MAIR Exam, part “Markov models for multi-agent learning” (35 min.)

- No open book. No electronic devices. Name: …………. ……….. Number: …………..
- All answers on this sheet. (Answers to multiple choice items can be encircled.)
- For every multiple choice item one answer is considered correct. If you think there are multiple correct answers, then choose the best one.
- Answer to open questions must be written in boxes. Do not hesitate to draft an answer on scratch paper first.

1. What is the meaning of “sample path”?
   (a) An example of an optimal policy.
   (b) A realisation of a probabilistic process.
   (c) A trial run in the process of learning a policy.
   (d) The correct answer is not listed here.

   Answer. A realisation of a probabilistic process.

2. i) Policy evaluation.
   ii) Policy iteration.

3. Consider the following Markov decision problem. Suppose a discount factor $\gamma = 0.5$.

   $$
   \begin{align*}
   s (P) : & \begin{cases}
   s (r = 6, p = 0.4) ; a (r = 7, p = 0.6), \\
   a (P) : & s(r = 3, p = 0.3) ; b (r = 0, p = 0.7), \\
   b (D) : & b (r = 2) ; c (r = -3), \\
   c (D) : & a (r = 4) ; b (r = -5) ; c (r = -1).
   \end{cases}
   \end{align*}
   $$

   The letter “D” indicates a decision node; the letter “P” indicates a probabilistic node.

   (a) Give a state transition diagram, including probabilities and rewards.

   ![State Transition Diagram](image)

   Answer.

   (b) Give the probability transition matrix, $P$, and the immediate reward matrix, $R$.

   $$
   P = \begin{bmatrix}
   s & a & b & c \\
   s & 0.4 & 0.6 & 0 & 0 \\
   a & 0.3 & 0 & 0.7 & 0 \\
   b & 0 & 0 & * & * \\
   c & 0 & * & * & *
   \end{bmatrix}, \quad R = \begin{bmatrix}
   s & a & b & c \\
   s & 6 & 7 & * & * \\
   a & 3 & 0 & * & * \\
   b & 2 & -3 & * & * \\
   c & 4 & -5 & -1 & *
   \end{bmatrix}.
   $$
Entries in $P$ that are controlled by the decision maker are starred. Blank entries are irrelevant.

(c) Express the vector of optimal values $v^* = (s, a, b, c)$ as a solution of a system of non-linear equations. It is not necessary to simplify this system of non-linear equations.

Answer.

\[
\begin{align*}
  s &= 0.4(6 + \gamma s) + 0.6(7 + \gamma a), \\
  a &= 0.3(3 + \gamma s) + 0.7(0 + \gamma b), \\
  b &= \max\{2 + \gamma b, -3 + \gamma c\}, \\
  c &= \max\{4 + \gamma a, -5 + \gamma b, -1 + \gamma c\}.
\end{align*}
\]

(d) Express the vector of optimal values $v^* = (s, a, b, c)$ as a solution of a linear program. It is not necessary to simplify this linear program.

Answer.

\[
\begin{align*}
  \text{Minimise:} & \quad s/4 + a/4 + b/4 + c/4, \\
  \text{Subject to:} & \quad s = 0.4(6 + \gamma s) + 0.6(7 + \gamma a), \\
  & \quad a = 0.3(3 + \gamma s) + 0.7(0 + \gamma b), \\
  & \quad b \geq 2 + \gamma b, \\
  & \quad b \geq -3 + \gamma c, \\
  & \quad c \geq 4 + \gamma a, \\
  & \quad c \geq -5 + \gamma b, \\
  & \quad c \geq -1 + \gamma c.
\end{align*}
\]

(e) Determine the mean hitting time for the set of all decision nodes.

Answer.

\[
\begin{align*}
  h_s &= 1 + 0.4h_s + 0.6h_a, \\
  h_a &= 1 + 0.3h_s + 0.7h_b, \\
  h_b &= 0, \\
  h_c &= 0, \\
  h_s &= 80/21 \approx 3.81, \\
  h_a &= 15/7 \approx 2.14, \\
  h_b &= 0, \\
  h_c &= 0.
\end{align*}
\]