Assignment 12 (Sol.)

Introduction to Machine Learning
Prof. B. Ravindran

1. If $\{A, B, C\}$, $\{B, C\}$ and $\{A\}$ are the only frequent closed itemsets, then which of the following statements are true? (Assume that A, B and C are the only items in our universe. Note that multiple statements could be true)

   (a) B and C always occur together in a transaction
   (b) $\{A, B\}$ and $\{A, C\}$ have the same frequency
   (c) $\{A, B\}$ is not frequent
   (d) B has a greater frequency than $\{B, C\}$
   (e) There are 2 maximal frequent itemsets

   Sol. (a) & (b)
   Since $\{B, C\}$ is closed, but B and C are not, the frequency of $\{B, C\}$ is equal to the frequencies of B and C. This means whenever B occurs, C occurs. Hence, (a) is true. Since $\{A, B\}$ and $\{A, C\}$ have the same frequency as $\{A, B, C\}$, they have the same frequency. Hence, (b) is true. (c) is false since $\{A, B, C\}$ being frequent, $\{A, B\}$ has to be. Since $\{B\}$ is not closed, (d) is false.

2. For a transactional database with 7 different items (A, B, C, D, E, F, G), how many possible itemsets would need to be examined if one were to determine frequent itemsets by exhaustive enumeration (brute force)?

   (a) 255
   (b) 128
   (c) 127
   (d) 256

   Sol. (c)
   There will be $2^7 - 1 = 127$ itemsets. The subtraction by 1 is since we do not include the null set.

3. Modelling a living organism as an RL agent, the environment encompasses

   (a) everything external to the organism
   (b) some portions internal to the organism as well

   Sol. (b)
   Consider that the reward signals are generated within the brain.

4. Suppose we want an RL agent to learn to play the game of golf. For training purposes, we make use of a golf simulator program. Assume that the original reward distribution gives a reward of +10 when the golf ball is hit into the hole and -1 for all other transitions. To aide the agents learning process, we propose to give an additional reward of +3 whenever the ball is within a 1 metre radius of the hole. Is this additional reward a good idea or not? Why?

(a) Yes. The additional reward will help speed-up learning.

(b) Yes. Getting the ball to within a metre of the hole is like a sub-goal and hence, should be rewarded.

(c) No. The additional reward may actually hinder learning.

(d) No. It violates the idea that a goal must be outside the agents direct control.

Sol. (c)
In this specific case, the additional reward will be detrimental to the learning process, as the agent will learn to accumulate rewards by keeping the ball within the 1 metre radius circle and not actually hitting the ball in the hole.

5. You face a particularly challenging RL problem, where the reward distribution keeps changing with time. In order to gain maximum reward in this scenario, does it make sense to stop exploration or continue exploration?

   (a) Stop exploration

   (b) Continue exploration

   Sol. (b)
   Ideally, we would like to continue exploring, since this allows us to adapt to the changing reward distribution.

6. This question is related to the one discussed in class. Recall the temporal difference learning approach to the tic-tac-toe problem. Suppose that the probability of winning at a particular state is 0.6, the max probability value in the next set of states is 0.8, and based on our exploration policy, we choose a next state which has probability value 0.4. Should you backup the current state's probability value based on this choice of next state (i.e., move probability value 0.6 closer to 0.4) or not, given that the agent never stops exploring (i.e., the agent always makes an exploratory move some fraction of the time)?

   (a) Backup the value

   (b) Do not backup the value

   Sol. (a)
   There are valid arguments for both backing up, as well as not backing up values on exploration steps. The argument in favour of backing up is that by backing up values, even on exploration steps, the value function (i.e., the function specifying the value at each state - here, the probability of winning) actually represents the behaviour you are following (because according to your behaviour, you do explore some fraction of the times). The argument against backing up is that exploration is performed only for learning purposes, and that if we were to actually go by our learned policy, say when the agent is deployed in a competition, we will perform only exploitative steps to maximise reward. In this sense, we should only backup values on greedy moves.

   Now, in the question, we are given that the agent never stops exploring. So, if we choose not to backup the values during exploration steps, we are actually learning a value function (and hence, policy), which is different from the actual policy the agent will follow. Thus, it makes sense to backup the values during all steps.

7. Consider a transactional database, where every item (or 1 itemset e.g A or $\{A\}$) is frequent. The height of the FP Tree constructed from this database is then equal to

   (a) The number of transactions in the database
   (b) One greater than the number of frequent itemsets in the database
   (c) One greater than the maximum number of distinct items occurring in a transaction
   (d) One greater than the average number of items per transaction in the database
   (e) One greater than the number of items in the database

   Sol. (c)
   By definition of the FP-Tree, the longest transaction (in terms of number of distinct items), will trace the longest path in the tree during the insertion. The +1 is with respect to the null node.

8. How many passes does the FP-Tree algorithm perform over the transactional database?

   (a) 1
   (b) 3
   (c) 2
   (d) 4

   Sol. (c)
   It performs 2 passes - one to find the frequencies of individual items, and the second to sort each transaction in that order and insert it into the tree.