1 Efficient Routing MDP

(a) rs $\in\{-5, -0.5, 0, 2\}$

rs = −5.

1.If we move to red square 7, game will be terminated with the reward of -10.

2. If we choose the strategy with 2 or more steps - reward will be worth.

3. If we choose the strategy to reach green square 33, it will be at least 5 steps and the best result will be -20.

The best policy is first one with reward -10.

Optimal policy is unique and it isn`t depend on discount factor.

rs = −0.5.

1.If we move to red square 7, game will be terminated with the reward of -5.5.

2. If we choose the strategy with 2 or more steps with the last one red - reward will be worth.

3. If we choose the strategy to reach green square 33 with 5 steps, the result will be +2.5.

4. If we choose the strategy to reach green square 33 with more than 5 steps, the result will be worth then +2.5.

The best policy is third one with reward +2.5. The way is: 2 > 9 > 16 > 21 > 28(26) > 33

Optimal policy isn`t unique and it isn`t depend on discount factor.

rs = 0.

Because of rs=0, the best reward will be 5 when we get 33 green square.

One more is discount factor, which means the longer path will be, the smaller rewerd wi will get.

So the best policy is with reward +5. The way is: 2 > 9 > 16 > 21 > 28(26) > 33

Optimal policy isn`t unique and it depends on discount factor.

rs = 2.

Here because of rs>0 we should go the longest path

Optimal policy is unique and it depends on discount factor.

b)

The shortest path to the green square will be for the values **rs** $\in$**{−0.5, 0}**

$$v_\pi(2) = r_s(1 + \gamma + \gamma^2 + \gamma^3 + \gamma^4) + r_g\gamma^5 = -0.5(1 + 0.9 + 0.9^2 + 0.9^3 + 0.9^4) + 5 \cdot 0.9^5 = 0.9049$$

$$v_\pi(13) = -5$$

$$v_\pi(21) = -0.5(1 + 0.9) + 5 \cdot 0.9^2 = 3.1$$

$$v_\pi(32) = r_s + r_r\gamma = -0.5 - 5 \cdot 0.9 = -5$$

c)

re = −5.

We should find the path to terminate the game faster. The path is 2 times go right to red square 14.

re = −0.5.

Optimal policy is the shortest path, move down one time and than go right to green square.

re = 0.

Optimal policy is the shortest path, the same is upper.

re = 2.

In this case we will search for the longest path. If we move from 2 down to 4, then right. This path will have 2 more moves, and it's the longest path possible. In this case we depend on γ

d)

For $\gamma = 0.9$ it will be approximately re > 0.0630

e)

- Using only efficient actions - states {5,17}.
- Using only inefficient actions - only state 33

f)

$$
\begin{aligned}
(v_\pi)_{\text{new}}(s) &= \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k \left(R_{t+k+1} + c \right) \middle| S_t = s \right] \\
&= \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} + c \sum_{k=0}^{\infty} \gamma^k \middle| S_t = s \right] \\
&= \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \middle| S_t = s \right] + c \sum_{k=0}^{\infty} \gamma^k
\end{aligned}
$$

g)

- avoiding stops or slow-downs
- moving at the speed, that is optimal for engine productivity

So, for the most sustainable route I would propose the rules:

- Avoid cities.

The biggest city - the largest is negative reward for the route. City has a lot of turns, stops, speed limit zones. All of these would require us to change the speed frequently, which will cause huge fuel consumption.

- Negative reward for the turns and changes of direction. Usually, we slow down before the turn. The larger the angle of the turn - the bigger penalty.
- Highways are the best for sustainable driving - give them high positive reward.
- Penalize traffic jam situations. In the modern maps we can receive the information about traffic jams in a real time.