**Department of Information and Computer Science**

**ICS 381: Principles of Artificial Intelligence
Second Semester 2018/2019 (182)**

**Solution of Quiz No. 4**

**Name: ID:**

**Name: ID:**

**Name: ID:**

**Name: ID:**

**Question 1:**

You are in-charge of a financial stock broker that needs to invest on behalf of your company clients. At any time, you can buy or sell shares. You get a reward of ***+10*** for ***buying*** ***shares*** and **+4** for ***selling shares***. Your investment strategy can be in one of the following three states: ***sure***, ***unsure***, ***broke***. ***Buying shares tends to make you unsure***, while ***selling shares tends to make surer***. ***If you get too unsure, you might end up in the broke state, for good***. The transitions of the modeled Markov decision process (MDP) is shown in the table below. Because your investment strategy depends on the stock market observations, there is a discount of ***γ = 0.9***.

|  |  |  |  |
| --- | --- | --- | --- |
| s | a | s’ | T(s, a, s’) |
| sure | sell | sure | 1 |
| sure | buy | sure | 1/4 |
| sure | buy | unsure | 3/4 |
| Unsure | sell | Sure | 1/4 |
| Unsure | sell | Unsure | 3/4 |
| Unsure | buy | Unsure | 7/8 |
| Unsure | buy | broke | 1/8 |

### How many possible deterministic stationary policies are there? There are 2 states, 2 possible actions in each state, so 2 · 2 = 4 policies.

### What is the value of the state *sure* under the policy that always sells the shares? 40. There are two straightforward ways to solve this problem. One is to note that: 4 + γ + 4 γ2 +…+ 4 γ ∞ = 4 / (1 - γ) = 40 (for γ = 0.9). It can also be found using:  V(sell) = 4 + γ V(sell) = 4 + 0.9 V(sell) which gives V(sell) = 40.

### Fill in the following table of depth-limited values from value iteration for this MDP.

|  |  |  |  |
| --- | --- | --- | --- |
| s | V0(s) | V1(s) | V2(s) |
| sure | 0 | 10 | 19 |
| unsure | 0 | 10 | 17.875 |
| broke | 0 | 0 | 0 |

### Note: This question concerns (optimal) value iteration, not evaluation of the always-sells policy.

### How many rounds of value iteration will it take for the values of all states to converge to their exact values? (State infinitely many if you think it will only have converged after infinitely many.)

### (sure, sell, 4) → sure(sure, buy, 10) → sure(sure, buy, 10) → sure(sure, buy, 10) → unsure(unsure, sell, 4) → sure

### What is the optimal policy for *γ = 0.9*?

|  |  |
| --- | --- |
| s | π\*(s) |
| sure | buy |
| unsure | sell |

### What are the optimal values for the optimal policy when *γ = 0.9*?

|  |  |
| --- | --- |
| s | V\*(s) |
| sure | 59.5 |
| unsure | 43.5 |

### Your investment manager is requesting faster profits. He tells you that the company clients do not care about the far future of their portfolios. In particular, they say that your discount parameter *γ* must be set *0.5*. What are the optimal policy and values now?

|  |  |
| --- | --- |
| s | V\*(s) |
| sure | 19.0476 |
| unsure | 17.7778 |

|  |  |
| --- | --- |
| s | π\*(s) |
| sure | buy |
| unsure | buy |