+1/1/60+

Continuous evaluation

IA (second part)
May 2019

Name :

.....

Question 1 In the context of supervised learning, for some data of the form $\mathbf{a_i} \in \mathbb{R}^n$, where $i \in [0\ m]$, we know the values of $f(\mathbf{a_i}) \in \mathbb{R}$. We would like to determine the values of $f(\mathbf{b_j})$ where $0 \leq j \leq t$ and $\forall j, b_j \notin \{a_i\}_{0 \leq i \leq m}$. Describe how you could do that using a neural network (input, output, what has to be learned). ☐0 ☐0.5 ☐1 ☐1.5

.....

.....

.....

.....

.....

Question 2 For the pong game, you tried to learn how to play using supervised learning. Why were the results so poor ? ☐0 ☐0.5 ☐1

.....

.....

Question 3 What is the main difference between supervised learning and reinforcement learning? ☐0 ☐0.5 ☐1

.....

.....

.....

Question 4 A MDP is described by 5 variables: (s, T, a, R, π). Describe in one line each of these variables. ☐0 ☐0.5 ☐1 ☐1.5 ☐2 ☐2.5

.....

.....

.....

.....

.....

Question 5 What is the optimal policy ? ☐0 ☐0.5 ☐1

.....

Question 6 Here is the second Bellman equation:

$$Q(s,a) = R(s,a) + \gamma \sum_{s'} T(s,a,s') \max_{a'} Q(s',a')$$

Describe in details what that equation is expressing. ☐0 ☐0.5 ☐1 ☐1.5 ☐2 ☐2.5

.....

.....

.....

.....

.....

Question 7 What if the main difference between the algorithm of policy gradient and the algorithm of Q-learning? ☐0 ☐0.5 ☐1

.....

.....

.....

Question 8 What is the purpose of the ϵ-greedy algorithm ? ☐0 ☐0.5 ☐1

.....

.....

Question 9 In some Q-learning implementation, it is possible to use a neural network(NN). What is this NN for (you can compare to implementation without NN) ? ☐0 ☐0.5 ☐1

.....

.....

Question 10 What is the main problem to deal with in reinforcement learning? Why are some games so difficult to solve? (you could refer to Montezuma's revenge)

☐0 ☐0.5 ☐1

.....

.....

Question 11 We play to a 3x3 frozen game. We remind you that you start at S position and your goal is to reach the G position (+1 point). F are safe places while H are very dangerous position you need to avoid: if you reach a H position, you die (-1 point). When you are in in position, you need to choose between going Up, Down, Left or Right (displacements are deterministic). Using Q-learning, compute the two first iterations, starting from 0 values and using $\gamma = 0.5$ in order to simplify computations.

☐0 ☐0.5 ☐1 ☐2