DYNAMIC PROGRAMMING EXERCISES

QUESTIONS FROM THE ONLINE FORM

FORM: HTTPS://FORMS.GLE/QHJMMNLWMHCZZIWK8

Sample-Based Coursera Course:

- Week 4 assignment due date is April 25th (Finals week)
- We will not cover this week of material in the course and the assignment is not counted

QUESTIONS FROM THE ONLINE FORM

FORM: HTTPS://FORMS.GLE/QHJMMNLWMHCZZIWK8

What does deterministic mean?

Deterministic means the probability distribution places all mass on one element:

Deterministic policy:

$$\pi(s) = a_1,\ \pi(a_1 \mid s) = \Pr(A_t = a_1 \mid S_t = s) = 1$$

Deterministic transition and reward:

$$p(s_2,1 \mid s_1, a) = 1 \text{ -> } \Pr(S_{t+1} = s_2, R_{t+1} = 1 \mid S_t = s, A_t = a) = 1$$

Deterministic transition:

$$p(s_1, a, s_2) = 1 \text{ -> } \Pr(S_{t+1} = s_2 \mid S_t = s, A_t = a) = 1$$

Deterministic reward

$$\Pr(R_{t+1} = r \mid S_t = s, A_t = a, S_{t+1} = s') = 1 \text{ -> } r(s, a, s') = r$$

QUESTIONS FROM THE ONLINE FORM

FORM: HTTPS://FORMS.GLE/QHJMMNLWMHCZZIWK8

What does deterministic mean?

When we say an MDP is deterministic, this means that all state transitions and reward transitions are deterministic.

If we say the transition dynamics are deterministic, this applies to all state-action pairs

If the reward function is deterministic, this could apply to all state-action pairs or state-action-next-state triples. The description of the rewards would indicate which case.

The agent receives zero reward unless it enters the goal state, then it gets a reward of 1

The reward is deterministic on s, a, s'

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $v_\pi(s) =$ <use q_π>
2. $q_\pi(s, a) =$ <use v_π>

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $v_\pi(s) = \sum_a \pi(a \mid s) q_\pi(s, a)$
2. $q_\pi(s, a) = r(s, a) + \gamma \sum_{s'} p(s, a, s') v_\pi(s')$

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $v_*(s) =$ <use q_* >
2. $q_*(s, a) =$ <use v_*>

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $v_*(s) = \sum_a \pi_*(a \mid s) q_*(s, a) = \max_a q_*(s, a)$

2. $q_*(s, a) = r(s, a) + \gamma \sum_{s'} p(s, a, s') v_*(s')$

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $\pi_*(s) \in$ <use v_*>
2. $\pi_*(s) \in$ <use q_*>

SWITCHING BETWEEN v AND q

WRITE AN EXPRESSION FOR THE FOLLOWING

1. $\pi_*(s) \in \arg\max_a r(s,a) + \gamma \sum_{s'} p(s,a,s')v_*(s')$
2. $\pi_*(s) \in q_*(s,a)$

ITERATIVE UPDATES

USING q ESTIMATES

Consider the update:

$$v_i^{k+1} = \sum_a \pi(a \mid s_i) \left(r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j^k \right)$$

How can we modify this equation to estimate q_π? Let $q^k \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{A}|}$, i.e,. $q_{i,a}^k \approx q_\pi(s_i, a)$

ITERATIVE UPDATES

USING q ESTIMATES

Consider the update:

$$v_i^{k+1} = \sum_a \pi(a \mid s_i) \left(r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j^k \right)$$

How can we modify this equation to estimate q_π? Let $q^k \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{A}|}$, i.e,. $q_{i,a}^k \approx q_\pi(s_i, a)$

$$q_{i,a}^{k+1} = r(s_i, a) + \gamma \sum_j p(s_i, a, s_j) \sum_{a'} \pi(a' \mid s_j) q_{j,a'}^k$$

ITERATIVE UPDATES

USING q ESTIMATES

Consider the update:

$$v_i^{k+1} = \sum_a \pi(a \mid s_i) \left(r(s, a) + \gamma \sum_j p(s_i, a, s_j) v_j^k \right)$$

How can we modify this equation to use the four arguments p, i.e., use $p(s', r \mid s, a)$ in the update?

ITERATIVE UPDATES

USING q ESTIMATES

Consider the update:

$$v_i^{k+1} = \sum_a \pi(a \mid s_i) \left(r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j^k \right)$$

How can we modify this equation to use the four arguments p, i.e., use $p(s', r \mid s, a)$ in the update?

$$v_i^{k+1} = \sum_a \pi(a \mid s_i) \sum_j \sum_r p(s_j, r \mid s_i, a) \left(r + \gamma v_j^k \right)$$

ITERATIVE UPDATES

USING q IN VALUE ITERATION

Consider the update:

$$v_i^{k+1} = \max_a r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j^k$$

How can we modify this equation to estimate q_*?

ITERATIVE UPDATES

USING q IN VALUE ITERATION

Consider the update:

$$v_i^{k+1} = \max_a r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j^k$$

How can we modify this equation to estimate q_*?

$$q_{i,a}^{k+1} = r(s,a) + \gamma \sum_j p(s_i, a, s_j) \max_{a'} q_{j,a'}^k$$

NEXT CLASS

WHAT YOU SHOULD DO

1. Programming assignment due tonight night

Friday: Midterm review. Bring questions you want answered