COMPSCI 687 Pop Quiz 4 — Solutions

Grading instructions. Each question is either correct or incorrect; there is no partial credit. Answers to Questions 1–4 must be perfect down to the symbol, except for the accepted variants listed below. For Questions 1 and 2, both the input/output sets and the function values are required. Quantifiers may be included or omitted in Questions 1–4. In definitions, either $=$ or $:=$ is accepted. Periods at the end of answers are optional. The typeface of $\mathbf{E}$ and $\mathbb{R}$ is not graded (e.g., $\mathbb{E}$, $\mathbf{E}$, or E is accepted), but square brackets and $;\pi$ inside each expectation are required. For Questions 5 and 6, equivalent mathematical statements or wording are accepted; the universal quantifiers must be stated in symbols or words.

1. Give the definition of v^π. Your answer should be two lines. The first specifies the input and output sets (e.g., $f : \mathcal{X} \to \mathcal{Y}$) and the second should specify the function values (e.g., $f(x) = \Pr(X = x)$). You may omit the quantifiers (e.g., $\forall x \in \mathcal{X}$).

   Required first line:

   $$v^\pi : \mathcal{S} \to \mathbb{R}.$$

   Accepted second lines: Any one of the following four forms is sufficient:

   $$v^\pi(s) = \mathbf{E}[G_t \mid S_t = s; \pi],$$
   $$v^\pi(s) = \mathbf{E}\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s; \pi\right],$$
   $$v^\pi(s) = \mathbf{E}[G \mid S_0 = s; \pi],$$
   $$v^\pi(s) = \mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \,\middle|\, S_0 = s; \pi\right].$$

   Any summation index may be renamed consistently without conflicting with a free index. All four forms must include $;\pi$ inside the expectation. The return and conditioning time must match: G_t goes with $S_t = s$, whereas G goes with $S_0 = s$.

2. Give the definition of q^π. Your answer should be two lines. The first specifies the input and output sets (e.g., $f : \mathcal{X} \to \mathcal{Y}$) and the second should specify the function values (e.g., $f(x) = \Pr(X = x)$). You may omit the quantifiers (e.g., $\forall x \in \mathcal{X}$).

   Required first line:

   $$q^\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}.$$

   Accepted second lines: Any one of the following four forms is sufficient:

   $$q^\pi(s, a) = \mathbf{E}[G_t \mid S_t = s, A_t = a; \pi],$$
   $$q^\pi(s, a) = \mathbf{E}\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, A_t = a; \pi\right],$$
   $$q^\pi(s, a) = \mathbf{E}[G \mid S_0 = s, A_0 = a; \pi],$$
   $$q^\pi(s, a) = \mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \,\middle|\, S_0 = s, A_0 = a; \pi\right].$$

   Summation indices may be renamed as in Question 1. All four forms must include $;\pi$ inside the expectation. Both the state and action conditions are required, and their time indices must match the return. The two conditions may be written in either order.

COMPSCI 687 Pop Quiz 4 — Solutions (continued)

3. Give the Bellman equation for v^π as a recursive expression for v^π using R, p, π, and γ. Start with $v^\pi(s) =$. You may omit the quantifier and the sets being summed over (e.g., writing $\sum_{a'}$ instead of $\sum_{a' \in \mathcal{A}}$).

$$v^\pi(s) = \sum_{a \in \mathcal{A}} \pi(s,a) \sum_{s' \in \mathcal{S}} p(s,a,s') \left(R(s,a) + \gamma v^\pi(s') \right),$$
$$v^\pi(s) = \sum_{a \in \mathcal{A}} \pi(s,a) \left(R(s,a) + \gamma \sum_{s' \in \mathcal{S}} p(s,a,s') v^\pi(s') \right).$$

For Questions 3 and 4, either displayed form is accepted. The sets being summed over, $\mathcal{A}$ and $\mathcal{S}$, may be omitted, and different variable names may be used for a and s'. The γ term may be placed in other positions as long as it does not change the value (e.g., in the second form it can be placed immediately before v^π).

4. Give the Bellman optimality equation for v^* as a recursive expression for v^* using R, p, and γ. Start with $v^*(s) =$. You may omit the quantifier and the sets being summed over (e.g., writing $\sum_{a'}$ instead of $\sum_{a' \in \mathcal{A}}$).

$$v^*(s) = \max_{a \in \mathcal{A}} \left(R(s,a) + \gamma \sum_{s' \in \mathcal{S}} p(s,a,s') v^*(s') \right),$$
$$v^*(s) = \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} p(s,a,s') \left(R(s,a) + \gamma v^*(s') \right).$$

The big parentheses in the first line are optional. Writing $\max_a$ instead of $\max_{a \in \mathcal{A}}$ is also accepted. In the first line, the γ term may be placed in other equivalent positions.

5. Give the definition of the preorder on policies that we defined in class. That is, $\pi' \geq \pi$ if and only if:

$$\forall s \in \mathcal{S}, \quad v^{\pi'}(s) \geq v^\pi(s).$$

Accepted wording: This answer need not be expressed in the same way, but the concept must be exactly and precisely conveyed. Students may write answers like: "For all states s, $v^{\pi'}(s) \geq v^\pi(s)$." The answer should be clear that this applies to every state and shouldn't require strict improvement ($\geq$ not $>$).

6. The definition of an optimal policy in terms of the preorder on policies defined by $\geq$ is:

A policy π^* is optimal if and only if

$$\forall \pi \in \Pi, \quad \pi^* \geq \pi.$$

Accepted wording: As with the previous question, different expressions of the exact same concept will be accepted, as long as they are clear and precise. The J-based definition will not be accepted.