**Ex1\_model\_free\_rl**

SARSA Mean Training Reward: 0.0702 | Mean Test Reward: 0.0712 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.548 | Mean Test Reward: 0.7788 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.999

SARSA Mean Training Reward: 0.5942 | Mean Test Reward: 0.625 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.7101 | Mean Test Reward: 0.522 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.4736 | Mean Test Reward: 0.526 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.999

SARSA Mean Training Reward: 0.6943 | Mean Test Reward: 0.7466 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.4476 | Mean Test Reward: 0.8098 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.2241 | Mean Test Reward: 0.2368 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.6827 | Mean Test Reward: 0.7774 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.5374 | Mean Test Reward: 0.781 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.7073 | Mean Test Reward: 0.7832 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.501 | Mean Test Reward: 0.8202 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.687 | Mean Test Reward: 0.7356 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.6827 | Mean Test Reward: 0.7872 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.4679 | Mean Test Reward: 0.7386 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.99

Expected SARSA Mean Training Reward: 0.0255 | Mean Test Reward: 0.025 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.995

C:\Users\nikla\Documents\Uni\MitStefRL\ReinforcementLearning\assignment\_2\code\_skeleton\ex1\_model\_free\_rl.py:85: RuntimeWarning: overflow encountered in multiply

expected\_next\_q = np.sum(self.Q[next\_state] \* self.policy(next\_state, is\_training=False))

C:\Users\nikla\Documents\Uni\MitStefRL\ReinforcementLearning\assignment\_2\code\_skeleton\ex1\_model\_free\_rl.py:87: RuntimeWarning: invalid value encountered in scalar add

self.Q[state][action] += self.alpha \* (target - self.Q[state][action])

Expected SARSA Mean Training Reward: 0.0028 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.1, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0003 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.99

C:\Users\nikla\Documents\Uni\MitStefRL\ReinforcementLearning\assignment\_2\code\_skeleton\ex1\_model\_free\_rl.py:87: RuntimeWarning: invalid value encountered in scalar subtract

self.Q[state][action] += self.alpha \* (target - self.Q[state][action])

Expected SARSA Mean Training Reward: 0.0021 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.995

Expected SARSA Mean Training Reward: 0.0018 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.5, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0003 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.99

Expected SARSA Mean Training Reward: 0.0016 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.995

C:\Users\nikla\Documents\Uni\venv\Lib\site-packages\numpy\core\fromnumeric.py:88: RuntimeWarning: overflow encountered in reduce

return ufunc.reduce(obj, axis, dtype, out, \*\*passkwargs)

Expected SARSA Mean Training Reward: 0.0014 | Mean Test Reward: 0.0 | gamma=0.95, alpha=0.9, eps=1, eps\_decay=0.999

SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.569 | Mean Test Reward: 0.61 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.5873 | Mean Test Reward: 0.6218 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.999

SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.1412 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.4565 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.999

SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.99

SARSA Mean Training Reward: 0.1136 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.995

SARSA Mean Training Reward: 0.3854 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.0894 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.7406 | Mean Test Reward: 0.7716 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.0109 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.0333 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.999

Q-Learning Mean Training Reward: 0.0014 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.99

Q-Learning Mean Training Reward: 0.0083 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.995

Q-Learning Mean Training Reward: 0.0258 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0005 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.99

Expected SARSA Mean Training Reward: 0.0008 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.995

Expected SARSA Mean Training Reward: 0.0027 | Mean Test Reward: 0.0 | gamma=1, alpha=0.1, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0004 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.99

Expected SARSA Mean Training Reward: 0.0005 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.995

C:\Users\nikla\Documents\Uni\MitStefRL\ReinforcementLearning\assignment\_2\code\_skeleton\ex1\_model\_free\_rl.py:85: RuntimeWarning: invalid value encountered in multiply

expected\_next\_q = np.sum(self.Q[next\_state] \* self.policy(next\_state, is\_training=False))

Expected SARSA Mean Training Reward: 0.0036 | Mean Test Reward: 0.0 | gamma=1, alpha=0.5, eps=1, eps\_decay=0.999

Expected SARSA Mean Training Reward: 0.0 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.99

Expected SARSA Mean Training Reward: 0.0028 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.995

Expected SARSA Mean Training Reward: 0.0026 | Mean Test Reward: 0.0 | gamma=1, alpha=0.9, eps=1, eps\_decay=0.999