Module 6

Ibrahim Sanogo Joel Brou Boni

December 2020

1 Introduction

The goal of this module is to learn more about reinforcement learning. Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward.Through 5 questions regarding different aspects of reinforcement learning such as Q-learning value iterations algorithms or decision trees.

2 Question 1

1- The optimal path of the MDP is EENNN with a total reward of 0. But there is an other one with is the same result which is EENNWE.

2- The optimal action in each state is the following :
S(0,0) = N or E
S(1,0) = E
S(0,1) = N or E
S(2,0) = N or W
S(0,2) = N or S or E
S(1,1) = All the directions
S(2,1) = E
S(2,2) = N
S(1,2) = E OR n
S(0,3) = E
S(1,3) = E
S(2,3) = Finish, absobring state

By looking a the result we get and the optimal path from question 1, we remarked that In case we have several choices the final reward may not be the same at the end. Even if the reward is the same at the next stage, it is possible that the choice we make may not lead us to the optimal path and therefore does not maximise the reward. In this case it is necessary to look further than the next stage in order to maximise our gain.To do so it is possible to find the

optimal policy using Bellman equation as writen in the notebook
3- The total reward can be calculated through the following equation :

$$V^{\pi}(s_1) = E\left[\sum_{t=1}^{T} r(s_t, a_t, s_{t+1}) \Big| s_1\right] = \sum_{t=1}^{T} r(s_t, a_t, s_{t+1}) p(s_{t+1} | a_t, s_t)$$

By applying the formula we get an excepted total reward of 0 which correspond to what we get with the optimal path.

3 Question 3

Cf the jupyter notebook

4 Question 4

b- Exploration consists of probing a larger portion of the search space in order to maybe find other promising solutions that have not be found yet. This can there for avoid us to be stuck to local optimums while better solutions can be found. For example,a group of friends on holiday in a city (the search space) used to go to a restaurant which they found on arrival and which they like (exploitation of a local optimum). They don't suspect it, but only 200m away is a restaurant that they would like even more and which is not only cheaper, but also the best in the city. Exploring the area a little more would have allowed them to find this place.

5 Question 5

Decision Trees are a non-parametric supervised learning method used for classification and regression. Decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model.They build classification or regression models by splitting a node into two or more sub-nodes.Each node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.To do so these types of algorithms uses different tools such as the entropy that is used to calculate the homogeneity of a sample or the information gain (The construction of a decision tree is about finding the feature that returns the highest information gain).In cases where a single tree is not sufficient for producing effective results it is possible to extend them to random forest. But one of the drawbacks on a signle decision tree is that it can lead to overfitting.Therefore Random Forest is a tree-based machine learning algorithm that leverages the power of multiple decision trees for making decisions. Contrary to a decision tree built on an entire dataset, using all the features/variables of interest, whereas a random forest randomly selects observations/rows and specific features/variables to build multiple decision trees from and then averages the results.

2-One of the main differences between reinforcement learning and supervised learning is that reinforcement learning make decision sequently where the supervised algorithm's decisions are made on the initial input or the input given at the start.Moreover In supervised learning the learning process is passive.Model learns a mapping from input to output space, without altering the input space as a consequence of its learning. So every sample from the input space is independent from each other where it is the opposite for reinforcement learning which have a active learning process. This means that the model learns by interacting sequently with its environment (exploitation, exploration).