements_colab.txt (line
1)) (1.15.0)
/content/gdrive/My Drive/DQN_atari2
```

```
In [ ]: ##title set up virtual display
```

```
from pyvirtualdisplay import Display

display = Display(visible=0, size=(1400, 900))
display.start()

# For later
from colab_utils import (
    wrap_env_demo,
    show_video_demo,
    show_video
)
)
```

```
In [ ]: ##title test virtual display
```

```
##markdown If you see a video, setup is complete!

import gym
import matplotlib

env = wrap_env_demo(gym.make(gym_name))

observation = env.reset()
for i in range(10):
    env.render(mode='rgb_array')
    obs, rew, term, _ = env.step(env.action_space.sample() )
```

```

        if term:
            break;

env.close()
print('Loading video...')
show_video_demo()

```

Loading video...

<IPython.core.display.HTML object>

In []: *#@title imports*

```

import os.path as osp
import sys, time
from functools import partial

import gym
from gym import wrappers

import numpy as np
import random

from atari_wrappers import *
from buffer import ReplayBuffer

from NN import Neural_Net
from env_utils import episode_step
from video_utils import learning_logger

%load_ext autoreload
%autoreload 2

```

In []: *#@title hyperparameters*

```

#@markdown Set the seed and define the hyperparameters for DQN.

# seed
seed = 0
np.random.seed(seed)
random.seed(seed)
np.random.RandomState(seed)

# Q-learning & network
N_iterations = 3000000 # 200 #

# discount factor
gamma = 0.99

# Q network update frequency
update_frequency = 4

# frame history length
agent_history_length = 4

```

```

use_target = True
# target network update frequency
target_update = 10000 # 100 #
minibatch_size = 32

# replay buffer parameters
replay_memory_size = 1000000 # 10000 #

# buffer prefilling steps
replay_start_size = 50000 # 500 #

# adam parameters
step_size = 1e-4
adam_beta1 = 0.9
adam_beta2 = 0.999
adam_eps = 1e-4
adam_params=dict(N_iterations=N_iterations,
                 step_size=step_size,
                 b1=adam_beta1,
                 b2=adam_beta2,
                 eps=adam_eps,
                 )

# exploration (epsilon-greedy) schedule
eps_schedule_step = [0, 1e6, 2.5e6]
#eps_schedule_val = [1.0, 0.1, 0.01]
eps_schedule_val = [0.2, 0.1, 0.01]
eps_schedule_args = dict(
    eps_schedule_step=eps_schedule_step, eps_schedule_val=eps_schedule_val
)

# video logging: default is to not log video so that logs are small enough
# in units of episodes
video_log_freq = 1000 #-1

In [ ]: #@title load the pacman gym environment

def get_env(seed):
    env = gym.make("MsPacman-v0")
    env.seed(seed)
    env.action_space.np_random.seed(seed)
    expt_dir = "./"

    # the video recorder only captures a sampling of episodes
    # (those with episode numbers which are perfect cubes: 1, 8, 27, 64, ...)
    →and then
    every [video_log_freq]-th).
    def capped_cubic_video_schedule(episode_id):
        if episode_id < video_log_freq:
            return int(round(episode_id ** (1.0 / 3))) ** 3 == episode_id
        else:
            return episode_id % video_log_freq == 0

    env = wrappers.Monitor(
        env,
        osp.join(expt_dir, "video"),

```

```

        force=True,
        video_callable=(capped_cubic_video_schedule if video_log_freq > 0 else_
→False),
    )

    # configure environment for DeepMind-style Atari
    env = wrap_deepmind(env)
    return env

```

```

##### Create a breakout environment
# fix env seeds
env = get_env(seed)
# reset environment to initial state
frame = env.reset()
# get the size of the action space
n_actions = env.action_space.n

# define logger
rl_logger = learning_logger(env, eps_schedule_args)

```

In []: *#@title create the data buffer*

```

frame_shape = (env.observation_space.shape[0], env.observation_space.shape[1])
replay_buffer = ReplayBuffer(replay_memory_size, agent_history_length,
→lander=False)
# channel last format of the input
input_shape = (1,) + frame_shape + (agent_history_length,)

```

In []: *#@build deep Q-network*

```

print("build the Q learning network.\n")
##### Create deep neural net
model = Neural_Net(
    n_actions,
    input_shape,
    adam_params,
    use_target=use_target,
    seed=seed
)

```

build the Q learning network.

DQN input shape: (1, 84, 84, 4).

In []: *#@title Pre-fill buffer*

```

print("Start prefilling the buffer.\n")

tot_time = time.time()

##### prefill buffer using the random policy
pre_iteration = 0
while pre_iteration < replay_start_size:
    # reset environment

```



```

state = env.reset()
is_terminal = False

while not is_terminal:

    # store state in buffer
    buffer_index = replay_buffer.store_frame(state)
    last_obs_encode = replay_buffer.encode_recent_observation()
    state_enc = np.expand_dims(last_obs_encode, 0)

    # take environment step and overwrite state; reward is not used to
→prefill
    buffer
    state, reward, is_terminal = episode_step(
        pre_iteration,
        env,
        model,
        replay_buffer,
        buffer_index,
        state_enc,
        prefill_buffer=True,
    )
    pre_iteration += 1

    print("\nFinished prefilling the buffer.\n")

```

Start prefilling the buffer.

Finished prefilling the buffer.

In []: *#@title train DQN*

```

# reset environment
state = env.reset()

#####
print("Start learning.\n")
##### run DQN
for iteration in range(N_iterations):

    # store state in buffer and compute its encoding
    buffer_index = replay_buffer.store_frame(state)
    last_obs_encode = replay_buffer.encode_recent_observation()
    state_enc = np.expand_dims(last_obs_encode, 0)

    # take one episode step
    state, reward, is_terminal = episode_step(
        iteration,
        env,
        model,
        replay_buffer,
        buffer_index,
        state_enc,
        eps_schedule_args=eps_schedule_args,

```

```

    )

    # update deep Q-net
    if iteration % update_frequency == 0:
        model.update_Qnet(replay_buffer, minibatch_size, gamma)

    # update target Q-net
    if iteration % target_update == 0:
        model.update_Qnet_target()

    if is_terminal:
        # print stats
        rl_logger.stats(iteration)

        # reset environment
        state = env.reset()

    print("\n\ntotal time: {}".format(time.time() - tot_time))

```

Streaming output truncated to the last 5000 lines.

Timestep 2951737

mean reward (100 episodes) 1900.500000

best mean reward 2123.500000

episodes 11261

exploration 0.010000

running time 0.005929

 [1590.0, 2170.0, 1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0]

 Timestep 2951869

mean reward (100 episodes) 1900.500000

best mean reward 2123.500000

episodes 11261

exploration 0.010000

running time 0.012838

 [1590.0, 2170.0, 1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0]

 Timestep 2951967

mean reward (100 episodes) 1900.500000

best mean reward 2123.500000

episodes 11261

exploration 0.010000

running time 0.009840

 [2170.0, 1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0]

 Timestep 2952017

mean reward (100 episodes) 1900.100000

best mean reward 2123.500000

episodes 11262

exploration 0.010000

running time 0.005109

[2170.0, 1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0]

Timestep 2952220
mean reward (100 episodes) 1900.100000
best mean reward 2123.500000
episodes 11262
exploration 0.010000
running time 0.019820

[2170.0, 1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0]

Timestep 2952294
mean reward (100 episodes) 1900.100000
best mean reward 2123.500000
episodes 11262
exploration 0.010000
running time 0.007293

[1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0]

Timestep 2952339
mean reward (100 episodes) 1885.400000
best mean reward 2123.500000
episodes 11263
exploration 0.010000
running time 0.004777

[1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0]

Timestep 2952492
mean reward (100 episodes) 1885.400000
best mean reward 2123.500000
episodes 11263
exploration 0.010000
running time 0.015232

[1710.0, 1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0]

Timestep 2952574
mean reward (100 episodes) 1885.400000
best mean reward 2123.500000
episodes 11263
exploration 0.010000
running time 0.008123

[1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0]

Timestep 2952618
mean reward (100 episodes) 1891.100000
best mean reward 2123.500000

episodes 11264
exploration 0.010000
running time 0.004279

[1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0]

Timestep 2952787
mean reward (100 episodes) 1891.100000
best mean reward 2123.500000
episodes 11264
exploration 0.010000
running time 0.016945

[1250.0, 1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0]

Timestep 2952840
mean reward (100 episodes) 1891.100000
best mean reward 2123.500000
episodes 11264
exploration 0.010000
running time 0.005668

[1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0]

Timestep 2952892
mean reward (100 episodes) 1879.800000
best mean reward 2123.500000
episodes 11265
exploration 0.010000
running time 0.005072

[1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0]

Timestep 2953077
mean reward (100 episodes) 1879.800000
best mean reward 2123.500000
episodes 11265
exploration 0.010000
running time 0.018298

[1420.0, 2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0]

Timestep 2953222
mean reward (100 episodes) 1879.800000
best mean reward 2123.500000
episodes 11265
exploration 0.010000
running time 0.013981

[2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0]

```
-----  
Timestep 2953270  
mean reward (100 episodes) 1888.100000  
best mean reward 2123.500000  
episodes 11266  
exploration 0.010000  
running time 0.004703  
-----  
[2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0]  
-----  
Timestep 2953498  
mean reward (100 episodes) 1888.100000  
best mean reward 2123.500000  
episodes 11266  
exploration 0.010000  
running time 0.022178  
-----  
[2480.0, 1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0]  
-----  
Timestep 2953593  
mean reward (100 episodes) 1888.100000  
best mean reward 2123.500000  
episodes 11266  
exploration 0.010000  
running time 0.009356  
-----  
[1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0]  
-----  
Timestep 2953621  
mean reward (100 episodes) 1886.500000  
best mean reward 2123.500000  
episodes 11267  
exploration 0.010000  
running time 0.003000  
-----  
[1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0]  
-----  
Timestep 2953715  
mean reward (100 episodes) 1886.500000  
best mean reward 2123.500000  
episodes 11267  
exploration 0.010000  
running time 0.009389  
-----  
[1290.0, 3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0]  
-----  
Timestep 2953818  
mean reward (100 episodes) 1886.500000  
best mean reward 2123.500000  
episodes 11267  
exploration 0.010000
```

running time 0.010282

[3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0]

Timestep 2953870

mean reward (100 episodes) 1884.300000

best mean reward 2123.500000

episodes 11268

exploration 0.010000

running time 0.005301

[3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0]

Timestep 2954052

mean reward (100 episodes) 1884.300000

best mean reward 2123.500000

episodes 11268

exploration 0.010000

running time 0.017801

[3000.0, 1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0]

Timestep 2954119

mean reward (100 episodes) 1884.300000

best mean reward 2123.500000

episodes 11268

exploration 0.010000

running time 0.006628

[1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0]

Timestep 2954158

mean reward (100 episodes) 1877.900000

best mean reward 2123.500000

episodes 11269

exploration 0.010000

running time 0.003958

[1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0]

Timestep 2954272

mean reward (100 episodes) 1877.900000

best mean reward 2123.500000

episodes 11269

exploration 0.010000

running time 0.011659

[1630.0, 1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0]

Timestep 2954315

mean reward (100 episodes) 1877.900000
best mean reward 2123.500000
episodes 11269
exploration 0.010000
running time 0.004559

[1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0]

Timestep 2954348
mean reward (100 episodes) 1868.000000
best mean reward 2123.500000
episodes 11270
exploration 0.010000
running time 0.003519

[1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0]

Timestep 2954445
mean reward (100 episodes) 1868.000000
best mean reward 2123.500000
episodes 11270
exploration 0.010000
running time 0.009599

[1230.0, 2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0]

Timestep 2954566
mean reward (100 episodes) 1868.000000
best mean reward 2123.500000
episodes 11270
exploration 0.010000
running time 0.012021

[2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0]

Timestep 2954660
mean reward (100 episodes) 1862.100000
best mean reward 2123.500000
episodes 11271
exploration 0.010000
running time 0.009335

[2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0]

Timestep 2954843
mean reward (100 episodes) 1862.100000
best mean reward 2123.500000
episodes 11271
exploration 0.010000
running time 0.017797

[2120.0, 1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0]

Timestep 2954909
mean reward (100 episodes) 1862.100000
best mean reward 2123.500000
episodes 11271
exploration 0.010000
running time 0.006830

[1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0]

Timestep 2954977
mean reward (100 episodes) 1866.700000
best mean reward 2123.500000
episodes 11272
exploration 0.010000
running time 0.006726

[1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0]

Timestep 2955111
mean reward (100 episodes) 1866.700000
best mean reward 2123.500000
episodes 11272
exploration 0.010000
running time 0.013344

[1760.0, 1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0]

Timestep 2955248
mean reward (100 episodes) 1866.700000
best mean reward 2123.500000
episodes 11272
exploration 0.010000
running time 0.013933

[1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0]

Timestep 2955311
mean reward (100 episodes) 1874.200000
best mean reward 2123.500000
episodes 11273
exploration 0.010000
running time 0.006076

[1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0]

Timestep 2955448
mean reward (100 episodes) 1874.200000
best mean reward 2123.500000

episodes 11273
exploration 0.010000
running time 0.013769

[1540.0, 1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0]

Timestep 2955536
mean reward (100 episodes) 1874.200000
best mean reward 2123.500000
episodes 11273
exploration 0.010000
running time 0.008750

[1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0]

Timestep 2955569
mean reward (100 episodes) 1874.300000
best mean reward 2123.500000
episodes 11274
exploration 0.010000
running time 0.003397

[1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0]

Timestep 2955713
mean reward (100 episodes) 1874.300000
best mean reward 2123.500000
episodes 11274
exploration 0.010000
running time 0.014110

[1140.0, 2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0]

Timestep 2955806
mean reward (100 episodes) 1874.300000
best mean reward 2123.500000
episodes 11274
exploration 0.010000
running time 0.008966

[2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0]

Timestep 2955854
mean reward (100 episodes) 1864.600000
best mean reward 2123.500000
episodes 11275
exploration 0.010000
running time 0.005028

[2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0]

```
-----  
Timestep 2956047  
mean reward (100 episodes) 1864.600000  
best mean reward 2123.500000  
episodes 11275  
exploration 0.010000  
running time 0.019119  
-----
```

```
[2040.0, 2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0]
```

```
-----  
Timestep 2956079  
mean reward (100 episodes) 1864.600000  
best mean reward 2123.500000  
episodes 11275  
exploration 0.010000  
running time 0.003347  
-----
```

```
[2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0]
```

```
-----  
Timestep 2956104  
mean reward (100 episodes) 1859.000000  
best mean reward 2123.500000  
episodes 11276  
exploration 0.010000  
running time 0.002737  
-----
```

```
[2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0]
```

```
-----  
Timestep 2956235  
mean reward (100 episodes) 1859.000000  
best mean reward 2123.500000  
episodes 11276  
exploration 0.010000  
running time 0.012819  
-----
```

```
[2450.0, 1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0]
```

```
-----  
Timestep 2956366  
mean reward (100 episodes) 1859.000000  
best mean reward 2123.500000  
episodes 11276  
exploration 0.010000  
running time 0.013174  
-----
```

```
[1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0]
```

```
-----  
Timestep 2956436  
mean reward (100 episodes) 1860.600000  
best mean reward 2123.500000  
episodes 11277  
exploration 0.010000  
-----
```

running time 0.007016

[1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0]

Timestep 2956568

mean reward (100 episodes) 1860.600000

best mean reward 2123.500000

episodes 11277

exploration 0.010000

running time 0.013185

[1300.0, 1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0]

Timestep 2956621

mean reward (100 episodes) 1860.600000

best mean reward 2123.500000

episodes 11277

exploration 0.010000

running time 0.005292

[1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0]

Timestep 2956669

mean reward (100 episodes) 1860.900000

best mean reward 2123.500000

episodes 11278

exploration 0.010000

running time 0.004822

[1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0]

Timestep 2956846

mean reward (100 episodes) 1860.900000

best mean reward 2123.500000

episodes 11278

exploration 0.010000

running time 0.017279

[1920.0, 1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0]

Timestep 2956881

mean reward (100 episodes) 1860.900000

best mean reward 2123.500000

episodes 11278

exploration 0.010000

running time 0.003411

[1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0]

Timestep 2956923

mean reward (100 episodes) 1845.200000
best mean reward 2123.500000
episodes 11279
exploration 0.010000
running time 0.004312

[1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0]

Timestep 2957071
mean reward (100 episodes) 1845.200000
best mean reward 2123.500000
episodes 11279
exploration 0.010000
running time 0.014570

[1080.0, 1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0]

Timestep 2957161
mean reward (100 episodes) 1845.200000
best mean reward 2123.500000
episodes 11279
exploration 0.010000
running time 0.008702

[1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0]

Timestep 2957198
mean reward (100 episodes) 1846.600000
best mean reward 2123.500000
episodes 11280
exploration 0.010000
running time 0.003777

[1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0]

Timestep 2957291
mean reward (100 episodes) 1846.600000
best mean reward 2123.500000
episodes 11280
exploration 0.010000
running time 0.009368

[1580.0, 2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0]

Timestep 2957369
mean reward (100 episodes) 1846.600000
best mean reward 2123.500000
episodes 11280
exploration 0.010000
running time 0.007875

[2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0]

Timestep 2957422
mean reward (100 episodes) 1841.200000
best mean reward 2123.500000
episodes 11281
exploration 0.010000
running time 0.005472

[2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0]

Timestep 2957529
mean reward (100 episodes) 1841.200000
best mean reward 2123.500000
episodes 11281
exploration 0.010000
running time 0.010465

[2260.0, 2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0]

Timestep 2957598
mean reward (100 episodes) 1841.200000
best mean reward 2123.500000
episodes 11281
exploration 0.010000
running time 0.006816

[2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0]

Timestep 2957706
mean reward (100 episodes) 1834.700000
best mean reward 2123.500000
episodes 11282
exploration 0.010000
running time 0.010542

[2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0]

Timestep 2957838
mean reward (100 episodes) 1834.700000
best mean reward 2123.500000
episodes 11282
exploration 0.010000
running time 0.013306

[2400.0, 1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0]

Timestep 2957988
mean reward (100 episodes) 1834.700000
best mean reward 2123.500000

episodes 11282
exploration 0.010000
running time 0.015055

[1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0]

Timestep 2958030
mean reward (100 episodes) 1848.600000
best mean reward 2123.500000
episodes 11283
exploration 0.010000
running time 0.004173

[1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0]

Timestep 2958129
mean reward (100 episodes) 1848.600000
best mean reward 2123.500000
episodes 11283
exploration 0.010000
running time 0.010281

[1450.0, 1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0]

Timestep 2958242
mean reward (100 episodes) 1848.600000
best mean reward 2123.500000
episodes 11283
exploration 0.010000
running time 0.011086

[1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0]

Timestep 2958294
mean reward (100 episodes) 1856.100000
best mean reward 2123.500000
episodes 11284
exploration 0.010000
running time 0.005193

[1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0]

Timestep 2958421
mean reward (100 episodes) 1856.100000
best mean reward 2123.500000
episodes 11284
exploration 0.010000
running time 0.012927

[1510.0, 1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0]

```
-----  
Timestep 2958463  
mean reward (100 episodes) 1856.100000  
best mean reward 2123.500000  
episodes 11284  
exploration 0.010000  
running time 0.004028  
-----
```

```
[1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0]
```

```
-----  
Timestep 2958491  
mean reward (100 episodes) 1847.200000  
best mean reward 2123.500000  
episodes 11285  
exploration 0.010000  
running time 0.002846  
-----
```

```
[1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0]
```

```
-----  
Timestep 2958596  
mean reward (100 episodes) 1847.200000  
best mean reward 2123.500000  
episodes 11285  
exploration 0.010000  
running time 0.010952  
-----
```

```
[1630.0, 1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0]
```

```
-----  
Timestep 2958692  
mean reward (100 episodes) 1847.200000  
best mean reward 2123.500000  
episodes 11285  
exploration 0.010000  
running time 0.009313  
-----
```

```
[1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0]
```

```
-----  
Timestep 2958784  
mean reward (100 episodes) 1850.600000  
best mean reward 2123.500000  
episodes 11286  
exploration 0.010000  
running time 0.009153  
-----
```

```
[1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0]
```

```
-----  
Timestep 2958926  
mean reward (100 episodes) 1850.600000  
best mean reward 2123.500000  
episodes 11286  
exploration 0.010000  
-----
```

running time 0.014253

[1490.0, 1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0]

Timestep 2958953

mean reward (100 episodes) 1850.600000

best mean reward 2123.500000

episodes 11286

exploration 0.010000

running time 0.002855

[1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0]

Timestep 2959003

mean reward (100 episodes) 1843.700000

best mean reward 2123.500000

episodes 11287

exploration 0.010000

running time 0.005021

[1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0]

Timestep 2959121

mean reward (100 episodes) 1843.700000

best mean reward 2123.500000

episodes 11287

exploration 0.010000

running time 0.011799

[1520.0, 1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0]

Timestep 2959164

mean reward (100 episodes) 1843.700000

best mean reward 2123.500000

episodes 11287

exploration 0.010000

running time 0.004247

[1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0]

Timestep 2959241

mean reward (100 episodes) 1842.500000

best mean reward 2123.500000

episodes 11288

exploration 0.010000

running time 0.007515

[1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0]

Timestep 2959354

mean reward (100 episodes) 1842.500000
best mean reward 2123.500000
episodes 11288
exploration 0.010000
running time 0.011177

[1430.0, 1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0]

Timestep 2959442
mean reward (100 episodes) 1842.500000
best mean reward 2123.500000
episodes 11288
exploration 0.010000
running time 0.008747

[1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0]

Timestep 2959486
mean reward (100 episodes) 1836.300000
best mean reward 2123.500000
episodes 11289
exploration 0.010000
running time 0.004545

[1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0]

Timestep 2959597
mean reward (100 episodes) 1836.300000
best mean reward 2123.500000
episodes 11289
exploration 0.010000
running time 0.011019

[1940.0, 1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0]

Timestep 2959705
mean reward (100 episodes) 1836.300000
best mean reward 2123.500000
episodes 11289
exploration 0.010000
running time 0.010512

[1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0]

Timestep 2959749
mean reward (100 episodes) 1840.400000
best mean reward 2123.500000
episodes 11290
exploration 0.010000
running time 0.004323

[1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0]

Timestep 2959965
mean reward (100 episodes) 1840.400000
best mean reward 2123.500000
episodes 11290
exploration 0.010000
running time 0.021164

[1380.0, 1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0]

Timestep 2960030
mean reward (100 episodes) 1840.400000
best mean reward 2123.500000
episodes 11290
exploration 0.010000
running time 0.006518

[1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0]

Timestep 2960070
mean reward (100 episodes) 1840.300000
best mean reward 2123.500000
episodes 11291
exploration 0.010000
running time 0.003996

[1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0]

Timestep 2960166
mean reward (100 episodes) 1840.300000
best mean reward 2123.500000
episodes 11291
exploration 0.010000
running time 0.009537

[1460.0, 2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0]

Timestep 2960289
mean reward (100 episodes) 1840.300000
best mean reward 2123.500000
episodes 11291
exploration 0.010000
running time 0.012238

[2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0]

Timestep 2960342
mean reward (100 episodes) 1840.800000
best mean reward 2123.500000

episodes 11292
exploration 0.010000
running time 0.005388

[2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0]

Timestep 2960450
mean reward (100 episodes) 1840.800000
best mean reward 2123.500000
episodes 11292
exploration 0.010000
running time 0.010783

[2730.0, 2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0]

Timestep 2960505
mean reward (100 episodes) 1840.800000
best mean reward 2123.500000
episodes 11292
exploration 0.010000
running time 0.005346

[2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0]

Timestep 2960567
mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11293
exploration 0.010000
running time 0.006058

[2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0]

Timestep 2960701
mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11293
exploration 0.010000
running time 0.013341

[2090.0, 1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0]

Timestep 2960743
mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11293
exploration 0.010000
running time 0.004144

[1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0]

```
-----  
Timestep 2960820  
mean reward (100 episodes) 1843.400000  
best mean reward 2123.500000  
episodes 11294  
exploration 0.010000  
running time 0.007938  
-----  
[1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0]  
-----  
Timestep 2960978  
mean reward (100 episodes) 1843.400000  
best mean reward 2123.500000  
episodes 11294  
exploration 0.010000  
running time 0.015893  
-----  
[1080.0, 1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0]  
-----  
Timestep 2961093  
mean reward (100 episodes) 1843.400000  
best mean reward 2123.500000  
episodes 11294  
exploration 0.010000  
running time 0.011767  
-----  
[1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0]  
-----  
Timestep 2961203  
mean reward (100 episodes) 1847.300000  
best mean reward 2123.500000  
episodes 11295  
exploration 0.010000  
running time 0.010838  
-----  
[1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0]  
-----  
Timestep 2961306  
mean reward (100 episodes) 1847.300000  
best mean reward 2123.500000  
episodes 11295  
exploration 0.010000  
running time 0.010412  
-----  
[1870.0, 1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0]  
-----  
Timestep 2961426  
mean reward (100 episodes) 1847.300000  
best mean reward 2123.500000  
episodes 11295  
exploration 0.010000
```

running time 0.011986

[1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0]

Timestep 2961466

mean reward (100 episodes) 1849.100000

best mean reward 2123.500000

episodes 11296

exploration 0.010000

running time 0.004223

[1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0]

Timestep 2961622

mean reward (100 episodes) 1849.100000

best mean reward 2123.500000

episodes 11296

exploration 0.010000

running time 0.015544

[1520.0, 1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0]

Timestep 2961680

mean reward (100 episodes) 1849.100000

best mean reward 2123.500000

episodes 11296

exploration 0.010000

running time 0.005958

[1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0]

Timestep 2961706

mean reward (100 episodes) 1849.800000

best mean reward 2123.500000

episodes 11297

exploration 0.010000

running time 0.002762

[1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0]

Timestep 2961959

mean reward (100 episodes) 1849.800000

best mean reward 2123.500000

episodes 11297

exploration 0.010000

running time 0.024555

[1840.0, 1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0]

Timestep 2962011

mean reward (100 episodes) 1849.800000
best mean reward 2123.500000
episodes 11297
exploration 0.010000
running time 0.005374

[1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0]

Timestep 2962043
mean reward (100 episodes) 1853.600000
best mean reward 2123.500000
episodes 11298
exploration 0.010000
running time 0.003495

[1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0]

Timestep 2962183
mean reward (100 episodes) 1853.600000
best mean reward 2123.500000
episodes 11298
exploration 0.010000
running time 0.013889

[1400.0, 1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0]

Timestep 2962342
mean reward (100 episodes) 1853.600000
best mean reward 2123.500000
episodes 11298
exploration 0.010000
running time 0.015456

[1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0]

Timestep 2962386
mean reward (100 episodes) 1857.400000
best mean reward 2123.500000
episodes 11299
exploration 0.010000
running time 0.004314

[1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0]

Timestep 2962486
mean reward (100 episodes) 1857.400000
best mean reward 2123.500000
episodes 11299
exploration 0.010000
running time 0.009874

[1680.0, 1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0]

Timestep 2962523
mean reward (100 episodes) 1857.400000
best mean reward 2123.500000
episodes 11299
exploration 0.010000
running time 0.003702

[1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0]

Timestep 2962567
mean reward (100 episodes) 1847.000000
best mean reward 2123.500000
episodes 11300
exploration 0.010000
running time 0.004243

[1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0]

Timestep 2962634
mean reward (100 episodes) 1847.000000
best mean reward 2123.500000
episodes 11300
exploration 0.010000
running time 0.006646

[1870.0, 1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0]

Timestep 2962851
mean reward (100 episodes) 1847.000000
best mean reward 2123.500000
episodes 11300
exploration 0.010000
running time 0.021458

[1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0]

Timestep 2962938
mean reward (100 episodes) 1856.500000
best mean reward 2123.500000
episodes 11301
exploration 0.010000
running time 0.008500

[1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0]

Timestep 2963065
mean reward (100 episodes) 1856.500000
best mean reward 2123.500000

episodes 11301
exploration 0.010000
running time 0.012936

[1970.0, 1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0]

Timestep 2963116
mean reward (100 episodes) 1856.500000
best mean reward 2123.500000
episodes 11301
exploration 0.010000
running time 0.005173

[1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0]

Timestep 2963161
mean reward (100 episodes) 1869.000000
best mean reward 2123.500000
episodes 11302
exploration 0.010000
running time 0.004630

[1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0]

Timestep 2963261
mean reward (100 episodes) 1869.000000
best mean reward 2123.500000
episodes 11302
exploration 0.010000
running time 0.010167

[1230.0, 1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0]

Timestep 2963303
mean reward (100 episodes) 1869.000000
best mean reward 2123.500000
episodes 11302
exploration 0.010000
running time 0.004054

[1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0]

Timestep 2963347
mean reward (100 episodes) 1857.300000
best mean reward 2123.500000
episodes 11303
exploration 0.010000
running time 0.004751

[1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0]


```
-----  
Timestep 2963480  
mean reward (100 episodes) 1857.300000  
best mean reward 2123.500000  
episodes 11303  
exploration 0.010000  
running time 0.013110  
-----
```

```
[1410.0, 1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0]
```

```
-----  
Timestep 2963538  
mean reward (100 episodes) 1857.300000  
best mean reward 2123.500000  
episodes 11303  
exploration 0.010000  
running time 0.006057  
-----
```

```
[1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0]
```

```
-----  
Timestep 2963600  
mean reward (100 episodes) 1860.400000  
best mean reward 2123.500000  
episodes 11304  
exploration 0.010000  
running time 0.006017  
-----
```

```
[1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0]
```

```
-----  
Timestep 2963768  
mean reward (100 episodes) 1860.400000  
best mean reward 2123.500000  
episodes 11304  
exploration 0.010000  
running time 0.016364  
-----
```

```
[1770.0, 1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0]
```

```
-----  
Timestep 2963822  
mean reward (100 episodes) 1860.400000  
best mean reward 2123.500000  
episodes 11304  
exploration 0.010000  
running time 0.005447  
-----
```

```
[1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0]
```

```
-----  
Timestep 2963913  
mean reward (100 episodes) 1858.800000  
best mean reward 2123.500000  
episodes 11305  
exploration 0.010000  
-----
```

running time 0.009253

[1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0]

Timestep 2964020

mean reward (100 episodes) 1858.800000

best mean reward 2123.500000

episodes 11305

exploration 0.010000

running time 0.010812

[1910.0, 1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0]

Timestep 2964108

mean reward (100 episodes) 1858.800000

best mean reward 2123.500000

episodes 11305

exploration 0.010000

running time 0.009038

[1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0]

Timestep 2964199

mean reward (100 episodes) 1850.100000

best mean reward 2123.500000

episodes 11306

exploration 0.010000

running time 0.008948

[1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0]

Timestep 2964309

mean reward (100 episodes) 1850.100000

best mean reward 2123.500000

episodes 11306

exploration 0.010000

running time 0.010873

[1540.0, 1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0]

Timestep 2964345

mean reward (100 episodes) 1850.100000

best mean reward 2123.500000

episodes 11306

exploration 0.010000

running time 0.003711

[1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0]

Timestep 2964626

mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11307
exploration 0.010000
running time 0.027569

[1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0]

Timestep 2964993
mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11307
exploration 0.010000
running time 0.037939

[1660.0, 1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0]

Timestep 2965032
mean reward (100 episodes) 1839.500000
best mean reward 2123.500000
episodes 11307
exploration 0.010000
running time 0.004287

[1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0]

Timestep 2965074
mean reward (100 episodes) 1852.400000
best mean reward 2123.500000
episodes 11308
exploration 0.010000
running time 0.004354

[1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0]

Timestep 2965255
mean reward (100 episodes) 1852.400000
best mean reward 2123.500000
episodes 11308
exploration 0.010000
running time 0.018571

[1870.0, 1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0]

Timestep 2965313
mean reward (100 episodes) 1852.400000
best mean reward 2123.500000
episodes 11308
exploration 0.010000
running time 0.005869

[1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0]

Timestep 2965355
mean reward (100 episodes) 1863.700000
best mean reward 2123.500000
episodes 11309
exploration 0.010000
running time 0.004482

[1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0]

Timestep 2965567
mean reward (100 episodes) 1863.700000
best mean reward 2123.500000
episodes 11309
exploration 0.010000
running time 0.021858

[1180.0, 2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0]

Timestep 2965657
mean reward (100 episodes) 1863.700000
best mean reward 2123.500000
episodes 11309
exploration 0.010000
running time 0.009676

[2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0]

Timestep 2965735
mean reward (100 episodes) 1863.000000
best mean reward 2123.500000
episodes 11310
exploration 0.010000
running time 0.008040

[2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0]

Timestep 2965892
mean reward (100 episodes) 1863.000000
best mean reward 2123.500000
episodes 11310
exploration 0.010000
running time 0.016408

[2280.0, 2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0]

Timestep 2965914
mean reward (100 episodes) 1863.000000
best mean reward 2123.500000

episodes 11310
exploration 0.010000
running time 0.002269

[2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0]

Timestep 2965954
mean reward (100 episodes) 1857.100000
best mean reward 2123.500000
episodes 11311
exploration 0.010000
running time 0.004300

[2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0]

Timestep 2966086
mean reward (100 episodes) 1857.100000
best mean reward 2123.500000
episodes 11311
exploration 0.010000
running time 0.013714

[2600.0, 1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0]

Timestep 2966131
mean reward (100 episodes) 1857.100000
best mean reward 2123.500000
episodes 11311
exploration 0.010000
running time 0.004757

[1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0]

Timestep 2966183
mean reward (100 episodes) 1856.000000
best mean reward 2123.500000
episodes 11312
exploration 0.010000
running time 0.005773

[1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0]

Timestep 2966457
mean reward (100 episodes) 1856.000000
best mean reward 2123.500000
episodes 11312
exploration 0.010000
running time 0.029160

[1120.0, 1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0]

```
-----  
Timestep 2966510  
mean reward (100 episodes) 1856.000000  
best mean reward 2123.500000  
episodes 11312  
exploration 0.010000  
running time 0.005428  
-----  
[1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0]  
-----  
Timestep 2966615  
mean reward (100 episodes) 1857.900000  
best mean reward 2123.500000  
episodes 11313  
exploration 0.010000  
running time 0.010717  
-----  
[1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0]  
-----  
Timestep 2966750  
mean reward (100 episodes) 1857.900000  
best mean reward 2123.500000  
episodes 11313  
exploration 0.010000  
running time 0.013364  
-----  
[1990.0, 1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0]  
-----  
Timestep 2966790  
mean reward (100 episodes) 1857.900000  
best mean reward 2123.500000  
episodes 11313  
exploration 0.010000  
running time 0.004184  
-----  
[1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0]  
-----  
Timestep 2966918  
mean reward (100 episodes) 1856.400000  
best mean reward 2123.500000  
episodes 11314  
exploration 0.010000  
running time 0.013046  
-----  
[1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0]  
-----  
Timestep 2967016  
mean reward (100 episodes) 1856.400000  
best mean reward 2123.500000  
episodes 11314  
exploration 0.010000
```

running time 0.010005

[1510.0, 1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0]

Timestep 2967236

mean reward (100 episodes) 1856.400000

best mean reward 2123.500000

episodes 11314

exploration 0.010000

running time 0.021679

[1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0]

Timestep 2967281

mean reward (100 episodes) 1854.000000

best mean reward 2123.500000

episodes 11315

exploration 0.010000

running time 0.004495

[1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0]

Timestep 2967489

mean reward (100 episodes) 1854.000000

best mean reward 2123.500000

episodes 11315

exploration 0.010000

running time 0.021026

[1500.0, 1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0]

Timestep 2967542

mean reward (100 episodes) 1854.000000

best mean reward 2123.500000

episodes 11315

exploration 0.010000

running time 0.005470

[1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0]

Timestep 2967622

mean reward (100 episodes) 1859.500000

best mean reward 2123.500000

episodes 11316

exploration 0.010000

running time 0.007804

[1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0]

Timestep 2967727

mean reward (100 episodes) 1859.500000
best mean reward 2123.500000
episodes 11316
exploration 0.010000
running time 0.010652

[1480.0, 3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0]

Timestep 2967759
mean reward (100 episodes) 1859.500000
best mean reward 2123.500000
episodes 11316
exploration 0.010000
running time 0.003328

[3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0]

Timestep 2967816
mean reward (100 episodes) 1851.100000
best mean reward 2123.500000
episodes 11317
exploration 0.010000
running time 0.006043

[3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0]

Timestep 2967909
mean reward (100 episodes) 1851.100000
best mean reward 2123.500000
episodes 11317
exploration 0.010000
running time 0.009905

[3340.0, 3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0]

Timestep 2968095
mean reward (100 episodes) 1851.100000
best mean reward 2123.500000
episodes 11317
exploration 0.010000
running time 0.018379

[3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0]

Timestep 2968329
mean reward (100 episodes) 1849.600000
best mean reward 2123.500000
episodes 11318
exploration 0.010000
running time 0.023681

[3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0]

Timestep 2968461
mean reward (100 episodes) 1849.600000
best mean reward 2123.500000
episodes 11318
exploration 0.010000
running time 0.013310

[3430.0, 1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0]

Timestep 2968489
mean reward (100 episodes) 1849.600000
best mean reward 2123.500000
episodes 11318
exploration 0.010000
running time 0.003174

[1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0]

Timestep 2968691
mean reward (100 episodes) 1849.300000
best mean reward 2123.500000
episodes 11319
exploration 0.010000
running time 0.020074

[1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0]

Timestep 2968859
mean reward (100 episodes) 1849.300000
best mean reward 2123.500000
episodes 11319
exploration 0.010000
running time 0.017083

[1700.0, 1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0]

Timestep 2968907
mean reward (100 episodes) 1849.300000
best mean reward 2123.500000
episodes 11319
exploration 0.010000
running time 0.004859

[1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0]

Timestep 2969081
mean reward (100 episodes) 1849.400000
best mean reward 2123.500000

episodes 11320
exploration 0.010000
running time 0.017249

[1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0]

Timestep 2969196
mean reward (100 episodes) 1849.400000
best mean reward 2123.500000
episodes 11320
exploration 0.010000
running time 0.011589

[1250.0, 1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0]

Timestep 2969239
mean reward (100 episodes) 1849.400000
best mean reward 2123.500000
episodes 11320
exploration 0.010000
running time 0.004433

[1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0]

Timestep 2969330
mean reward (100 episodes) 1847.300000
best mean reward 2123.500000
episodes 11321
exploration 0.010000
running time 0.009174

[1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0]

Timestep 2969434
mean reward (100 episodes) 1847.300000
best mean reward 2123.500000
episodes 11321
exploration 0.010000
running time 0.010379

[1580.0, 2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0]

Timestep 2969528
mean reward (100 episodes) 1847.300000
best mean reward 2123.500000
episodes 11321
exploration 0.010000
running time 0.009663

[2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0]

```
-----  
Timestep 2969579  
mean reward (100 episodes) 1834.300000  
best mean reward 2123.500000  
episodes 11322  
exploration 0.010000  
running time 0.005074  
-----  
[2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0]  
-----  
Timestep 2969713  
mean reward (100 episodes) 1834.300000  
best mean reward 2123.500000  
episodes 11322  
exploration 0.010000  
running time 0.013703  
-----  
[2570.0, 1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0]  
-----  
Timestep 2969754  
mean reward (100 episodes) 1834.300000  
best mean reward 2123.500000  
episodes 11322  
exploration 0.010000  
running time 0.004111  
-----  
[1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0]  
-----  
Timestep 2969799  
mean reward (100 episodes) 1828.000000  
best mean reward 2123.500000  
episodes 11323  
exploration 0.010000  
running time 0.004428  
-----  
[1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0]  
-----  
Timestep 2969937  
mean reward (100 episodes) 1828.000000  
best mean reward 2123.500000  
episodes 11323  
exploration 0.010000  
running time 0.013728  
-----  
[1600.0, 1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0]  
-----  
Timestep 2969989  
mean reward (100 episodes) 1828.000000  
best mean reward 2123.500000  
episodes 11323  
exploration 0.010000
```

running time 0.005194

[1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0]

Timestep 2970025

mean reward (100 episodes) 1821.700000

best mean reward 2123.500000

episodes 11324

exploration 0.010000

running time 0.003609

[1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0]

Timestep 2970131

mean reward (100 episodes) 1821.700000

best mean reward 2123.500000

episodes 11324

exploration 0.010000

running time 0.010547

[1710.0, 2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0]

Timestep 2970252

mean reward (100 episodes) 1821.700000

best mean reward 2123.500000

episodes 11324

exploration 0.010000

running time 0.012176

[2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0]

Timestep 2970308

mean reward (100 episodes) 1821.000000

best mean reward 2123.500000

episodes 11325

exploration 0.010000

running time 0.005407

[2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0]

Timestep 2970445

mean reward (100 episodes) 1821.000000

best mean reward 2123.500000

episodes 11325

exploration 0.010000

running time 0.013510

[2320.0, 1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0]

Timestep 2970478

mean reward (100 episodes) 1821.000000
best mean reward 2123.500000
episodes 11325
exploration 0.010000
running time 0.003334

[1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0]

Timestep 2970509
mean reward (100 episodes) 1814.000000
best mean reward 2123.500000
episodes 11326
exploration 0.010000
running time 0.003404

[1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0]

Timestep 2970617
mean reward (100 episodes) 1814.000000
best mean reward 2123.500000
episodes 11326
exploration 0.010000
running time 0.010893

[1130.0, 1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0]

Timestep 2970796
mean reward (100 episodes) 1814.000000
best mean reward 2123.500000
episodes 11326
exploration 0.010000
running time 0.017712

[1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0]

Timestep 2970844
mean reward (100 episodes) 1810.000000
best mean reward 2123.500000
episodes 11327
exploration 0.010000
running time 0.005008

[1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0]

Timestep 2970954
mean reward (100 episodes) 1810.000000
best mean reward 2123.500000
episodes 11327
exploration 0.010000
running time 0.010774

[1790.0, 1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0]

Timestep 2971051
mean reward (100 episodes) 1810.000000
best mean reward 2123.500000
episodes 11327
exploration 0.010000
running time 0.009686

[1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0]

Timestep 2971097
mean reward (100 episodes) 1805.400000
best mean reward 2123.500000
episodes 11328
exploration 0.010000
running time 0.004931

[1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0]

Timestep 2971201
mean reward (100 episodes) 1805.400000
best mean reward 2123.500000
episodes 11328
exploration 0.010000
running time 0.010510

[1680.0, 2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0]

Timestep 2971298
mean reward (100 episodes) 1805.400000
best mean reward 2123.500000
episodes 11328
exploration 0.010000
running time 0.009790

[2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0]

Timestep 2971444
mean reward (100 episodes) 1805.100000
best mean reward 2123.500000
episodes 11329
exploration 0.010000
running time 0.014505

[2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0]

Timestep 2971629
mean reward (100 episodes) 1805.100000
best mean reward 2123.500000

episodes 11329
exploration 0.010000
running time 0.018518

[2090.0, 2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0]

Timestep 2971647
mean reward (100 episodes) 1805.100000
best mean reward 2123.500000
episodes 11329
exploration 0.010000
running time 0.001882

[2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0]

Timestep 2971710
mean reward (100 episodes) 1799.900000
best mean reward 2123.500000
episodes 11330
exploration 0.010000
running time 0.006304

[2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0]

Timestep 2971876
mean reward (100 episodes) 1799.900000
best mean reward 2123.500000
episodes 11330
exploration 0.010000
running time 0.016867

[2070.0, 1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0]

Timestep 2971891
mean reward (100 episodes) 1799.900000
best mean reward 2123.500000
episodes 11330
exploration 0.010000
running time 0.001653

[1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0]

Timestep 2971909
mean reward (100 episodes) 1801.500000
best mean reward 2123.500000
episodes 11331
exploration 0.010000
running time 0.001982

[1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0]

```
-----  
Timestep 2972050  
mean reward (100 episodes) 1801.500000  
best mean reward 2123.500000  
episodes 11331  
exploration 0.010000  
running time 0.013998  
-----  
[1420.0, 1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0]  
-----  
Timestep 2972110  
mean reward (100 episodes) 1801.500000  
best mean reward 2123.500000  
episodes 11331  
exploration 0.010000  
running time 0.006068  
-----  
[1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0]  
-----  
Timestep 2972149  
mean reward (100 episodes) 1805.700000  
best mean reward 2123.500000  
episodes 11332  
exploration 0.010000  
running time 0.003944  
-----  
[1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0]  
-----  
Timestep 2972354  
mean reward (100 episodes) 1805.700000  
best mean reward 2123.500000  
episodes 11332  
exploration 0.010000  
running time 0.020438  
-----  
[1340.0, 1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0]  
-----  
Timestep 2972465  
mean reward (100 episodes) 1805.700000  
best mean reward 2123.500000  
episodes 11332  
exploration 0.010000  
running time 0.010950  
-----  
[1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0]  
-----  
Timestep 2972522  
mean reward (100 episodes) 1798.800000  
best mean reward 2123.500000  
episodes 11333  
exploration 0.010000
```


running time 0.006283

[1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0]

Timestep 2972649

mean reward (100 episodes) 1798.800000

best mean reward 2123.500000

episodes 11333

exploration 0.010000

running time 0.012793

[1520.0, 1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0]

Timestep 2972707

mean reward (100 episodes) 1798.800000

best mean reward 2123.500000

episodes 11333

exploration 0.010000

running time 0.005992

[1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0]

Timestep 2972738

mean reward (100 episodes) 1794.500000

best mean reward 2123.500000

episodes 11334

exploration 0.010000

running time 0.003150

[1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0]

Timestep 2972844

mean reward (100 episodes) 1794.500000

best mean reward 2123.500000

episodes 11334

exploration 0.010000

running time 0.010796

[1990.0, 1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0]

Timestep 2972937

mean reward (100 episodes) 1794.500000

best mean reward 2123.500000

episodes 11334

exploration 0.010000

running time 0.009230

[1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0]

Timestep 2972992

mean reward (100 episodes) 1803.900000
best mean reward 2123.500000
episodes 11335
exploration 0.010000
running time 0.005623

[1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0]

Timestep 2973225
mean reward (100 episodes) 1803.900000
best mean reward 2123.500000
episodes 11335
exploration 0.010000
running time 0.023202

[1210.0, 1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0]

Timestep 2973265
mean reward (100 episodes) 1803.900000
best mean reward 2123.500000
episodes 11335
exploration 0.010000
running time 0.004254

[1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0]

Timestep 2973333
mean reward (100 episodes) 1789.700000
best mean reward 2123.500000
episodes 11336
exploration 0.010000
running time 0.007130

[1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0]

Timestep 2973450
mean reward (100 episodes) 1789.700000
best mean reward 2123.500000
episodes 11336
exploration 0.010000
running time 0.011839

[1310.0, 1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0]

Timestep 2973499
mean reward (100 episodes) 1789.700000
best mean reward 2123.500000
episodes 11336
exploration 0.010000
running time 0.004994

[1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0]

Timestep 2973554
mean reward (100 episodes) 1787.600000
best mean reward 2123.500000
episodes 11337
exploration 0.010000
running time 0.005728

[1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0]

Timestep 2973812
mean reward (100 episodes) 1787.600000
best mean reward 2123.500000
episodes 11337
exploration 0.010000
running time 0.025925

[1530.0, 2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0]

Timestep 2973868
mean reward (100 episodes) 1787.600000
best mean reward 2123.500000
episodes 11337
exploration 0.010000
running time 0.005754

[2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0]

Timestep 2973909
mean reward (100 episodes) 1784.000000
best mean reward 2123.500000
episodes 11338
exploration 0.010000
running time 0.004194

[2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0]

Timestep 2974028
mean reward (100 episodes) 1784.000000
best mean reward 2123.500000
episodes 11338
exploration 0.010000
running time 0.012052

[2090.0, 1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0]

Timestep 2974091
mean reward (100 episodes) 1784.000000
best mean reward 2123.500000

episodes 11338
exploration 0.010000
running time 0.006209

[1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0]

Timestep 2974124
mean reward (100 episodes) 1784.300000
best mean reward 2123.500000
episodes 11339
exploration 0.010000
running time 0.003437

[1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0]

Timestep 2974251
mean reward (100 episodes) 1784.300000
best mean reward 2123.500000
episodes 11339
exploration 0.010000
running time 0.012879

[1470.0, 1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0]

Timestep 2974366
mean reward (100 episodes) 1784.300000
best mean reward 2123.500000
episodes 11339
exploration 0.010000
running time 0.011704

[1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0]

Timestep 2974478
mean reward (100 episodes) 1807.500000
best mean reward 2123.500000
episodes 11340
exploration 0.010000
running time 0.010942

[1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0]

Timestep 2974589
mean reward (100 episodes) 1807.500000
best mean reward 2123.500000
episodes 11340
exploration 0.010000
running time 0.011026

[1390.0, 1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0]

```
-----  
Timestep 2974691  
mean reward (100 episodes) 1807.500000  
best mean reward 2123.500000  
episodes 11340  
exploration 0.010000  
running time 0.010272  
-----
```

```
[1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0]
```

```
-----  
Timestep 2974735  
mean reward (100 episodes) 1811.900000  
best mean reward 2123.500000  
episodes 11341  
exploration 0.010000  
running time 0.004394  
-----
```

```
[1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0]
```

```
-----  
Timestep 2974873  
mean reward (100 episodes) 1811.900000  
best mean reward 2123.500000  
episodes 11341  
exploration 0.010000  
running time 0.013864  
-----
```

```
[1410.0, 2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0]
```

```
-----  
Timestep 2975111  
mean reward (100 episodes) 1811.900000  
best mean reward 2123.500000  
episodes 11341  
exploration 0.010000  
running time 0.023865  
-----
```

```
[2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0]
```

```
-----  
Timestep 2975167  
mean reward (100 episodes) 1811.300000  
best mean reward 2123.500000  
episodes 11342  
exploration 0.010000  
running time 0.005880  
-----
```

```
[2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0]
```

```
-----  
Timestep 2975309  
mean reward (100 episodes) 1811.300000  
best mean reward 2123.500000  
episodes 11342  
exploration 0.010000  
-----
```

running time 0.014561

[2060.0, 1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0]

Timestep 2975411

mean reward (100 episodes) 1811.300000

best mean reward 2123.500000

episodes 11342

exploration 0.010000

running time 0.010236

[1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0]

Timestep 2975459

mean reward (100 episodes) 1809.600000

best mean reward 2123.500000

episodes 11343

exploration 0.010000

running time 0.004719

[1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0]

Timestep 2975595

mean reward (100 episodes) 1809.600000

best mean reward 2123.500000

episodes 11343

exploration 0.010000

running time 0.013487

[1290.0, 2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0]

Timestep 2975635

mean reward (100 episodes) 1809.600000

best mean reward 2123.500000

episodes 11343

exploration 0.010000

running time 0.003990

[2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0]

Timestep 2975658

mean reward (100 episodes) 1781.200000

best mean reward 2123.500000

episodes 11344

exploration 0.010000

running time 0.002417

[2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0]

Timestep 2975797

mean reward (100 episodes) 1781.200000
best mean reward 2123.500000
episodes 11344
exploration 0.010000
running time 0.013911

[2430.0, 2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0]

Timestep 2975984
mean reward (100 episodes) 1781.200000
best mean reward 2123.500000
episodes 11344
exploration 0.010000
running time 0.018365

[2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0]

Timestep 2976030
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11345
exploration 0.010000
running time 0.004784

[2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0]

Timestep 2976166
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11345
exploration 0.010000
running time 0.013429

[2210.0, 1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0]

Timestep 2976202
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11345
exploration 0.010000
running time 0.003757

[1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0]

Timestep 2976243
mean reward (100 episodes) 1782.500000
best mean reward 2123.500000
episodes 11346
exploration 0.010000
running time 0.004422

[1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0]

Timestep 2976344
mean reward (100 episodes) 1782.500000
best mean reward 2123.500000
episodes 11346
exploration 0.010000
running time 0.009949

[1480.0, 1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0]

Timestep 2976443
mean reward (100 episodes) 1782.500000
best mean reward 2123.500000
episodes 11346
exploration 0.010000
running time 0.009717

[1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0]

Timestep 2976490
mean reward (100 episodes) 1787.400000
best mean reward 2123.500000
episodes 11347
exploration 0.010000
running time 0.004744

[1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0]

Timestep 2976657
mean reward (100 episodes) 1787.400000
best mean reward 2123.500000
episodes 11347
exploration 0.010000
running time 0.016542

[1720.0, 1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0]

Timestep 2976767
mean reward (100 episodes) 1787.400000
best mean reward 2123.500000
episodes 11347
exploration 0.010000
running time 0.010958

[1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0]

Timestep 2976865
mean reward (100 episodes) 1784.200000
best mean reward 2123.500000

episodes 11348
exploration 0.010000
running time 0.009582

[1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0]

Timestep 2976981
mean reward (100 episodes) 1784.200000
best mean reward 2123.500000
episodes 11348
exploration 0.010000
running time 0.011535

[1540.0, 3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0]

Timestep 2977080
mean reward (100 episodes) 1784.200000
best mean reward 2123.500000
episodes 11348
exploration 0.010000
running time 0.009934

[3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0]

Timestep 2977113
mean reward (100 episodes) 1781.800000
best mean reward 2123.500000
episodes 11349
exploration 0.010000
running time 0.003992

[3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0]

Timestep 2977254
mean reward (100 episodes) 1781.800000
best mean reward 2123.500000
episodes 11349
exploration 0.010000
running time 0.014183

[3630.0, 1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0]

Timestep 2977302
mean reward (100 episodes) 1781.800000
best mean reward 2123.500000
episodes 11349
exploration 0.010000
running time 0.004691

[1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0]

```
-----  
Timestep 2977335  
mean reward (100 episodes) 1773.300000  
best mean reward 2123.500000  
episodes 11350  
exploration 0.010000  
running time 0.003259  
-----
```

```
[1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0]
```

```
-----  
Timestep 2977473  
mean reward (100 episodes) 1773.300000  
best mean reward 2123.500000  
episodes 11350  
exploration 0.010000  
running time 0.013822  
-----
```

```
[1650.0, 2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0]
```

```
-----  
Timestep 2977487  
mean reward (100 episodes) 1773.300000  
best mean reward 2123.500000  
episodes 11350  
exploration 0.010000  
running time 0.001673  
-----
```

```
[2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0]
```

```
-----  
Timestep 2977544  
mean reward (100 episodes) 1776.100000  
best mean reward 2123.500000  
episodes 11351  
exploration 0.010000  
running time 0.006055  
-----
```

```
[2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0]
```

```
-----  
Timestep 2977738  
mean reward (100 episodes) 1776.100000  
best mean reward 2123.500000  
episodes 11351  
exploration 0.010000  
running time 0.019310  
-----
```

```
[2020.0, 1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0]
```

```
-----  
Timestep 2977797  
mean reward (100 episodes) 1776.100000  
best mean reward 2123.500000  
episodes 11351  
exploration 0.010000  
-----
```

running time 0.005974

[1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0]

Timestep 2977982

mean reward (100 episodes) 1778.200000

best mean reward 2123.500000

episodes 11352

exploration 0.010000

running time 0.018177

[1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0]

Timestep 2978090

mean reward (100 episodes) 1778.200000

best mean reward 2123.500000

episodes 11352

exploration 0.010000

running time 0.010832

[1990.0, 1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0]

Timestep 2978173

mean reward (100 episodes) 1778.200000

best mean reward 2123.500000

episodes 11352

exploration 0.010000

running time 0.008083

[1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0]

Timestep 2978219

mean reward (100 episodes) 1774.100000

best mean reward 2123.500000

episodes 11353

exploration 0.010000

running time 0.004734

[1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0]

Timestep 2978466

mean reward (100 episodes) 1774.100000

best mean reward 2123.500000

episodes 11353

exploration 0.010000

running time 0.024461

[1670.0, 2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0]

Timestep 2978533

mean reward (100 episodes) 1774.100000
best mean reward 2123.500000
episodes 11353
exploration 0.010000
running time 0.006686

[2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0]

Timestep 2978643
mean reward (100 episodes) 1775.100000
best mean reward 2123.500000
episodes 11354
exploration 0.010000
running time 0.010719

[2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0]

Timestep 2978804
mean reward (100 episodes) 1775.100000
best mean reward 2123.500000
episodes 11354
exploration 0.010000
running time 0.016376

[2090.0, 1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0]

Timestep 2978844
mean reward (100 episodes) 1775.100000
best mean reward 2123.500000
episodes 11354
exploration 0.010000
running time 0.004093

[1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0]

Timestep 2978867
mean reward (100 episodes) 1777.500000
best mean reward 2123.500000
episodes 11355
exploration 0.010000
running time 0.002245

[1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0]

Timestep 2979003
mean reward (100 episodes) 1777.500000
best mean reward 2123.500000
episodes 11355
exploration 0.010000
running time 0.014000

[1320.0, 1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0]

Timestep 2979066
mean reward (100 episodes) 1777.500000
best mean reward 2123.500000
episodes 11355
exploration 0.010000
running time 0.006564

[1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0]

Timestep 2979267
mean reward (100 episodes) 1781.400000
best mean reward 2123.500000
episodes 11356
exploration 0.010000
running time 0.019610

[1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0]

Timestep 2979372
mean reward (100 episodes) 1781.400000
best mean reward 2123.500000
episodes 11356
exploration 0.010000
running time 0.010741

[1760.0, 1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0]

Timestep 2979462
mean reward (100 episodes) 1781.400000
best mean reward 2123.500000
episodes 11356
exploration 0.010000
running time 0.009181

[1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0]

Timestep 2979567
mean reward (100 episodes) 1771.800000
best mean reward 2123.500000
episodes 11357
exploration 0.010000
running time 0.010880

[1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0]

Timestep 2979795
mean reward (100 episodes) 1771.800000
best mean reward 2123.500000

episodes 11357
exploration 0.010000
running time 0.022755

[1880.0, 1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0]

Timestep 2979843
mean reward (100 episodes) 1771.800000
best mean reward 2123.500000
episodes 11357
exploration 0.010000
running time 0.004973

[1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0]

Timestep 2979882
mean reward (100 episodes) 1776.300000
best mean reward 2123.500000
episodes 11358
exploration 0.010000
running time 0.004067

[1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0]

Timestep 2980061
mean reward (100 episodes) 1776.300000
best mean reward 2123.500000
episodes 11358
exploration 0.010000
running time 0.017949

[1760.0, 1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0]

Timestep 2980122
mean reward (100 episodes) 1776.300000
best mean reward 2123.500000
episodes 11358
exploration 0.010000
running time 0.006068

[1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0]

Timestep 2980184
mean reward (100 episodes) 1762.300000
best mean reward 2123.500000
episodes 11359
exploration 0.010000
running time 0.006362

[1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0]

```
-----  
Timestep 2980457  
mean reward (100 episodes) 1762.300000  
best mean reward 2123.500000  
episodes 11359  
exploration 0.010000  
running time 0.027706  
-----
```

```
[1270.0, 2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0]
```

```
-----  
Timestep 2980503  
mean reward (100 episodes) 1762.300000  
best mean reward 2123.500000  
episodes 11359  
exploration 0.010000  
running time 0.004717  
-----
```

```
[2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0]
```

```
-----  
Timestep 2980582  
mean reward (100 episodes) 1768.500000  
best mean reward 2123.500000  
episodes 11360  
exploration 0.010000  
running time 0.007822  
-----
```

```
[2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0]
```

```
-----  
Timestep 2980704  
mean reward (100 episodes) 1768.500000  
best mean reward 2123.500000  
episodes 11360  
exploration 0.010000  
running time 0.012308  
-----
```

```
[2630.0, 1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0]
```

```
-----  
Timestep 2980810  
mean reward (100 episodes) 1768.500000  
best mean reward 2123.500000  
episodes 11360  
exploration 0.010000  
running time 0.010586  
-----
```

```
[1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0]
```

```
-----  
Timestep 2980879  
mean reward (100 episodes) 1777.600000  
best mean reward 2123.500000  
episodes 11361  
exploration 0.010000  
-----
```

running time 0.006973

[1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0]

Timestep 2981018

mean reward (100 episodes) 1777.600000

best mean reward 2123.500000

episodes 11361

exploration 0.010000

running time 0.014234

[1800.0, 1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0]

Timestep 2981065

mean reward (100 episodes) 1777.600000

best mean reward 2123.500000

episodes 11361

exploration 0.010000

running time 0.005045

[1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0]

Timestep 2981095

mean reward (100 episodes) 1769.000000

best mean reward 2123.500000

episodes 11362

exploration 0.010000

running time 0.003099

[1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0]

Timestep 2981286

mean reward (100 episodes) 1769.000000

best mean reward 2123.500000

episodes 11362

exploration 0.010000

running time 0.019252

[1760.0, 1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0]

Timestep 2981318

mean reward (100 episodes) 1769.000000

best mean reward 2123.500000

episodes 11362

exploration 0.010000

running time 0.003365

[1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0]

Timestep 2981370

mean reward (100 episodes) 1771.200000
best mean reward 2123.500000
episodes 11363
exploration 0.010000
running time 0.005149

[1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0]

Timestep 2981557
mean reward (100 episodes) 1771.200000
best mean reward 2123.500000
episodes 11363
exploration 0.010000
running time 0.018653

[1810.0, 1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0]

Timestep 2981576
mean reward (100 episodes) 1771.200000
best mean reward 2123.500000
episodes 11363
exploration 0.010000
running time 0.002141

[1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0]

Timestep 2981714
mean reward (100 episodes) 1774.600000
best mean reward 2123.500000
episodes 11364
exploration 0.010000
running time 0.013540

[1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0]

Timestep 2981910
mean reward (100 episodes) 1774.600000
best mean reward 2123.500000
episodes 11364
exploration 0.010000
running time 0.019959

[1490.0, 1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0]

Timestep 2981973
mean reward (100 episodes) 1774.600000
best mean reward 2123.500000
episodes 11364
exploration 0.010000
running time 0.006333

[1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0]

Timestep 2982074
mean reward (100 episodes) 1788.300000
best mean reward 2123.500000
episodes 11365
exploration 0.010000
running time 0.010603

[1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0]

Timestep 2982235
mean reward (100 episodes) 1788.300000
best mean reward 2123.500000
episodes 11365
exploration 0.010000
running time 0.015896

[1810.0, 1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0]

Timestep 2982293
mean reward (100 episodes) 1788.300000
best mean reward 2123.500000
episodes 11365
exploration 0.010000
running time 0.005841

[1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0]

Timestep 2982415
mean reward (100 episodes) 1784.900000
best mean reward 2123.500000
episodes 11366
exploration 0.010000
running time 0.012126

[1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0]

Timestep 2982537
mean reward (100 episodes) 1784.900000
best mean reward 2123.500000
episodes 11366
exploration 0.010000
running time 0.012437

[1520.0, 1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0]

Timestep 2982631
mean reward (100 episodes) 1784.900000
best mean reward 2123.500000

episodes 11366
exploration 0.010000
running time 0.009639

[1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0]

Timestep 2982686
mean reward (100 episodes) 1775.400000
best mean reward 2123.500000
episodes 11367
exploration 0.010000
running time 0.005568

[1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0]

Timestep 2982925
mean reward (100 episodes) 1775.400000
best mean reward 2123.500000
episodes 11367
exploration 0.010000
running time 0.023797

[1740.0, 1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0]

Timestep 2983035
mean reward (100 episodes) 1775.400000
best mean reward 2123.500000
episodes 11367
exploration 0.010000
running time 0.010761

[1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0]

Timestep 2983094
mean reward (100 episodes) 1805.900000
best mean reward 2123.500000
episodes 11368
exploration 0.010000
running time 0.005973

[1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0]

Timestep 2983198
mean reward (100 episodes) 1805.900000
best mean reward 2123.500000
episodes 11368
exploration 0.010000
running time 0.010490

[1600.0, 2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0]

```
-----  
Timestep 2983290  
mean reward (100 episodes) 1805.900000  
best mean reward 2123.500000  
episodes 11368  
exploration 0.010000  
running time 0.009231  
-----
```

```
[2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0]
```

```
-----  
Timestep 2983381  
mean reward (100 episodes) 1800.100000  
best mean reward 2123.500000  
episodes 11369  
exploration 0.010000  
running time 0.009151  
-----
```

```
[2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0]
```

```
-----  
Timestep 2983533  
mean reward (100 episodes) 1800.100000  
best mean reward 2123.500000  
episodes 11369  
exploration 0.010000  
running time 0.015224  
-----
```

```
[2250.0, 2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0]
```

```
-----  
Timestep 2983590  
mean reward (100 episodes) 1800.100000  
best mean reward 2123.500000  
episodes 11369  
exploration 0.010000  
running time 0.005937  
-----
```

```
[2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0]
```

```
-----  
Timestep 2983645  
mean reward (100 episodes) 1804.100000  
best mean reward 2123.500000  
episodes 11370  
exploration 0.010000  
running time 0.005463  
-----
```

```
[2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0]
```

```
-----  
Timestep 2983800  
mean reward (100 episodes) 1804.100000  
best mean reward 2123.500000  
episodes 11370  
exploration 0.010000  
-----
```

running time 0.015691

[2140.0, 1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0]

Timestep 2983815

mean reward (100 episodes) 1804.100000

best mean reward 2123.500000

episodes 11370

exploration 0.010000

running time 0.001623

[1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0]

Timestep 2983871

mean reward (100 episodes) 1808.400000

best mean reward 2123.500000

episodes 11371

exploration 0.010000

running time 0.005617

[1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0]

Timestep 2983992

mean reward (100 episodes) 1808.400000

best mean reward 2123.500000

episodes 11371

exploration 0.010000

running time 0.012391

[1260.0, 1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0]

Timestep 2984081

mean reward (100 episodes) 1808.400000

best mean reward 2123.500000

episodes 11371

exploration 0.010000

running time 0.009212

[1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0]

Timestep 2984129

mean reward (100 episodes) 1801.400000

best mean reward 2123.500000

episodes 11372

exploration 0.010000

running time 0.005112

[1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0]

Timestep 2984286

mean reward (100 episodes) 1801.400000
best mean reward 2123.500000
episodes 11372
exploration 0.010000
running time 0.015947

[1980.0, 1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0]

Timestep 2984326
mean reward (100 episodes) 1801.400000
best mean reward 2123.500000
episodes 11372
exploration 0.010000
running time 0.004267

[1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0]

Timestep 2984386
mean reward (100 episodes) 1789.400000
best mean reward 2123.500000
episodes 11373
exploration 0.010000
running time 0.006077

[1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0]

Timestep 2984515
mean reward (100 episodes) 1789.400000
best mean reward 2123.500000
episodes 11373
exploration 0.010000
running time 0.013134

[1880.0, 2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0]

Timestep 2984530
mean reward (100 episodes) 1789.400000
best mean reward 2123.500000
episodes 11373
exploration 0.010000
running time 0.001818

[2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0]

Timestep 2984544
mean reward (100 episodes) 1783.000000
best mean reward 2123.500000
episodes 11374
exploration 0.010000
running time 0.001743

[2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0]

Timestep 2984659
mean reward (100 episodes) 1783.000000
best mean reward 2123.500000
episodes 11374
exploration 0.010000
running time 0.011549

[2510.0, 1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0]

Timestep 2984704
mean reward (100 episodes) 1783.000000
best mean reward 2123.500000
episodes 11374
exploration 0.010000
running time 0.004560

[1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0]

Timestep 2984800
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11375
exploration 0.010000
running time 0.009781

[1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0]

Timestep 2984991
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11375
exploration 0.010000
running time 0.018941

[1700.0, 1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0]

Timestep 2985031
mean reward (100 episodes) 1786.300000
best mean reward 2123.500000
episodes 11375
exploration 0.010000
running time 0.004012

[1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0]

Timestep 2985101
mean reward (100 episodes) 1788.000000
best mean reward 2123.500000

episodes 11376
exploration 0.010000
running time 0.006940

[1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0]

Timestep 2985233
mean reward (100 episodes) 1788.000000
best mean reward 2123.500000
episodes 11376
exploration 0.010000
running time 0.013120

[1500.0, 4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0]

Timestep 2985287
mean reward (100 episodes) 1788.000000
best mean reward 2123.500000
episodes 11376
exploration 0.010000
running time 0.005394

[4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0]

Timestep 2985437
mean reward (100 episodes) 1805.600000
best mean reward 2123.500000
episodes 11377
exploration 0.010000
running time 0.014918

[4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0]

Timestep 2985651
mean reward (100 episodes) 1805.600000
best mean reward 2123.500000
episodes 11377
exploration 0.010000
running time 0.021467

[4350.0, 1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0]

Timestep 2985697
mean reward (100 episodes) 1805.600000
best mean reward 2123.500000
episodes 11377
exploration 0.010000
running time 0.004894

[1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0]


```
-----  
Timestep 2985737  
mean reward (100 episodes) 1807.400000  
best mean reward 2123.500000  
episodes 11378  
exploration 0.010000  
running time 0.004114  
-----  
[1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0]
```

```
-----  
Timestep 2985898  
mean reward (100 episodes) 1807.400000  
best mean reward 2123.500000  
episodes 11378  
exploration 0.010000  
running time 0.015914  
-----  
[1340.0, 1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0]
```

```
-----  
Timestep 2985993  
mean reward (100 episodes) 1807.400000  
best mean reward 2123.500000  
episodes 11378  
exploration 0.010000  
running time 0.009895  
-----  
[1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0]
```

```
-----  
Timestep 2986143  
mean reward (100 episodes) 1815.200000  
best mean reward 2123.500000  
episodes 11379  
exploration 0.010000  
running time 0.015685  
-----  
[1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0]
```

```
-----  
Timestep 2986268  
mean reward (100 episodes) 1815.200000  
best mean reward 2123.500000  
episodes 11379  
exploration 0.010000  
running time 0.013382  
-----  
[1480.0, 2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0]
```

```
-----  
Timestep 2986306  
mean reward (100 episodes) 1815.200000  
best mean reward 2123.500000  
episodes 11379  
exploration 0.010000
```

```
running time 0.003958
-----
[2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0]
-----
Timestep 2986423
mean reward (100 episodes) 1818.800000
best mean reward 2123.500000
episodes 11380
exploration 0.010000
running time 0.012435
-----
[2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0]
-----
Timestep 2986552
mean reward (100 episodes) 1818.800000
best mean reward 2123.500000
episodes 11380
exploration 0.010000
running time 0.012737
-----
[2010.0, 1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0]
-----
Timestep 2986585
mean reward (100 episodes) 1818.800000
best mean reward 2123.500000
episodes 11380
exploration 0.010000
running time 0.003608
-----
[1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0]
-----
Timestep 2986638
mean reward (100 episodes) 1818.700000
best mean reward 2123.500000
episodes 11381
exploration 0.010000
running time 0.005769
-----
[1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0]
-----
Timestep 2986770
mean reward (100 episodes) 1818.700000
best mean reward 2123.500000
episodes 11381
exploration 0.010000
running time 0.013327
-----
[1560.0, 1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0]
-----
Timestep 2986800
```

mean reward (100 episodes) 1818.700000
best mean reward 2123.500000
episodes 11381
exploration 0.010000
running time 0.002985

[1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0]

Timestep 2986906
mean reward (100 episodes) 1847.900000
best mean reward 2123.500000
episodes 11382
exploration 0.010000
running time 0.010759

[1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0]

Timestep 2987066
mean reward (100 episodes) 1847.900000
best mean reward 2123.500000
episodes 11382
exploration 0.010000
running time 0.015926

[1200.0, 810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0]

Timestep 2987127
mean reward (100 episodes) 1847.900000
best mean reward 2123.500000
episodes 11382
exploration 0.010000
running time 0.005966

[810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0]

Timestep 2987139
mean reward (100 episodes) 1832.300000
best mean reward 2123.500000
episodes 11383
exploration 0.010000
running time 0.001457

[810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0]

Timestep 2987232
mean reward (100 episodes) 1832.300000
best mean reward 2123.500000
episodes 11383
exploration 0.010000
running time 0.009428

[810.0, 1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0]

Timestep 2987348
mean reward (100 episodes) 1832.300000
best mean reward 2123.500000
episodes 11383
exploration 0.010000
running time 0.011697

[1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0]

Timestep 2987391
mean reward (100 episodes) 1827.900000
best mean reward 2123.500000
episodes 11384
exploration 0.010000
running time 0.004532

[1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0]

Timestep 2987517
mean reward (100 episodes) 1827.900000
best mean reward 2123.500000
episodes 11384
exploration 0.010000
running time 0.012744

[1840.0, 1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0]

Timestep 2987556
mean reward (100 episodes) 1827.900000
best mean reward 2123.500000
episodes 11384
exploration 0.010000
running time 0.004019

[1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0]

Timestep 2987658
mean reward (100 episodes) 1841.700000
best mean reward 2123.500000
episodes 11385
exploration 0.010000
running time 0.010177

[1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0]

Timestep 2987760
mean reward (100 episodes) 1841.700000
best mean reward 2123.500000

episodes 11385
exploration 0.010000
running time 0.010456

[1800.0, 3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0]

Timestep 2987852
mean reward (100 episodes) 1841.700000
best mean reward 2123.500000
episodes 11385
exploration 0.010000
running time 0.009046

[3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0]

Timestep 2987969
mean reward (100 episodes) 1845.700000
best mean reward 2123.500000
episodes 11386
exploration 0.010000
running time 0.011815

[3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0]

Timestep 2988087
mean reward (100 episodes) 1845.700000
best mean reward 2123.500000
episodes 11386
exploration 0.010000
running time 0.011996

[3250.0, 1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0]

Timestep 2988161
mean reward (100 episodes) 1845.700000
best mean reward 2123.500000
episodes 11386
exploration 0.010000
running time 0.007455

[1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0]

Timestep 2988256
mean reward (100 episodes) 1860.400000
best mean reward 2123.500000
episodes 11387
exploration 0.010000
running time 0.009585

[1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0]

```
-----  
Timestep 2988366  
mean reward (100 episodes) 1860.400000  
best mean reward 2123.500000  
episodes 11387  
exploration 0.010000  
running time 0.011131  
-----
```

```
[1700.0, 2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0]
```

```
-----  
Timestep 2988446  
mean reward (100 episodes) 1860.400000  
best mean reward 2123.500000  
episodes 11387  
exploration 0.010000  
running time 0.008201  
-----
```

```
[2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0]
```

```
-----  
Timestep 2988501  
mean reward (100 episodes) 1856.000000  
best mean reward 2123.500000  
episodes 11388  
exploration 0.010000  
running time 0.005442  
-----
```

```
[2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0]
```

```
-----  
Timestep 2988585  
mean reward (100 episodes) 1856.000000  
best mean reward 2123.500000  
episodes 11388  
exploration 0.010000  
running time 0.008692  
-----
```

```
[2210.0, 2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0]
```

```
-----  
Timestep 2988671  
mean reward (100 episodes) 1856.000000  
best mean reward 2123.500000  
episodes 11388  
exploration 0.010000  
running time 0.008534  
-----
```

```
[2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0]
```

```
-----  
Timestep 2988755  
mean reward (100 episodes) 1856.100000  
best mean reward 2123.500000  
episodes 11389  
exploration 0.010000  
-----
```

running time 0.008708

[2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0]

Timestep 2988934

mean reward (100 episodes) 1856.100000

best mean reward 2123.500000

episodes 11389

exploration 0.010000

running time 0.018037

[2300.0, 1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0]

Timestep 2988948

mean reward (100 episodes) 1856.100000

best mean reward 2123.500000

episodes 11389

exploration 0.010000

running time 0.001647

[1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0]

Timestep 2988962

mean reward (100 episodes) 1866.900000

best mean reward 2123.500000

episodes 11390

exploration 0.010000

running time 0.001585

[1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0]

Timestep 2989158

mean reward (100 episodes) 1866.900000

best mean reward 2123.500000

episodes 11390

exploration 0.010000

running time 0.019852

[1370.0, 4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0]

Timestep 2989212

mean reward (100 episodes) 1866.900000

best mean reward 2123.500000

episodes 11390

exploration 0.010000

running time 0.005624

[4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0]

Timestep 2989351

mean reward (100 episodes) 1871.400000
best mean reward 2123.500000
episodes 11391
exploration 0.010000
running time 0.014017

[4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0]

Timestep 2989453
mean reward (100 episodes) 1871.400000
best mean reward 2123.500000
episodes 11391
exploration 0.010000
running time 0.010583

[4380.0, 1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0]

Timestep 2989673
mean reward (100 episodes) 1871.400000
best mean reward 2123.500000
episodes 11391
exploration 0.010000
running time 0.022021

[1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0]

Timestep 2989784
mean reward (100 episodes) 1879.100000
best mean reward 2123.500000
episodes 11392
exploration 0.010000
running time 0.011148

[1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0]

Timestep 2989875
mean reward (100 episodes) 1879.100000
best mean reward 2123.500000
episodes 11392
exploration 0.010000
running time 0.008826

[1170.0, 1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0]

Timestep 2989954
mean reward (100 episodes) 1879.100000
best mean reward 2123.500000
episodes 11392
exploration 0.010000
running time 0.008308

[1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0]

Timestep 2990021
mean reward (100 episodes) 1886.700000
best mean reward 2123.500000
episodes 11393
exploration 0.010000
running time 0.007057

[1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0]

Timestep 2990146
mean reward (100 episodes) 1886.700000
best mean reward 2123.500000
episodes 11393
exploration 0.010000
running time 0.012520

[1650.0, 2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0]

Timestep 2990188
mean reward (100 episodes) 1886.700000
best mean reward 2123.500000
episodes 11393
exploration 0.010000
running time 0.004517

[2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0]

Timestep 2990244
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000
episodes 11394
exploration 0.010000
running time 0.005879

[2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0]

Timestep 2990386
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000
episodes 11394
exploration 0.010000
running time 0.013994

[2460.0, 2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0]

Timestep 2990453
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000

episodes 11394
exploration 0.010000
running time 0.006925

[2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0]

Timestep 2990513
mean reward (100 episodes) 1884.500000
best mean reward 2123.500000
episodes 11395
exploration 0.010000
running time 0.006125

[2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0]

Timestep 2990632
mean reward (100 episodes) 1884.500000
best mean reward 2123.500000
episodes 11395
exploration 0.010000
running time 0.012186

[2270.0, 2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0]

Timestep 2990724
mean reward (100 episodes) 1884.500000
best mean reward 2123.500000
episodes 11395
exploration 0.010000
running time 0.009377

[2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0]

Timestep 2990769
mean reward (100 episodes) 1880.400000
best mean reward 2123.500000
episodes 11396
exploration 0.010000
running time 0.004697

[2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0]

Timestep 2990942
mean reward (100 episodes) 1880.400000
best mean reward 2123.500000
episodes 11396
exploration 0.010000
running time 0.017345

[2990.0, 1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0]

```
-----  
Timestep 2990985  
mean reward (100 episodes) 1880.400000  
best mean reward 2123.500000  
episodes 11396  
exploration 0.010000  
running time 0.004718  
-----  
[1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0]  
-----  
Timestep 2991004  
mean reward (100 episodes) 1877.800000  
best mean reward 2123.500000  
episodes 11397  
exploration 0.010000  
running time 0.002542  
-----  
[1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0]  
-----  
Timestep 2991235  
mean reward (100 episodes) 1877.800000  
best mean reward 2123.500000  
episodes 11397  
exploration 0.010000  
running time 0.023167  
-----  
[1400.0, 1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0]  
-----  
Timestep 2991339  
mean reward (100 episodes) 1877.800000  
best mean reward 2123.500000  
episodes 11397  
exploration 0.010000  
running time 0.010472  
-----  
[1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0]  
-----  
Timestep 2991407  
mean reward (100 episodes) 1880.900000  
best mean reward 2123.500000  
episodes 11398  
exploration 0.010000  
running time 0.006942  
-----  
[1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0]  
-----  
Timestep 2991510  
mean reward (100 episodes) 1880.900000  
best mean reward 2123.500000  
episodes 11398  
exploration 0.010000
```

running time 0.010374

[1410.0, 2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0]

Timestep 2991546

mean reward (100 episodes) 1880.900000

best mean reward 2123.500000

episodes 11398

exploration 0.010000

running time 0.003649

[2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0]

Timestep 2991642

mean reward (100 episodes) 1875.300000

best mean reward 2123.500000

episodes 11399

exploration 0.010000

running time 0.009951

[2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0]

Timestep 2991761

mean reward (100 episodes) 1875.300000

best mean reward 2123.500000

episodes 11399

exploration 0.010000

running time 0.012285

[2760.0, 2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0]

Timestep 2991923

mean reward (100 episodes) 1875.300000

best mean reward 2123.500000

episodes 11399

exploration 0.010000

running time 0.016306

[2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0]

Timestep 2991970

mean reward (100 episodes) 1885.900000

best mean reward 2123.500000

episodes 11400

exploration 0.010000

running time 0.004622

[2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0]

Timestep 2992073

mean reward (100 episodes) 1885.900000
best mean reward 2123.500000
episodes 11400
exploration 0.010000
running time 0.010655

[2320.0, 2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0]

Timestep 2992207
mean reward (100 episodes) 1885.900000
best mean reward 2123.500000
episodes 11400
exploration 0.010000
running time 0.013109

[2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0]

Timestep 2992329
mean reward (100 episodes) 1882.300000
best mean reward 2123.500000
episodes 11401
exploration 0.010000
running time 0.012088

[2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0]

Timestep 2992484
mean reward (100 episodes) 1882.300000
best mean reward 2123.500000
episodes 11401
exploration 0.010000
running time 0.015940

[2740.0, 1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0]

Timestep 2992543
mean reward (100 episodes) 1882.300000
best mean reward 2123.500000
episodes 11401
exploration 0.010000
running time 0.006114

[1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0]

Timestep 2992594
mean reward (100 episodes) 1877.500000
best mean reward 2123.500000
episodes 11402
exploration 0.010000
running time 0.005350

[1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0]

Timestep 2992690
mean reward (100 episodes) 1877.500000
best mean reward 2123.500000
episodes 11402
exploration 0.010000
running time 0.009652

[1990.0, 1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0]

Timestep 2992775
mean reward (100 episodes) 1877.500000
best mean reward 2123.500000
episodes 11402
exploration 0.010000
running time 0.008281

[1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0]

Timestep 2992821
mean reward (100 episodes) 1885.800000
best mean reward 2123.500000
episodes 11403
exploration 0.010000
running time 0.004739

[1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0]

Timestep 2992992
mean reward (100 episodes) 1885.800000
best mean reward 2123.500000
episodes 11403
exploration 0.010000
running time 0.017157

[1460.0, 1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0]

Timestep 2993046
mean reward (100 episodes) 1885.800000
best mean reward 2123.500000
episodes 11403
exploration 0.010000
running time 0.005242

[1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0]

Timestep 2993074
mean reward (100 episodes) 1881.200000
best mean reward 2123.500000

episodes 11404
exploration 0.010000
running time 0.003070

[1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0]

Timestep 2993164
mean reward (100 episodes) 1881.200000
best mean reward 2123.500000
episodes 11404
exploration 0.010000
running time 0.009296

[1500.0, 1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0]

Timestep 2993215
mean reward (100 episodes) 1881.200000
best mean reward 2123.500000
episodes 11404
exploration 0.010000
running time 0.005242

[1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0]

Timestep 2993393
mean reward (100 episodes) 1900.700000
best mean reward 2123.500000
episodes 11405
exploration 0.010000
running time 0.017915

[1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0]

Timestep 2993526
mean reward (100 episodes) 1900.700000
best mean reward 2123.500000
episodes 11405
exploration 0.010000
running time 0.013120

[1500.0, 1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0]

Timestep 2993618
mean reward (100 episodes) 1900.700000
best mean reward 2123.500000
episodes 11405
exploration 0.010000
running time 0.009243

[1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0]

```
-----  
Timestep 2993675  
mean reward (100 episodes) 1903.200000  
best mean reward 2123.500000  
episodes 11406  
exploration 0.010000  
running time 0.005776  
-----
```

```
[1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0]
```

```
-----  
Timestep 2993800  
mean reward (100 episodes) 1903.200000  
best mean reward 2123.500000  
episodes 11406  
exploration 0.010000  
running time 0.012822  
-----
```

```
[1280.0, 1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0]
```

```
-----  
Timestep 2993845  
mean reward (100 episodes) 1903.200000  
best mean reward 2123.500000  
episodes 11406  
exploration 0.010000  
running time 0.004482  
-----
```

```
[1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0]
```

```
-----  
Timestep 2993952  
mean reward (100 episodes) 1907.100000  
best mean reward 2123.500000  
episodes 11407  
exploration 0.010000  
running time 0.011357  
-----
```

```
[1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0]
```

```
-----  
Timestep 2994122  
mean reward (100 episodes) 1907.100000  
best mean reward 2123.500000  
episodes 11407  
exploration 0.010000  
running time 0.017416  
-----
```

```
[1970.0, 1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0]
```

```
-----  
Timestep 2994168  
mean reward (100 episodes) 1907.100000  
best mean reward 2123.500000  
episodes 11407  
exploration 0.010000  
-----
```


running time 0.004863

[1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0]

Timestep 2994213

mean reward (100 episodes) 1894.900000

best mean reward 2123.500000

episodes 11408

exploration 0.010000

running time 0.004745

[1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0]

Timestep 2994375

mean reward (100 episodes) 1894.900000

best mean reward 2123.500000

episodes 11408

exploration 0.010000

running time 0.016494

[1310.0, 2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0]

Timestep 2994446

mean reward (100 episodes) 1894.900000

best mean reward 2123.500000

episodes 11408

exploration 0.010000

running time 0.007440

[2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0]

Timestep 2994509

mean reward (100 episodes) 1879.100000

best mean reward 2123.500000

episodes 11409

exploration 0.010000

running time 0.006513

[2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0]

Timestep 2994621

mean reward (100 episodes) 1879.100000

best mean reward 2123.500000

episodes 11409

exploration 0.010000

running time 0.011505

[2240.0, 1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0]

Timestep 2994659

mean reward (100 episodes) 1879.100000
best mean reward 2123.500000
episodes 11409
exploration 0.010000
running time 0.003875

[1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0]

Timestep 2994814
mean reward (100 episodes) 1876.400000
best mean reward 2123.500000
episodes 11410
exploration 0.010000
running time 0.015931

[1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0]

Timestep 2995033
mean reward (100 episodes) 1876.400000
best mean reward 2123.500000
episodes 11410
exploration 0.010000
running time 0.021923

[1920.0, 2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0]

Timestep 2995063
mean reward (100 episodes) 1876.400000
best mean reward 2123.500000
episodes 11410
exploration 0.010000
running time 0.003206

[2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0]

Timestep 2995112
mean reward (100 episodes) 1884.400000
best mean reward 2123.500000
episodes 11411
exploration 0.010000
running time 0.005301

[2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0]

Timestep 2995204
mean reward (100 episodes) 1884.400000
best mean reward 2123.500000
episodes 11411
exploration 0.010000
running time 0.009493

[2120.0, 1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0]

Timestep 2995350
mean reward (100 episodes) 1884.400000
best mean reward 2123.500000
episodes 11411
exploration 0.010000
running time 0.014435

[1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0]

Timestep 2995410
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000
episodes 11412
exploration 0.010000
running time 0.006098

[1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0]

Timestep 2995686
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000
episodes 11412
exploration 0.010000
running time 0.027264

[1950.0, 1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0]

Timestep 2995721
mean reward (100 episodes) 1887.200000
best mean reward 2123.500000
episodes 11412
exploration 0.010000
running time 0.003689

[1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0]

Timestep 2995837
mean reward (100 episodes) 1878.700000
best mean reward 2123.500000
episodes 11413
exploration 0.010000
running time 0.011953

[1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0]

Timestep 2995964
mean reward (100 episodes) 1878.700000
best mean reward 2123.500000

episodes 11413
exploration 0.010000
running time 0.012778

[1530.0, 3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0]

Timestep 2995980
mean reward (100 episodes) 1878.700000
best mean reward 2123.500000
episodes 11413
exploration 0.010000
running time 0.001779

[3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0]

Timestep 2996052
mean reward (100 episodes) 1874.400000
best mean reward 2123.500000
episodes 11414
exploration 0.010000
running time 0.007318

[3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0]

Timestep 2996152
mean reward (100 episodes) 1874.400000
best mean reward 2123.500000
episodes 11414
exploration 0.010000
running time 0.009894

[3460.0, 1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0]

Timestep 2996304
mean reward (100 episodes) 1874.400000
best mean reward 2123.500000
episodes 11414
exploration 0.010000
running time 0.015219

[1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0]

Timestep 2996347
mean reward (100 episodes) 1876.000000
best mean reward 2123.500000
episodes 11415
exploration 0.010000
running time 0.004245

[1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0]

```
-----  
Timestep 2996430  
mean reward (100 episodes) 1876.000000  
best mean reward 2123.500000  
episodes 11415  
exploration 0.010000  
running time 0.008503  
-----  
[1750.0, 1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0]  
-----  
Timestep 2996519  
mean reward (100 episodes) 1876.000000  
best mean reward 2123.500000  
episodes 11415  
exploration 0.010000  
running time 0.009205  
-----  
[1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0]  
-----  
Timestep 2996547  
mean reward (100 episodes) 1866.200000  
best mean reward 2123.500000  
episodes 11416  
exploration 0.010000  
running time 0.003106  
-----  
[1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0]  
-----  
Timestep 2996740  
mean reward (100 episodes) 1866.200000  
best mean reward 2123.500000  
episodes 11416  
exploration 0.010000  
running time 0.019288  
-----  
[1870.0, 2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0]  
-----  
Timestep 2996828  
mean reward (100 episodes) 1866.200000  
best mean reward 2123.500000  
episodes 11416  
exploration 0.010000  
running time 0.008807  
-----  
[2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0]  
-----  
Timestep 2996865  
mean reward (100 episodes) 1872.100000  
best mean reward 2123.500000  
episodes 11417  
exploration 0.010000
```

running time 0.003552

[2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0]

Timestep 2996953

mean reward (100 episodes) 1872.100000

best mean reward 2123.500000

episodes 11417

exploration 0.010000

running time 0.009000

[2120.0, 1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0]

Timestep 2997048

mean reward (100 episodes) 1872.100000

best mean reward 2123.500000

episodes 11417

exploration 0.010000

running time 0.009557

[1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0]

Timestep 2997151

mean reward (100 episodes) 1877.400000

best mean reward 2123.500000

episodes 11418

exploration 0.010000

running time 0.010107

[1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0]

Timestep 2997382

mean reward (100 episodes) 1877.400000

best mean reward 2123.500000

episodes 11418

exploration 0.010000

running time 0.023071

[1850.0, 1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0]

Timestep 2997430

mean reward (100 episodes) 1877.400000

best mean reward 2123.500000

episodes 11418

exploration 0.010000

running time 0.004976

[1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0]

Timestep 2997505

mean reward (100 episodes) 1880.500000
best mean reward 2123.500000
episodes 11419
exploration 0.010000
running time 0.007508

[1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0]

Timestep 2997664
mean reward (100 episodes) 1880.500000
best mean reward 2123.500000
episodes 11419
exploration 0.010000
running time 0.016225

[1430.0, 2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0]

Timestep 2997725
mean reward (100 episodes) 1880.500000
best mean reward 2123.500000
episodes 11419
exploration 0.010000
running time 0.006072

[2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0]

Timestep 2997757
mean reward (100 episodes) 1873.900000
best mean reward 2123.500000
episodes 11420
exploration 0.010000
running time 0.003365

[2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0]

Timestep 2997907
mean reward (100 episodes) 1873.900000
best mean reward 2123.500000
episodes 11420
exploration 0.010000
running time 0.015238

[2050.0, 1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0]

Timestep 2997955
mean reward (100 episodes) 1873.900000
best mean reward 2123.500000
episodes 11420
exploration 0.010000
running time 0.004836

[1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0]

Timestep 2998007
mean reward (100 episodes) 1865.400000
best mean reward 2123.500000
episodes 11421
exploration 0.010000
running time 0.005252

[1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0]

Timestep 2998130
mean reward (100 episodes) 1865.400000
best mean reward 2123.500000
episodes 11421
exploration 0.010000
running time 0.012451

[1860.0, 1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0]

Timestep 2998200
mean reward (100 episodes) 1865.400000
best mean reward 2123.500000
episodes 11421
exploration 0.010000
running time 0.006935

[1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0]

Timestep 2998260
mean reward (100 episodes) 1865.100000
best mean reward 2123.500000
episodes 11422
exploration 0.010000
running time 0.006232

[1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0]

Timestep 2998459
mean reward (100 episodes) 1865.100000
best mean reward 2123.500000
episodes 11422
exploration 0.010000
running time 0.019563

[1720.0, 1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0]

Timestep 2998514
mean reward (100 episodes) 1865.100000
best mean reward 2123.500000

episodes 11422
exploration 0.010000
running time 0.005474

[1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0]

Timestep 2998561
mean reward (100 episodes) 1865.700000
best mean reward 2123.500000
episodes 11423
exploration 0.010000
running time 0.004707

[1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0]

Timestep 2998757
mean reward (100 episodes) 1865.700000
best mean reward 2123.500000
episodes 11423
exploration 0.010000
running time 0.019625

[1170.0, 1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0]

Timestep 2998790
mean reward (100 episodes) 1865.700000
best mean reward 2123.500000
episodes 11423
exploration 0.010000
running time 0.003458

[1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0]

Timestep 2998929
mean reward (100 episodes) 1883.700000
best mean reward 2123.500000
episodes 11424
exploration 0.010000
running time 0.013643

[1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0]

Timestep 2999056
mean reward (100 episodes) 1883.700000
best mean reward 2123.500000
episodes 11424
exploration 0.010000
running time 0.012637

[1870.0, 1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0]

```
-----  
Timestep 2999102  
mean reward (100 episodes) 1883.700000  
best mean reward 2123.500000  
episodes 11424  
exploration 0.010000  
running time 0.004761  
-----
```

```
[1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0]
```

```
-----  
Timestep 2999143  
mean reward (100 episodes) 1877.200000  
best mean reward 2123.500000  
episodes 11425  
exploration 0.010000  
running time 0.004334  
-----
```

```
[1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0]
```

```
-----  
Timestep 2999260  
mean reward (100 episodes) 1877.200000  
best mean reward 2123.500000  
episodes 11425  
exploration 0.010000  
running time 0.011721  
-----
```

```
[1340.0, 1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0]
```

```
-----  
Timestep 2999390  
mean reward (100 episodes) 1877.200000  
best mean reward 2123.500000  
episodes 11425  
exploration 0.010000  
running time 0.012867  
-----
```

```
[1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0, 1840.0]
```

```
-----  
Timestep 2999459  
mean reward (100 episodes) 1883.500000  
best mean reward 2123.500000  
episodes 11426  
exploration 0.010000  
running time 0.006945  
-----
```

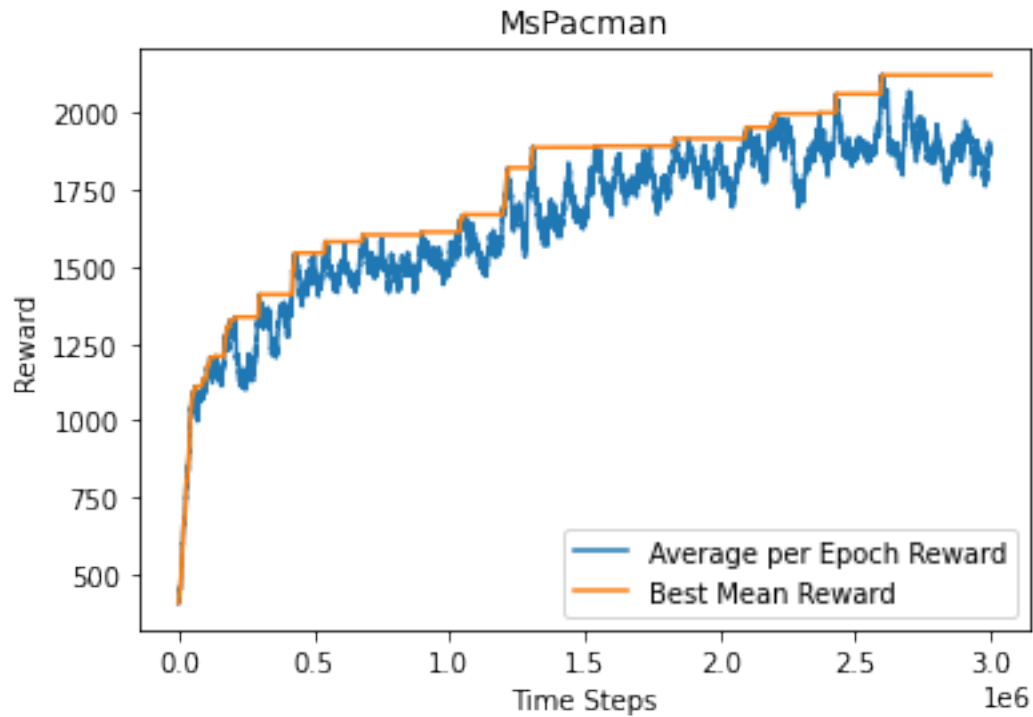
```
[1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0, 1840.0]
```

```
-----  
Timestep 2999661  
mean reward (100 episodes) 1883.500000  
best mean reward 2123.500000  
episodes 11426  
exploration 0.010000  
-----
```

```
running time 0.020031
-----
[1720.0, 2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0, 1840.0]
-----
Timestep 2999709
mean reward (100 episodes) 1883.500000
best mean reward 2123.500000
episodes 11426
exploration 0.010000
running time 0.004914
-----
[2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0, 1840.0, 1550.0]
-----
Timestep 2999748
mean reward (100 episodes) 1885.900000
best mean reward 2123.500000
episodes 11427
exploration 0.010000
running time 0.004156
-----
[2320.0, 1990.0, 1430.0, 1220.0, 1390.0, 1400.0, 3320.0, 1340.0, 1840.0, 1550.0]
-----
Timestep 2999965
mean reward (100 episodes) 1885.900000
best mean reward 2123.500000
episodes 11427
exploration 0.010000
running time 0.021556
-----

total time: 18287.27543926239
```

```
In [ ]: #@title plot learning curves
        # plot results
        rl_logger.plot(env.spec._env_name)
```



1. Questions

1. Try learning without the target network. What do you observe?
2. Play with the hyperparameters to see if you can substantially improve the performance of the algorithm.
3. Implement Double DQN; Does it perform better?
4. Check if the Q-network makes correct predictions, i.e. if the predicted expected return matches the observed values.
5. Try out Boltzmann exploration, where instead of an ε -greedy policy, one can use $\pi_\beta \propto \exp(Q(s, a)/T)$ for some temperature T which follows some decay schedule.