INFO8006 Introduction to Artificial Intelligence

Exam of January 2021

Instructions

- *The exam starts at 8:30 AM.*
- *Questions must be answered on paper.*
- *Answer the questions on separate sheets, labeled with the question number, your first name, last name and student id.*
- *Answer in English or in French.*
- *Your answers must be scanned and submitted on eCampus by 12:45 AM.*
- *Your answers must be submitted in PDF in 5 distinct files named as*
  - *LASTNAME_Firstname_Q1.pdf*
  - *LASTNAME_Firstname_Q2.pdf*
  - *LASTNAME_Firstname_Q3.pdf*
  - *LASTNAME_Firstname_Q4.pdf*
  - *LASTNAME_Firstname_Q5.pdf*

Question 1 [4 points]

Multiple choice questions. Choose one of the four choices. Correct answers are graded $+\frac{4}{10}$*, wrong answers are graded* $-\frac{2}{15}$ *and the absence of answers is graded* 0*. The total of your grade for Question 1 is bounded below at* 0/4*.*

1. An approximate Q-Learning agent based on a linear model of the Q-table is an example of
   (a) learning simple reflex agent.
   (b) learning model-based reflex agent.
   (c) learning goal-based planning agent.
   (d) learning utility-based planning agent.
2. A^* necessarily reduces to uniform-cost search when
   (a) the heuristic h is random.
   (b) the heuristic h is always null.
   (c) the heuristic h is admissible.
   (d) the heuristic h is consistent.
3. In adversarial search,
   (a) the horizon effect arises when the evaluation function is perfect.
   (b) the horizon effect is negligible unless TD learning's learning rate α tends to 1.
   (c) the deeper in the tree the evaluation function is buried, the more the quality of the evaluation matters.
   (d) if not looked deep enough, bad moves may appear as good moves, because their consequences are hidden beyond the search horizon.
4. Given that $A \perp B|C$, $\mathbf{P}(A, B|C)$ is equivalent to
   (a) $\frac{\mathbf{P}(C|A)\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(C)}$.
   (b) $\frac{\mathbf{P}(B,C|A)\mathbf{P}(A)}{\mathbf{P}(B,C)}$.
   (c) $\frac{\mathbf{P}(A,C)\mathbf{P}(B|C)}{\mathbf{P}(C)}$.
   (d) $\frac{\mathbf{P}(C|A)\mathbf{P}(C|B)}{\mathbf{P}(C)}$.
5. In the Kalman filter, the prior $\mathbf{P}(X_0)$

(a) must be a uniform distribution over the discrete state values $1, \ldots, S$.

(b) is arbitrary, since it depends on one's initial beliefs.

(c) must be a Gaussian distribution over the state values.

(d) can be any continuous distribution, as long as its mean is zero.

6. Which of following is wrong? Convolutional neural networks...

(a) are usually trained by solving a maximum likelihood estimation problem.

(b) are fit for processing spatially structured data, such as images or sequences.

(c) usually count thousands to millions of parameters.

(d) have a topology that encodes conditional independence assumptions between random variables.

7. In a Markov Decision Process, if rewards r are associated to transition tuples (s, a, s'), i.e. $r = R(s, a, s')$, then the Bellman equations should be expressed as

(a) $V(s) = R(s, a, s') + \gamma \max_a \sum_{s'} P(s'|s, a)V(s')$.

(b) $V(s) = \sum_{s'} P(s'|s, a) \left[R(s, a, s') + \gamma \max_a V(s')\right]$.

(c) $V(s) = \max_a \left[R(s, a, s') + \gamma \sum_{s'} P(s'|s, a)V(s')\right]$.

(d) $V(s) = \max_a \sum_{s'} P(s'|s, a) \left[R(s, a, s') + \gamma V(s')\right]$.

8. Which of the following is true? In reinforcement learning, ...

(a) the Q-table has size equal to $|\mathcal{S} \times \mathcal{A}|$.

(b) the state-action-value $Q(s, a)$ of the q-state (s, a) is the maximum utility starting out having taken action a from s and thereafter acting optimally.

(c) direct utility estimation is an efficient algorithm for estimating V^π since its time complexity (for reaching a desired level of precision of the estimate) grows sub-linearly with the size of the state-action space.

(d) ϵ-greedy is an exploration policy that avoids visiting states s for which $V(s) < \epsilon$.

9. In DQN (Mnih et al, 2015), the Q-table is approximated with a so-called Q-network. This network is

(a) a Bayesian network.

(b) a dynamic Bayesian network.

(c) a convolutional neural network.

(d) a decision network.

10. In DQN (Mnih et al, 2015), the Q-network is trained using a variant of

(a) value iteration.

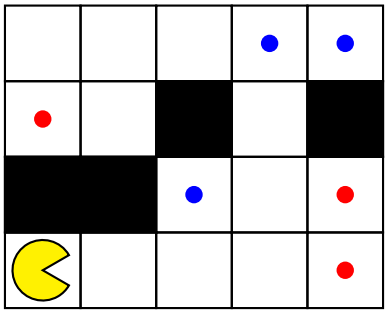
(b) policy iteration.

(c) direct utility estimation.

(d) temporal difference learning.

Question 2 [4 points]

Let us assume a grid-world in which there are two kinds of food pellets, each with a different color (blue and red). Mrs Pacman is only interested in tasting the two different kinds of food: the game ends when she has eaten (at least) 1 blue pellet and (at least) 1 red pellet. Mrs Pacman has four actions: moving north, south, east, or west; although she cannot move into any of the B walls. There are K blue pellets and K red pellets and the dimensions of the grid-world are $N \times M$.



(a) Formally define an *efficient* search problem for Mrs Pacman, for which an optimal solution would minimize the number of moves while achieving the goal. Assume she starts in location (i, j).

(b) Provide a tight upper bound on the size of the state space.

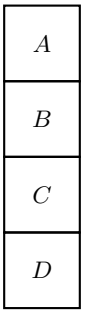
(c) Provide a tight upper bound on the branching factor of the search problem.

(d) For each of the following heuristics, indicate (yes/no) and motivate briefly (1-2 sentences) whether or not it is admissible.

- h_1: the number of remaining pellets.
- h_2: the Euclidean distance to the closest remaining pellet.
- h_3: the maximum Euclidean distance between any two remaining pellets.
- h_4: the minimum Euclidean distance between any two remaining pellets of opposite colors.

Question 3 [4 points]

Pinky the ghost escaped from Pacman and tries to hide from him. Pinky randomly moves around floors A, B, C, and D. Pinky's location at time t is X_t. At the end of each timestep, Pinky stays on the same floor with probability 0.5, goes upstairs with probability 0.3 , and goes downstairs with probability 0.2. If Pinky is on floor A, he goes down with probability 0.2 and stays put with probability 0.8. If Pinky is on floor D, he goes upstairs with probability 0.3 and stays put with probability 0.7.



(a) Given the prior $\mathbf{P}(X_0)$ below, determine the distribution of Pinky's location at time $t = 1$.

X_0	$P(X_0)$
A	0.1
B	0.2
C	0.3
D	0.4

(b) Determine the distribution of Pinky's location at time $t = \infty$.

To aid the search, a sensor S^r is installed on the roof and a sensor S^b is put in the basement. Both sensors detect either a ghost $(+g)$ or no ghost $(-g)$. The distribution of sensor measurements is determined by d, the number of floors between Pinky and the sensor. For example, if Pinky is on floor B, then $d_b = 2$ because there are two floors (C and D) between floor B and the basement, and $d_r = 1$ because there is one floor (A) between floor B and the roof. Pinky will not go onto the roof nor into the basement.

S^r	$P(S^r \mid d_r)$
$+g$	$1 - 0.3d_r$
$-g$	$0.3d_r$

S^b	$P(S^b \mid d_b)$
$+g$	$1 - 0.2d_b$
$-g$	$0.2d_b$

(c) You decide to track Pinky by particle filtering with 3 particles. At time t, the particles (1), (2) and (3) are at positions $X_t^{(1)} = A$, $X_t^{(2)} = B$ and $X_t^{(3)} = C$. In Step 1 of particle filtering, without incorporating any sensory information, what is the (joint) probability that the particles will be resampled respectively to B, B and C when projected through the transition model?

(d) Assume the particles have been resampled to B, B and C in Step 1. At $t+1$, the sensors observe $S_{t+1}^r = +g$ and $S_{t+1}^b = -g$. In Step 3 of particle filtering, now taking this sensory information into account, what is the (joint) probability that the particles will be such that $X_{t+1}^{(1)} = B$, $X_{t+1}^{(2)} = B$ and $X_{t+1}^{(3)} = B$?

(e) You find out that particle filtering with 3 particles is not very reliable and decide instead to make use of a Bayes filter. Express recursively the probability of X_t given all the measurements (up to timestep t) from the two sensors.

Question 4 [4 points]

You observe a grandmaster playing Pacman and wish to learn from her games. To this end, you write down in a table all the state-action pairs (s, a) played by the grandmaster, together with their corresponding Q-values $Q(s, a)$. You describe each state-action pair with six features: the horizontal and vertical position of Pacman (x_P, y_P) and of the ghost (x_G, y_G); the distance d to the closest food pellet; and an action feature. Unfortunately, you did not sleep too well the night before and make random errors when computing and reporting the Q-values. Assuming Gaussian errors (of zero mean and unit variance), how would you learn a model of the Q-function using your data?

(a) Describe formally the learning problem you would have to solve in the case of a linear model of the Q-function (i.e., the data, the model, and its parameters). Write down the optimization problem to estimate the model parameters.

(b) Assuming a Gaussian prior (of zero mean and unit variance) on the model parameters, revise your optimization problem above. What is the name of the resulting estimator?

(c) Derive a closed-form formula for computing the solution of this optimization problem.

(d) For the data in the table below, compute the parameters of the linear model.

x_P	y_P	x_G	y_G	d	a	Q
2	4	4	2	4	1	-1
2	2	2	2	-2	10	1
0	0	4	4	3	3	1

Question 5 [4 points]

Let us consider a simplified version of Blackjack where the deck is infinite and the dealer does not play. The deck contains cards 2 through 10, J, Q, K, and A, which are all equally likely to be drawn. Each card is worth the number of points shown on it, except for the cards J, Q, and K that are worth 10 points, and A that is worth 11. At each turn, you may either draw or stop. If you choose to draw, you are dealt with an additional card and move to the next turn. You don't receive any immediate reward. If you stop, you receive a reward of 0 if the total number of points for the cards you hold is exactly 15, 10 if it is higher than 15 but not higher than 21, and -10 otherwise (i.e., lower than 15, or larger than 21). After taking the stop action, the game ends. If your total number of points reaches 22 or higher, then you have failed: you may only choose the stop action, in which case you lose and receive a reward of -10.

Let us assume the state space $\mathcal{S}$ to be the set $\{0, 2, ..., 21, \text{fail}, \text{end}\}$, such that each state indicates either the ongoing total number of points of the player, whether the player has failed ("fail"), or if the game has ended ("end").

(a) Assume you have already performed j iterations of Value Iteration. Compute $V_{j+1}(12)$ given the table below for $V_j(s)$. The discount factor is $\gamma = 0.5$.

s	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	fail	end
$V_j(s)$	0	0	0	0	0	0	1	1	1	1	2	2	10	10	10	10	10	10	10	10	-10	0

(b) You are informed that the cards do not actually appear with equal probability and decide to use Q-learning instead of value iteration. Explain the Q-Learning algorithm as precisely as possible.

(c) Given the incomplete table of initial Q-values below, update the Q-values after the following episode occurred. Assume a learning rate of $\alpha = 0.5$ and a discount factor of $\gamma = 0.5$. Do not consider the values that Q-learning does not update.

s	a	$Q(s,a)$
19	draw	-2
19	stop	5
20	draw	-4
20	stop	7
21	draw	-6
21	stop	8
fail	stop	-8

(a) Initial values.

s	a	r	s	a	r	s	a	r
19	draw	0	21	draw	0	fail	stop	-10

(b) Episode.

(d) Dissatisfied with tabular Q-learning, you decide to model your Q-values with a linear model, representing them as $\sum_i w_i f_i(s,a)$. First consider the two feature functions

$$f_1(s,a) = \begin{cases} 0, \text{if a = stop} \\ 1, \text{if a = draw and } s \geq 15 \\ -1, \text{if a = draw and } s < 15 \end{cases} \qquad f_2(s,a) = \begin{cases} 0, \text{if a = stop} \\ 1, \text{if a = draw and } s \geq 18 \\ -1, \text{if a = draw and } s < 18 \end{cases} . \tag{1}$$

For which of the following partial policy tables is it possible to represent Q-values in the form $w_1 f_1(s,a) + w_2 f_2(s,a)$ while implying that policy unambiguously? Explain in one sentence.

s	$\pi(s)$
14	draw
15	draw
16	draw
17	draw
18	draw
19	draw

(a)

s	$\pi(s)$
14	stop
15	draw
16	draw
17	draw
18	stop
19	stop

(b)

s	$\pi(s)$
14	draw
15	draw
16	draw
17	draw
18	stop
19	stop

(c)

s	$\pi(s)$
14	draw
15	draw
16	draw
17	draw
18	draw
19	stop

(d)

s	$\pi(s)$
14	draw
15	draw
16	draw
17	stop
18	draw
19	stop

(e)