In [1]:

```
import gym
import numpy as np
import matplotlib.pyplot as plt
import sys
```

Semi-Gradient Sarsa(0)

In [2]:

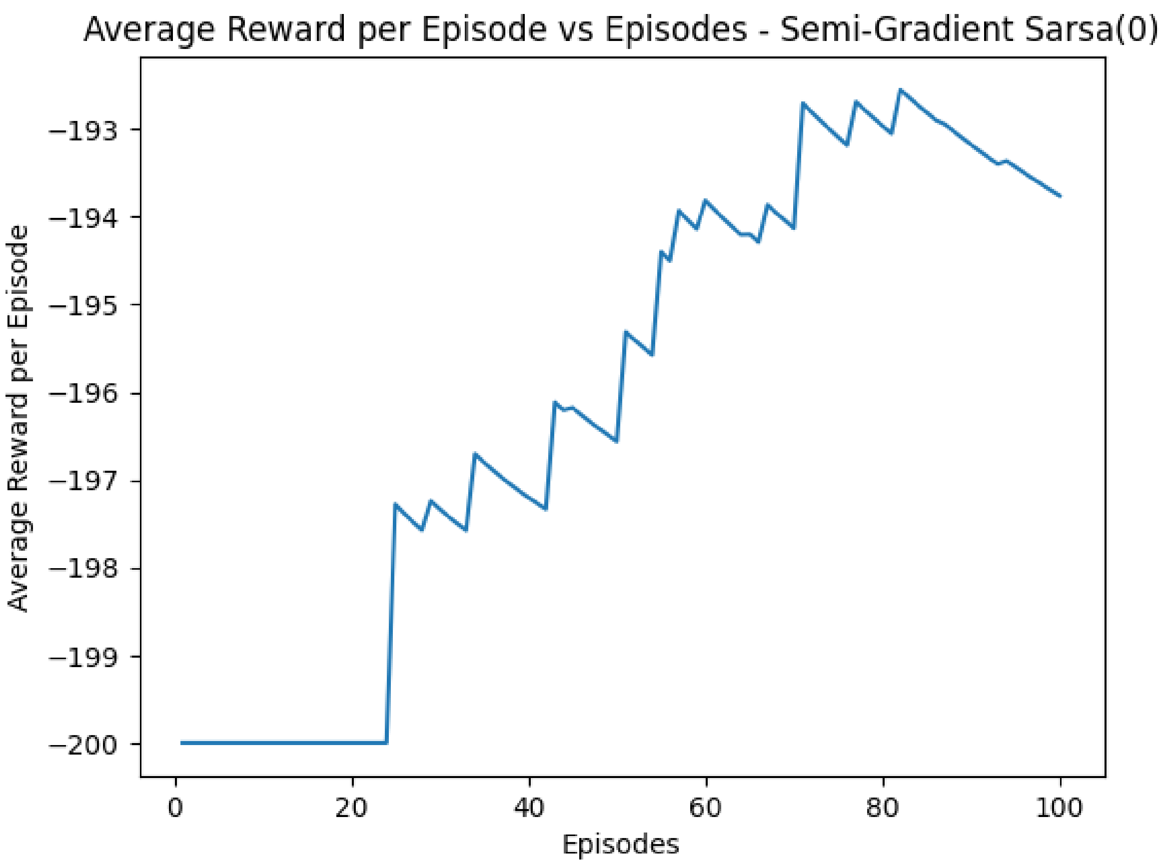
```
# Create the Mountain Car environment
env = gym.make('MountainCar-v0')

# Set up the action and observation space
action_space = env.action_space.n
observation_space = env.observation_space.shape[0]

# Define the weight vector for linear function approximation
weights_sarsa = np.random.rand(observation_space + 1)  # Add 1 for the action

# Define the hyperparameters
alpha = 0.1  # learning rate
gamma = 0.99  # discount factor
epsilon = 0.1  # exploration rate
episodes = 100  # number of episodes
steps_per_episode = 200  # maximum number of steps per episode

# Define the epsilon-greedy policy function
def epsilon_greedy(weights, state, epsilon):
    if np.random.rand() < epsilon:
        return np.random.randint(action_space)
    else:
        q_values = [np.dot(weights, get_features(state, a)) for a in range(action_
        return np.argmax(q_values)

# Define the feature extraction function
def get_features(state, action=None):
    if action is None:
        return np.append(state, 1)  # Append 1 for the action
    else:
        return np.append(state, action)

# Initialize lists to store rewards and steps
rewards_sarsa = []

# Implement the Semi-Gradient Sarsa(0) algorithm with linear function approximatio
for episode in range(episodes):
    state = env.reset()
    action_sarsa = epsilon_greedy(weights_sarsa, state, epsilon)

    episode_reward = 0

    for step in range(steps_per_episode):
        # Take action and observe the next state and reward
        next_state, reward, done, _ = env.step(action_sarsa)

        # Choose next action using epsilon-greedy policy for Sarsa(0)
        next_action_sarsa = epsilon_greedy(weights_sarsa, next_state, epsilon)

        # Compute the TD error for Sarsa(0)
        td_error = reward + gamma * np.dot(weights_sarsa, get_features(next_state,

        # Update the weight vector for Sarsa(0) using the gradient descent update r
        weights_sarsa += alpha * td_error * get_features(state, action_sarsa)
```

```
        # Update the state and action for Sarsa(0)
        state = next_state
        action_sarsa = next_action_sarsa

        episode_reward += reward

        if done:
            break

    # Store the total reward for the current episode
    rewards_sarsa.append(episode_reward)

    # Print the current state during training
    print("Episode:", episode + 1, "Steps:", step + 1, "State:", state, "Average R

# Test the learned policy
total_reward = 0
state = env.reset()

for step in range(steps_per_episode):
    action = epsilon_greedy(weights_sarsa, state, 0)  # Set exploration rate to 0 f
    state, reward, done, _ = env.step(action)
    total_reward += reward

    if done:
        break

print("Total reward:", total_reward)

# Plot the average reward per episode vs episodes
avg_rewards_sarsa = [np.mean(rewards_sarsa[max(0, i-99):i+1]) for i in range(len(re
plt.plot(range(1, episodes + 1), avg_rewards_sarsa)
plt.xlabel('Episodes')
plt.ylabel('Average Reward per Episode')
plt.title('Average Reward per Episode vs Episodes - Semi-Gradient Sarsa(0)')
plt.show()

# Close the environment
env.close()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarn
ing: `should_run_async` will not call `transform_cell` automatically in the futur
e. Please pass the result to `transformed_cell` argument and any exception that ha
ppen during thetransform in `preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
/usr/local/lib/python3.10/dist-packages/gym/core.py:317: DeprecationWarning: WARN:
Initializing wrapper in old step API which returns one bool instead of two. It is
recommended to set `new_step_api=True` to use new step API. This will be the defau
lt behaviour in future.
  deprecation(
/usr/local/lib/python3.10/dist-packages/gym/wrappers/step_api_compatibility.py:39:
DeprecationWarning: WARN: Initializing environment in old step API which returns o
ne bool instead of two. It is recommended to set `new_step_api=True` to use new st
ep API. This will be the default behaviour in future.
  deprecation(
```

```
Episode: 1 Steps: 200 State: [-8.4259486e-01 -2.6320363e-04] Average Reward: -200.
0
Episode: 2 Steps: 200 State: [-0.6806075   0.01010762] Average Reward: -200.0
Episode: 3 Steps: 200 State: [-1.0174215   0.00299919] Average Reward: -200.0
Episode: 4 Steps: 200 State: [-0.8358097   0.00302388] Average Reward: -200.0
Episode: 5 Steps: 200 State: [-0.9640592   0.00152048] Average Reward: -200.0
Episode: 6 Steps: 200 State: [-9.235712e-01  8.131184e-04] Average Reward: -200.0
Episode: 7 Steps: 200 State: [-0.72480965  0.00979286] Average Reward: -200.0
Episode: 8 Steps: 200 State: [-0.66587687  0.00788568] Average Reward: -200.0
Episode: 9 Steps: 200 State: [-0.6595023   0.00802438] Average Reward: -200.0
Episode: 10 Steps: 200 State: [-0.734279   -0.04016607] Average Reward: -200.0
Episode: 11 Steps: 200 State: [-0.8773557   0.00485843] Average Reward: -200.0
Episode: 12 Steps: 200 State: [-0.60650396 -0.03249706] Average Reward: -200.0
Episode: 13 Steps: 200 State: [-0.7255928   0.00459492] Average Reward: -200.0
Episode: 14 Steps: 200 State: [-0.6643349  -0.03597895] Average Reward: -200.0
Episode: 15 Steps: 200 State: [-0.7352178  0.0021055] Average Reward: -200.0
Episode: 16 Steps: 200 State: [-0.92205733 -0.02568693] Average Reward: -200.0
Episode: 17 Steps: 200 State: [-6.7989928e-01 -6.0048847e-06] Average Reward: -20
0.0
Episode: 18 Steps: 200 State: [-0.69207263  0.00095853] Average Reward: -200.0
Episode: 19 Steps: 200 State: [-0.8629312  -0.02998947] Average Reward: -200.0
Episode: 20 Steps: 200 State: [-0.7410086  0.0016352] Average Reward: -200.0
Episode: 21 Steps: 200 State: [-0.3933155  -0.02485566] Average Reward: -200.0
Episode: 22 Steps: 200 State: [-0.7146811  -0.04508664] Average Reward: -200.0
Episode: 23 Steps: 200 State: [-1.1465876   0.01498606] Average Reward: -200.0
Episode: 24 Steps: 200 State: [0.338816   0.00038226] Average Reward: -200.0
Episode: 25 Steps: 132 State: [0.51488835 0.03676482] Average Reward: -132.0
Episode: 26 Steps: 200 State: [-1.1149721  -0.02395261] Average Reward: -200.0
Episode: 27 Steps: 200 State: [-0.51093775 -0.03158752] Average Reward: -200.0
Episode: 28 Steps: 200 State: [-0.8242753  -0.03257569] Average Reward: -200.0
Episode: 29 Steps: 188 State: [0.51015043 0.03766095] Average Reward: -188.0
Episode: 30 Steps: 200 State: [-0.8875526   0.02936159] Average Reward: -200.0
Episode: 31 Steps: 200 State: [0.41605997 0.02835641] Average Reward: -200.0
Episode: 32 Steps: 200 State: [-1.1000296  0.0175474] Average Reward: -200.0
Episode: 33 Steps: 200 State: [-0.48996037 -0.02559725] Average Reward: -200.0
Episode: 34 Steps: 168 State: [0.50676763 0.02170159] Average Reward: -168.0
Episode: 35 Steps: 200 State: [-1.1907398  -0.01330119] Average Reward: -200.0
Episode: 36 Steps: 200 State: [-0.1283285   0.01236466] Average Reward: -200.0
Episode: 37 Steps: 200 State: [-1.0994567   0.01589999] Average Reward: -200.0
Episode: 38 Steps: 200 State: [-0.71396494  0.0063156 ] Average Reward: -200.0
Episode: 39 Steps: 200 State: [-0.81217724  0.00679169] Average Reward: -200.0
Episode: 40 Steps: 200 State: [0.31388873 0.02309423] Average Reward: -200.0
Episode: 41 Steps: 200 State: [-1.1655792   0.00962622] Average Reward: -200.0
Episode: 42 Steps: 200 State: [-0.00240222 -0.00829094] Average Reward: -200.0
Episode: 43 Steps: 145 State: [0.5076362  0.02036635] Average Reward: -145.0
Episode: 44 Steps: 200 State: [-0.46902502  0.0206257 ] Average Reward: -200.0
Episode: 45 Steps: 195 State: [0.52178204 0.03640092] Average Reward: -195.0
Episode: 46 Steps: 200 State: [0.1879756 0.042589 ] Average Reward: -200.0
Episode: 47 Steps: 200 State: [-0.004007    0.03010839] Average Reward: -200.0
Episode: 48 Steps: 200 State: [-1.1544678   0.01025899] Average Reward: -200.0
Episode: 49 Steps: 200 State: [-0.8802302   0.00091193] Average Reward: -200.0
Episode: 50 Steps: 200 State: [0.23687948 0.01489573] Average Reward: -200.0
Episode: 51 Steps: 133 State: [0.5044651  0.03174096] Average Reward: -133.0
Episode: 52 Steps: 200 State: [0.13448797 0.0427868 ] Average Reward: -200.0
Episode: 53 Steps: 200 State: [0.47189558 0.00875743] Average Reward: -200.0
Episode: 54 Steps: 200 State: [-0.54675424  0.02320059] Average Reward: -200.0
Episode: 55 Steps: 131 State: [0.5211556  0.03957938] Average Reward: -131.0
Episode: 56 Steps: 200 State: [-0.6469838  0.0180815] Average Reward: -200.0
Episode: 57 Steps: 162 State: [0.50427556 0.01491424] Average Reward: -162.0
Episode: 58 Steps: 200 State: [-0.33361968  0.02923066] Average Reward: -200.0
Episode: 59 Steps: 200 State: [-1.0820873   0.01736933] Average Reward: -200.0
Episode: 60 Steps: 175 State: [0.5289303  0.04546462] Average Reward: -175.0
Episode: 61 Steps: 200 State: [-0.4716298   0.04524658] Average Reward: -200.0
Episode: 62 Steps: 200 State: [-1.1969926  -0.04883662] Average Reward: -200.0
```

```
Episode: 63 Steps: 200 State: [-0.7507747  0.0038181] Average Reward: -200.0
Episode: 64 Steps: 200 State: [-0.88045454 -0.03203848] Average Reward: -200.0
Episode: 65 Steps: 194 State: [0.5230265  0.02667838] Average Reward: -194.0
Episode: 66 Steps: 200 State: [-0.3646996   0.05281581] Average Reward: -200.0
Episode: 67 Steps: 166 State: [0.5073271  0.01907358] Average Reward: -166.0
Episode: 68 Steps: 200 State: [-0.3944424   0.04453599] Average Reward: -200.0
Episode: 69 Steps: 200 State: [-1.18752     0.00500816] Average Reward: -200.0
Episode: 70 Steps: 200 State: [-0.47701612 -0.02751889] Average Reward: -200.0
Episode: 71 Steps: 93 State: [0.50371903 0.01831982] Average Reward: -93.0
Episode: 72 Steps: 200 State: [-0.23867162  0.04318878] Average Reward: -200.0
Episode: 73 Steps: 200 State: [-0.1609821  0.0418838] Average Reward: -200.0
Episode: 74 Steps: 200 State: [ 0.30889738 -0.00580547] Average Reward: -200.0
Episode: 75 Steps: 200 State: [-0.91103005 -0.02858634] Average Reward: -200.0
Episode: 76 Steps: 200 State: [0.26154855 0.03212797] Average Reward: -200.0
Episode: 77 Steps: 155 State: [0.51786655 0.01841644] Average Reward: -155.0
Episode: 78 Steps: 200 State: [-1.1987581  0.0012419] Average Reward: -200.0
Episode: 79 Steps: 200 State: [-1.0555344  -0.02109323] Average Reward: -200.0
Episode: 80 Steps: 200 State: [-0.43440753 -0.02904705] Average Reward: -200.0
Episode: 81 Steps: 200 State: [ 0.03229128 -0.00863347] Average Reward: -200.0
Episode: 82 Steps: 152 State: [0.5035765  0.01753113] Average Reward: -152.0
Episode: 83 Steps: 200 State: [-1.0811756  -0.00946005] Average Reward: -200.0
Episode: 84 Steps: 200 State: [-0.3751748  -0.01593433] Average Reward: -200.0
Episode: 85 Steps: 200 State: [-0.22679679 -0.02833044] Average Reward: -200.0
Episode: 86 Steps: 200 State: [-0.7837595   0.00344506] Average Reward: -200.0
Episode: 87 Steps: 197 State: [0.51475066 0.02153336] Average Reward: -197.0
Episode: 88 Steps: 200 State: [-0.7023657  0.0136055] Average Reward: -200.0
Episode: 89 Steps: 200 State: [-0.07260948 -0.00314452] Average Reward: -200.0
Episode: 90 Steps: 200 State: [-0.48694718  0.03973457] Average Reward: -200.0
Episode: 91 Steps: 200 State: [-1.0918249  0.0147572] Average Reward: -200.0
Episode: 92 Steps: 200 State: [-0.654301   -0.04968017] Average Reward: -200.0
Episode: 93 Steps: 200 State: [-0.597553  -0.0441918] Average Reward: -200.0
Episode: 94 Steps: 190 State: [0.5273151  0.03047346] Average Reward: -190.0
Episode: 95 Steps: 200 State: [-0.40278888  0.03766304] Average Reward: -200.0
Episode: 96 Steps: 200 State: [-1.1587042   0.01092957] Average Reward: -200.0
Episode: 97 Steps: 200 State: [-0.54112226 -0.03886005] Average Reward: -200.0
Episode: 98 Steps: 200 State: [0.4761966  0.03388999] Average Reward: -200.0
Episode: 99 Steps: 200 State: [-1.1544678   0.01025899] Average Reward: -200.0
Episode: 100 Steps: 200 State: [-0.11698714 -0.01035212] Average Reward: -200.0
Total reward: -200.0
```



Semi-Gradient TD(λ)

In [3]:
```
# Set up the action and observation space
action_space = env.action_space.n
observation_space = env.observation_space.shape[0]

# Define the weight vector for linear function approximation
weights_td = np.random.rand(observation_space + 1)  # Add 1 for the action

# Define the hyperparameters
alpha = 0.1  # learning rate
gamma = 0.99  # discount factor
epsilon = 0.1  # exploration rate
episodes = 100  # number of episodes
steps_per_episode = 200  # maximum number of steps per episode
lambda_ = 0.5  # eligibility trace parameter

# Define the epsilon-greedy policy function
def epsilon_greedy(weights, state, epsilon):
    if np.random.rand() < epsilon:
        return np.random.randint(action_space)
    else:
        q_values = [np.dot(weights, get_features(state, a)) for a in range(action_
        return np.argmax(q_values)

# Define the feature extraction function
def get_features(state, action=None):
    if action is None:
        return np.append(state, 1)  # Append 1 for the action
    else:
        return np.append(state, action)

# Initialize lists to store rewards and steps
rewards_td = []
```

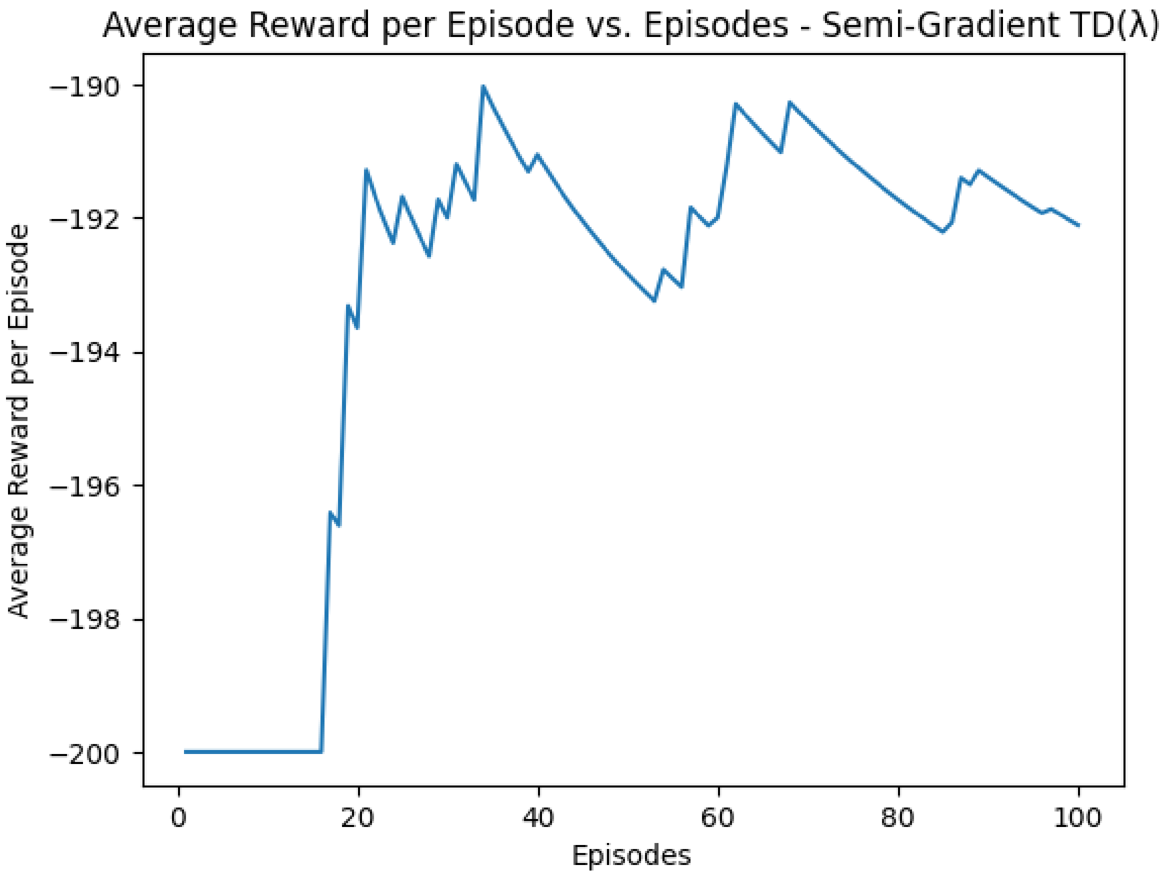
```
# Initialize lists to store average rewards and steps
avg_rewards_td = []
episode_steps = []

# Implement the Semi-Gradient TD(λ) algorithm with linear function approximation
for episode in range(episodes):
    state = env.reset()
    action_td = epsilon_greedy(weights_td, state, epsilon)

    eligibility_trace_td = np.zeros_like(weights_td)  # Initialize eligibility tra

    episode_reward = 0  # Track the total reward in each episode

    for step in range(steps_per_episode):
        # Take action and observe the next state and reward
        next_state, reward, done, _ = env.step(action_td)

        # Choose next action using epsilon-greedy policy for TD(λ)
        next_action_td = epsilon_greedy(weights_td, next_state, epsilon)

        # Compute the TD error for TD(λ)
        td_error_td = reward + gamma * np.dot(weights_td, get_features(next_state)

        # Update the eligibility trace for TD(λ)
        eligibility_trace_td = gamma * lambda_ * eligibility_trace_td + get_featur

        # Update the weight vector for TD(λ) using the gradient descent update rule
        weights_td += alpha * td_error_td * eligibility_trace_td

        # Update the state and action for TD(λ)
        state = next_state
        action_td = next_action_td

        episode_reward += reward  # Accumulate the reward

        if done:
            break

    # Store the total reward for the current episode
    rewards_td.append(episode_reward)

    # Calculate the average reward per episode
    if len(rewards_td) >= 100:
        avg_reward = sum(rewards_td[-100:]) / 100
    else:
        avg_reward = sum(rewards_td) / len(rewards_td)

    avg_rewards_td.append(avg_reward)

    # Store the number of steps in the episode
    episode_steps.append(step + 1)

    # Print the current state during training
    print("Episode:", episode + 1, "Steps:", step + 1, "State:", state, "Average R

# Test the learned policy
total_reward = 0
state = env.reset()

for step in range(steps_per_episode):
    action = epsilon_greedy(weights_td, state, 0)  # Set exploration rate to 0 for
    state, reward, done, _ = env.step(action)
    total_reward += reward
```

```python
    if done:
        break

print("Total reward:", total_reward)

# Plot the average reward per episode vs. episodes
plt.plot(range(1, episodes + 1), avg_rewards_td)
plt.xlabel('Episodes')
plt.ylabel('Average Reward per Episode')
plt.title('Average Reward per Episode vs. Episodes - Semi-Gradient TD(λ)')
plt.show()

# Close the environment
env.close()
```

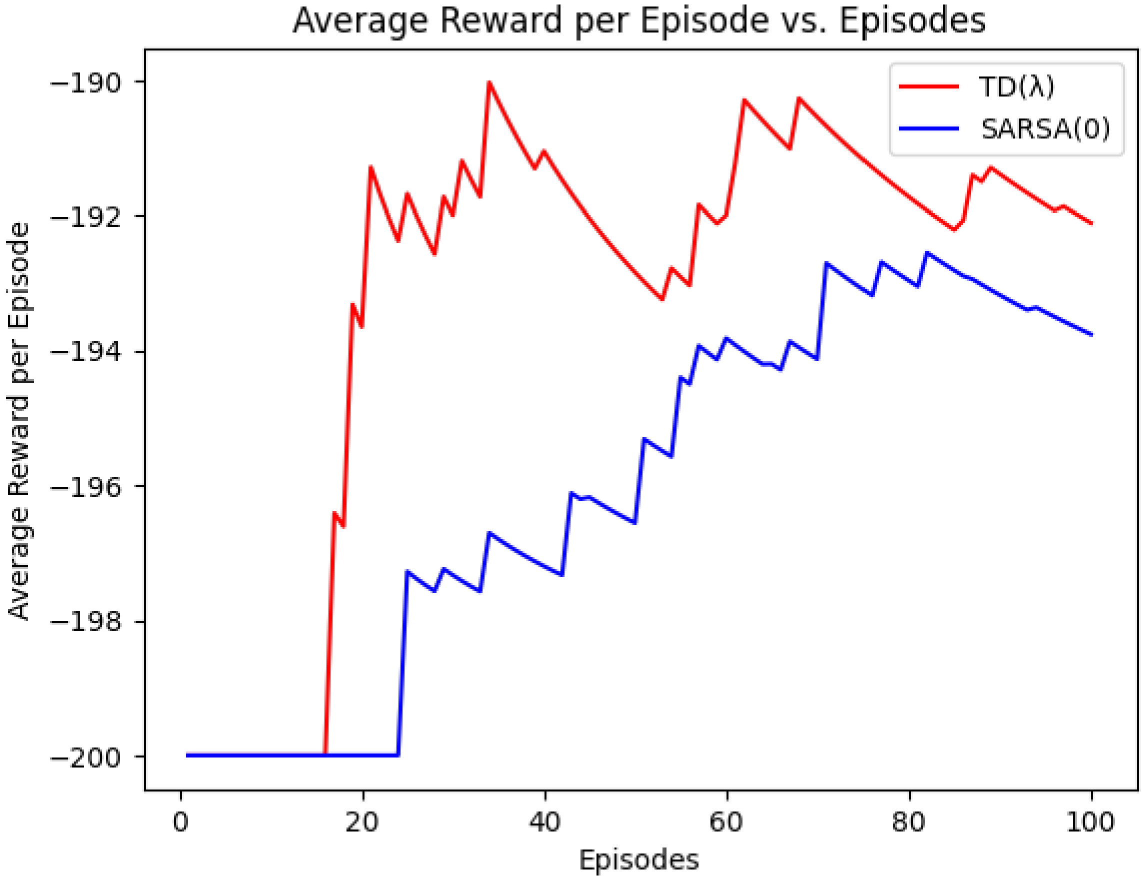
```
Episode: 1 Steps: 200 State: [-0.92463166 -0.00286642] Average Reward: -200.0
Episode: 2 Steps: 200 State: [-0.74627227  0.00914829] Average Reward: -200.0
Episode: 3 Steps: 200 State: [-0.7114813   0.00189998] Average Reward: -200.0
Episode: 4 Steps: 200 State: [0.2611268  0.01088297] Average Reward: -200.0
Episode: 5 Steps: 200 State: [-0.03951322  0.00756995] Average Reward: -200.0
Episode: 6 Steps: 200 State: [-1.1123488   0.01545782] Average Reward: -200.0
Episode: 7 Steps: 200 State: [-0.82883275  0.00811999] Average Reward: -200.0
Episode: 8 Steps: 200 State: [-0.7341943   0.01214671] Average Reward: -200.0
Episode: 9 Steps: 200 State: [-0.45200142 -0.05459183] Average Reward: -200.0
Episode: 10 Steps: 200 State: [-0.9599345  -0.05304709] Average Reward: -200.0
Episode: 11 Steps: 200 State: [-0.64362967 -0.03579873] Average Reward: -200.0
Episode: 12 Steps: 200 State: [-1.182204   -0.03403521] Average Reward: -200.0
Episode: 13 Steps: 200 State: [0.43392906 0.00911236] Average Reward: -200.0
Episode: 14 Steps: 200 State: [-0.37315223  0.00163654] Average Reward: -200.0
Episode: 15 Steps: 200 State: [-4.1128936e-01 -2.5509340e-05] Average Reward: -20
0.0
Episode: 16 Steps: 200 State: [-0.4247219   0.00428661] Average Reward: -200.0
Episode: 17 Steps: 139 State: [0.52490693 0.03002022] Average Reward: -196.4117647
0588235
Episode: 18 Steps: 200 State: [-0.48762628  0.02103901] Average Reward: -196.61111
111111111
Episode: 19 Steps: 134 State: [0.5068088 0.0310818] Average Reward: -193.315789473
68422
Episode: 20 Steps: 200 State: [-0.40814114  0.00262107] Average Reward: -193.65
Episode: 21 Steps: 144 State: [0.5055911  0.02501375] Average Reward: -191.2857142
8571428
Episode: 22 Steps: 200 State: [-0.4345183  -0.00149484] Average Reward: -191.68181
81818182
Episode: 23 Steps: 200 State: [-0.45787796  0.0464466 ] Average Reward: -192.04347
826086956
Episode: 24 Steps: 200 State: [-0.9982282  -0.00192826] Average Reward: -192.375
Episode: 25 Steps: 175 State: [0.53817767 0.04194646] Average Reward: -191.68
Episode: 26 Steps: 200 State: [-0.51387686  0.03065731] Average Reward: -192.0
Episode: 27 Steps: 200 State: [-0.73576796 -0.0350589 ] Average Reward: -192.29629
62962963
Episode: 28 Steps: 200 State: [-0.83635    -0.03243763] Average Reward: -192.57142
857142858
Episode: 29 Steps: 168 State: [0.502831   0.04043475] Average Reward: -191.7241379
3103448
Episode: 30 Steps: 200 State: [-0.54176843 -0.02201995] Average Reward: -192.0
Episode: 31 Steps: 167 State: [0.5296794  0.03839257] Average Reward: -191.1935483
8709677
Episode: 32 Steps: 200 State: [-0.7962345   0.03228319] Average Reward: -191.46875
Episode: 33 Steps: 200 State: [-0.7004834   0.00938237] Average Reward: -191.72727
272727272
Episode: 34 Steps: 134 State: [0.5144252  0.02964413] Average Reward: -190.0294117
6470588
Episode: 35 Steps: 200 State: [-1.0942681  -0.06088854] Average Reward: -190.31428
571428572
Episode: 36 Steps: 200 State: [-0.73905456  0.00362168] Average Reward: -190.58333
333333334
Episode: 37 Steps: 200 State: [-1.2  0. ] Average Reward: -190.83783783783784
Episode: 38 Steps: 200 State: [-0.01024785 -0.03653956] Average Reward: -191.07894
736842104
Episode: 39 Steps: 200 State: [-0.6647864  -0.04176928] Average Reward: -191.30769
230769232
Episode: 40 Steps: 181 State: [0.50571036 0.03368236] Average Reward: -191.05
Episode: 41 Steps: 200 State: [-0.37446812  0.00410289] Average Reward: -191.26829
268292684
Episode: 42 Steps: 200 State: [-0.616329   0.0369092] Average Reward: -191.4761904
7619048
Episode: 43 Steps: 200 State: [-0.23150732  0.00171207] Average Reward: -191.67441
860465115
Episode: 44 Steps: 200 State: [-0.80081314  0.03307518] Average Reward: -191.86363
```

```
636363637
Episode: 45 Steps: 200 State: [-1.0291226   0.02147224] Average Reward: -192.04444
444444445
Episode: 46 Steps: 200 State: [-0.46976477 -0.04257053] Average Reward: -192.21739
13043478
Episode: 47 Steps: 200 State: [-0.24228156 -0.01111466] Average Reward: -192.38297
872340425
Episode: 48 Steps: 200 State: [-0.9382387  -0.02536759] Average Reward: -192.54166
666666666
Episode: 49 Steps: 200 State: [-0.3388645  -0.00561276] Average Reward: -192.69387
755102042
Episode: 50 Steps: 200 State: [-0.85473037 -0.02868152] Average Reward: -192.84
Episode: 51 Steps: 200 State: [-0.5842732  -0.03126747] Average Reward: -192.98039
215686273
Episode: 52 Steps: 200 State: [0.4167754 0.0294074] Average Reward: -193.115384615
3846
Episode: 53 Steps: 200 State: [-0.46721974 -0.01517898] Average Reward: -193.24528
301886792
Episode: 54 Steps: 168 State: [0.5171817  0.03260274] Average Reward: -192.7777777
7777777
Episode: 55 Steps: 200 State: [-1.1450701  -0.00434856] Average Reward: -192.90909
09090909
Episode: 56 Steps: 200 State: [-0.20742415  0.0046476 ] Average Reward: -193.03571
428571428
Episode: 57 Steps: 125 State: [0.5355561  0.04260675] Average Reward: -191.8421052
631579
Episode: 58 Steps: 200 State: [-0.5937683   0.04714421] Average Reward: -191.98275
862068965
Episode: 59 Steps: 200 State: [-1.0440333   0.02191837] Average Reward: -192.11864
406779662
Episode: 60 Steps: 185 State: [0.53325003 0.04860929] Average Reward: -192.0
Episode: 61 Steps: 145 State: [0.5021369  0.01725516] Average Reward: -191.2295081
967213
Episode: 62 Steps: 133 State: [0.51274043 0.02787882] Average Reward: -190.2903225
8064515
Episode: 63 Steps: 200 State: [-0.36717895  0.00085132] Average Reward: -190.44444
444444446
Episode: 64 Steps: 200 State: [-0.8998434  0.0286513] Average Reward: -190.59375
Episode: 65 Steps: 200 State: [-1.1388116   0.01264561] Average Reward: -190.73846
153846154
Episode: 66 Steps: 200 State: [-0.66062456  0.0120718 ] Average Reward: -190.87878
787878788
Episode: 67 Steps: 200 State: [-0.22668093  0.03656664] Average Reward: -191.01492
537313433
Episode: 68 Steps: 140 State: [0.5008336 0.0265551] Average Reward: -190.264705882
35293
Episode: 69 Steps: 200 State: [-0.8544397   0.00975345] Average Reward: -190.40579
710144928
Episode: 70 Steps: 200 State: [-0.94659704 -0.02338161] Average Reward: -190.54285
714285714
Episode: 71 Steps: 200 State: [-0.7262785   0.00801378] Average Reward: -190.67605
633802816
Episode: 72 Steps: 200 State: [-0.20947197  0.02346583] Average Reward: -190.80555
555555554
Episode: 73 Steps: 200 State: [-0.70176494  0.03506107] Average Reward: -190.93150
684931507
Episode: 74 Steps: 200 State: [-0.5608737  -0.06749154] Average Reward: -191.05405
405405406
Episode: 75 Steps: 200 State: [-1.0653682  -0.03941942] Average Reward: -191.17333
333333335
Episode: 76 Steps: 200 State: [-0.7973031   0.00229879] Average Reward: -191.28947
368421052
Episode: 77 Steps: 200 State: [-0.25568175 -0.05383581] Average Reward: -191.40259
74025974
```

```
Episode: 78 Steps: 200 State: [-7.1491307e-01 -7.0210337e-04] Average Reward: -19
1.51282051282053
Episode: 79 Steps: 200 State: [-0.93798643 -0.02775243] Average Reward: -191.62025
316455697
Episode: 80 Steps: 200 State: [-1.127172   -0.04335022] Average Reward: -191.725
Episode: 81 Steps: 200 State: [-0.9738132  -0.03864793] Average Reward: -191.82716
049382717
Episode: 82 Steps: 200 State: [-0.86410034 -0.03052386] Average Reward: -191.92682
92682927
Episode: 83 Steps: 200 State: [-1.1925281   0.00374203] Average Reward: -192.02409
638554218
Episode: 84 Steps: 200 State: [-0.7424698   0.00447596] Average Reward: -192.11904
761904762
Episode: 85 Steps: 200 State: [-0.811014  -0.0345521] Average Reward: -192.2117647
0588236
Episode: 86 Steps: 180 State: [0.5126368  0.01680115] Average Reward: -192.0697674
4186048
Episode: 87 Steps: 134 State: [0.5206378  0.03228648] Average Reward: -191.4022988
505747
Episode: 88 Steps: 200 State: [-0.65177345  0.03238655] Average Reward: -191.5
Episode: 89 Steps: 173 State: [0.5159471  0.03352075] Average Reward: -191.2921348
314607
Episode: 90 Steps: 200 State: [-0.37552056 -0.00290903] Average Reward: -191.38888
888888889
Episode: 91 Steps: 200 State: [-0.9386321   0.02823923] Average Reward: -191.48351
64835165
Episode: 92 Steps: 200 State: [-0.8567966   0.01189915] Average Reward: -191.57608
695652175
Episode: 93 Steps: 200 State: [-1.018026    0.02285265] Average Reward: -191.66666
666666666
Episode: 94 Steps: 200 State: [-0.7698058  -0.03107499] Average Reward: -191.75531
914893617
Episode: 95 Steps: 200 State: [-1.1812301  0.0062899] Average Reward: -191.8421052
631579
Episode: 96 Steps: 200 State: [-0.49724802 -0.02242005] Average Reward: -191.92708
333333334
Episode: 97 Steps: 186 State: [0.5136781  0.03999509] Average Reward: -191.8659793
814433
Episode: 98 Steps: 200 State: [-0.99768156  0.02334484] Average Reward: -191.94897
959183675
Episode: 99 Steps: 200 State: [-0.7664023   0.00768417] Average Reward: -192.03030
303030303
Episode: 100 Steps: 200 State: [-1.1962702  0.0024879] Average Reward: -192.11
Total reward: -200.0
```




```
In [4]: # Comparison
        # Plot the average reward per episode vs. episodes for TD(λ)
        plt.plot(range(1, episodes + 1), avg_rewards_td, color='red', label='TD(λ)')

        # Plot the average reward per episode vs. episodes for SARSA(0)
        avg_rewards_sarsa = [np.mean(rewards_sarsa[max(0, i-99):i+1]) for i in range(len(r
        plt.plot(range(1, episodes + 1), avg_rewards_sarsa, color='blue', label='SARSA(0)'

        plt.xlabel('Episodes')
        plt.ylabel('Average Reward per Episode')
        plt.title('Average Reward per Episode vs. Episodes')
        plt.legend()
        plt.show()
```



Average Reward per Episode vs. Episodes
TD(λ)
SARSA(0)
Average Reward per Episode
−190
−192
−194
−196
−198
−200
0
20
40
60
80
100
Episodes