Step 1: Complete Problem 1 : Conceptual  
a) How would you define Reinforcement Learning? How is it different from regular supervised  
or unsupervised learning? [2 points]

Reinforcement Learning is a way to define an environment and build an optimal solution, path, or methodology to interact with that environment. To put it in layman’s terms, RL is teaching a computer the “rules of the game” and allowing the machine to find the best way to “win”. The “rules” include things like variables, states, transition probabilities, constraints, and optimization functions.

Supervised or unsupervised learning both focus on either classification or regression, they produce some sort of generalized equation. In contrast, RL focuses on building a Markov Decision process. The use cases are different, and therefore the inputs are different. Supervised/unsupervised learning focus on working with given dataset or samples, and RL focuses on interacting with the environment.

b) Can you think of three possible applications of RL that were not mentioned in the lecture?  
For each of them, what is the environment? What is the agent? What are possible actions?  
What are the rewards? [3 points]

c) What is the discount rate? Can the optimal policy change if you modify the discount rate? [1 points]

d) How do you measure the performance of a Reinforcement Learning agent? [1 points]

e) What is the credit assignment problem? When does it occur? How can you alleviate it? [2 points]

f) What is the point of using a replay memory? [1 points]

g) What is an off-policy RL algorithm? [1 points]

Step 2: Complete Problem 2: Markov Decision Processes (MDP)  
A robot operates on a hill and uses photovoltaic cell to recharge. That robot can be in one of four states: low, medium, high and top on the hill. If it spins its wheels slowly, it climbs the hill in each time step (from low to medium , from medium to high, from high to top) with a probability of 0.3. It slides down the hill to low with a probability of 0.7. If it spin its wheels rapidly, it climbs the hill in each time step from low to medium , from medium to high, from high to top) with a probability of 0.5. It slides down the slope to low with a probability of 0.5.

Spinning its wheels slowly uses one unit of energy per time step while spinning its wheels rapidly uses two units of energy per time step. The robot is low on the hill and wants to reach the top with minimum energy usage.

a) Draw a diagram of Markov Decision Process [3 points].

b) Solve the Markov Decision Process using undiscounted value iteration for the first 5 iterations (clearly outline the process) [5 points].

c) Describe the optimal policy [1 points].