EBU4203 Introduction to AI – Week 2 Tutorial 2023

Q1: STATE the elements of Neural networks and EXPLAIN the functionalities of them.

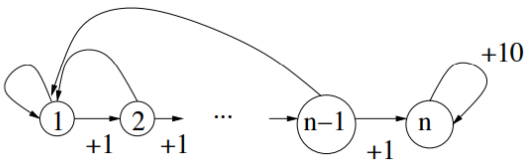
Q2: STATE and DISCUSS the four types of activation functions commonly used in the neural networks.

Q3: Given the input vector [a1,…,ak], weights [w1,…,wk] and bias b, assuming the activation function as σ, please explain what a single neuron in the neural networks does with formulas and diagrams.

Q4: How does reinforcement learning work? Please IDENTIFY the key elements and EXPLAIN the learning process.

Q5: Consider the n-state MDP in the figure below. In state n there is just one action that collects a reward of +10, and terminates the episode. In all the other states there are two actions: float, which moves deterministically one step to the right, and reset, which deterministically goes back to state 1. There is a reward of +1 for a float and 0 for reset. The discount factor is γ =1/2.

i) Compute the optimal value function, V*(k) for all k=1,….,n-1.
ii) Assuming V*(n)=20, V*(1)=1, instead of taking deterministic strategy, the agent now has 0.8 probability to move to right, and 0.2 probability to reset, calculate the V*(n-1).



Q6: This Gridworld problem is shown in Fig.1. The states are grid squares, dentified by their row and column number (row first). The agent always starts in state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. (The reward for a state is received as the agent moves into the state.) The transition function is such that the intended agent movement (North, South, West, or East) happens with probability 0.8. With probability 0.1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state. Please answer the following questions.

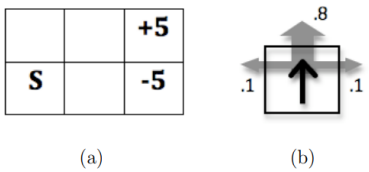


Figure 1: (a) Gridworld MDP. (b) Transition function.

(i). Draw the optimal policy for this grid.

(ii). Suppose the agent knows the transition probabilities. Give the first two rounds of value iteration updates for each state, with a discount factor γ= 0.9. (Assume V0 is 0 everywhere and compute Vi for times i = 1, 2).

(iii). Suppose the agent does not know the transition probabilities. What does it need to be able do (or have available) to learn the optimal policy?

(iv). When using Q-learning to solve this GridWolrd problem, how do you formulate it as a Markov decision process (MDP)?

(v) Based on the formulated MDP above, please create the Q-table. When assuming the agent moves two steps towards right, calculate the Q-value and update the Q-table. (learning rate $\alpha = 0.1, \gamma = 0.9$)