MANIPAL SCHOOL OF INFORMATION SCIENCES
MANIPAL
(A constituent unit of MAHE, Manipal)

AML 5204 | Reinforcement Learning | Sessional-1

1. [10 points] [L5, CO 1] For each of the following scenarios, decide which learning type (supervised/unsupervised/reinforcement learning) is most appropriate with justification not more than two lines for each:

   - credit scoring to assess the credit worthiness of individuals. To assign a score to individuals depending on past history of their transactions and debt repayments; Supervised
   - optimize a manufacturing process by learning from trial and error; R·L
   - have a robot train itself to self balance; R·L
   - a movie recommendation system based on categorizing customers into different groups;
     Unsupervised
   - predicting the time a patient would continue to spend in ICU based on their test results.
     Supervised

2. [10 points] [L5, CO 1] For each of the following scenarios, identify whether it is an episodic task, decide the rewards that may be assigned to the agent, and choose the discount factor γ:

   - you are trying to set up an environment and train an agent to play Golf. The goal of the agent is to hit the ball into the flag post in the least number of tries; Episodic, +ve for reaching goal and −ve for every other step. $\gamma = 1$
   - robot trying to balance itself (trying not to fall);
     Non-Episodic, +1 for every step. −ve for falling. $\gamma = 0.9$
   - robot trying to navigate out of a maze.
     Episodic, +ve if navigates out, $\gamma = 1$

   Note: justify your answer in not more than 3 lines for each case.

3. [10 points] [L3, CO 1] Consider a scenario where an individual can be either well or unwell on a particular day. The individual can either decide to rest or to work on the same day with the following possibilities:

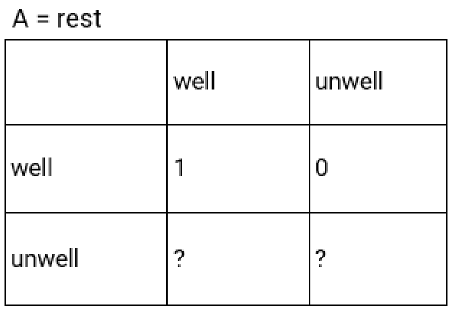
   - When they are well, and decide to work, there is a 5% chance that they become unwell on the next day. However, if the person rests given that he/she is well, there is a 100% chance of being well on the next day also.

- When the person is unwell, if they rest, there is a 90% chance of recovery becoming well the next day; otherwise, if they decided to work, there is a 50% chance of still being unwell on the next day.

We are also given the following: State space $S = \{well, unwell\}$ and Action space $A = \{rest, work\}$.

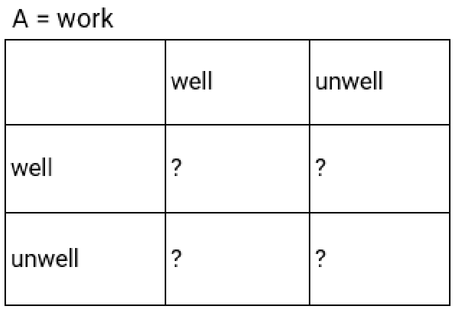
(a) Draw a state transition diagram representing the states and transition probabilities for each action the individual takes.

(b) Write down the transition probabilities in the form of a transition matrix for the two possible actions as shown below:



A = rest

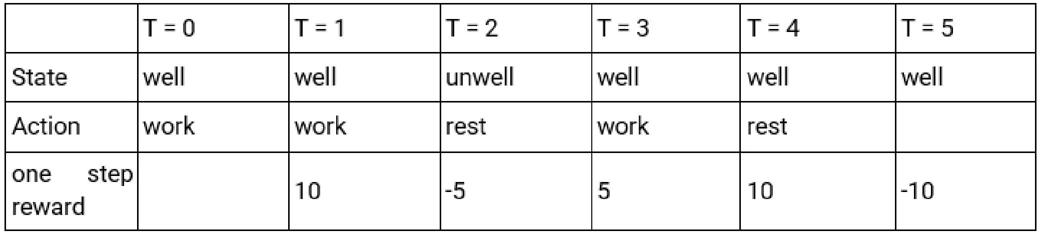
	well	unwell
well	1	0
unwell	?	?



A = work

	well	unwell
well	?	?
unwell	?	?

4. [10 points] [L3, CO 3] Continuing from the setup of the previous question, a particular individual undergoes the following state transitions over 5 days because of the actions they took as follows:



	T = 0	T = 1	T = 2	T = 3	T = 4	T = 5
State	well	well	unwell	well	well	well
Action	work	work	rest	work	rest	
one step reward		10	-5	5	10	-10

Use the discount factor $\gamma = 0.1$ and calculate for this individual

(a) G_0, G_1, G_2, G_3, G_4;

(b) values of the states *well* and *unwell* (that is, $v_\pi(well)$ and $v_\pi(unwell)$).

5. [10 points] [L3, CO 3] There is a robot that performs a task of moving inventory from one location to another. The robot's battery charge level can either be *high* or *low* or *dead*. The robot can either take a decision to *work* or to *charge* itself. Charging the robot will make the charge level of batteries transition to *high*. When the robot decides to *work*, there is a 10% chance that the charge decreases by one level

($high \rightarrow low$ or $low \rightarrow dead$) and a 90% chance that the battery level remains the same. The robot gets a reward of +10 if it performs the task ($work$) and does not get any reward if it decides to $charge$. However, if its batteries reach the $dead$ state, as a penalty the robot gets a reward of -20.

Given that we model the rewards that the agent gets as an MDP, specify what the following are:

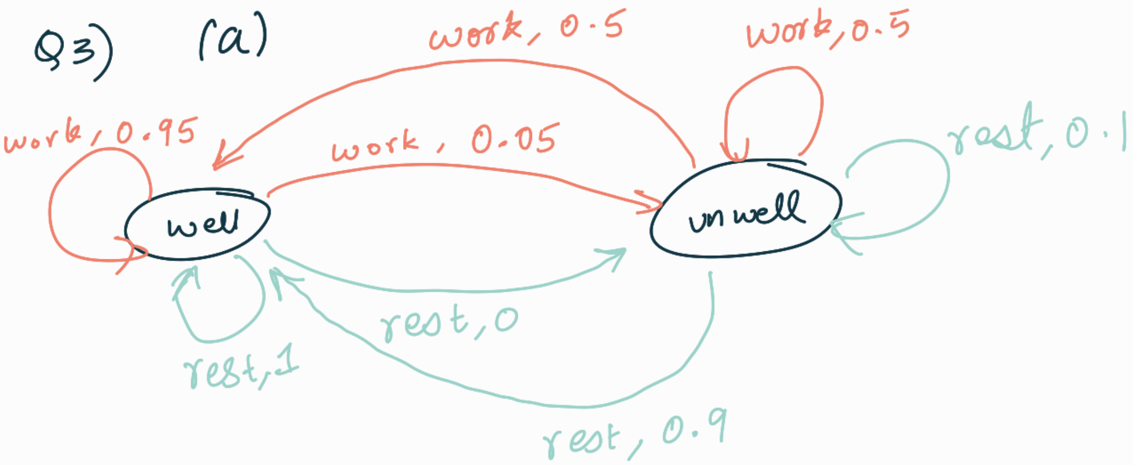
(a) state space

(b) action space

(c) transition probabilities for all transitions such as $P(high|high, charge)$ (ignore the transitions that have zero transition probability)

(d) one step rewards $r(high, work, high), r(high, charge, high), r(low, charge, high), r(low, work, dead)$.

Q3) (a)



(b)

A = rest

	well	unwell
well	1	0
unwell	0.9	0.1

A = work

	well	unwell
well	0.95	0.05
unwell	0.5	0.5

Q4) (a) $G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots\dots$

$G_t = R_{t+1} + \gamma G_{t+1}$

$G_4 = R_5 + \gamma G_5 = -10 + 0.1(0) = -10$

$G_3 = R_4 + \gamma G_4 = 10 + 0.1(-10) = 9$

$G_2 = R_3 + \gamma G_3 = 5 + 0.1(9) = 5.9$

$G_1 = R_2 + \gamma G_2 = -5 + 0.1(5.9) = -4.41$

$G_0 = R_1 + \gamma G_1 = 10 + 0.1(-4.41) = 9.559$

(b) $v_\pi(\text{well}) = E[G_t | S_t = \text{well}]$

$= \text{Avg}(G_0, G_1, G_3, G_4)$

$= \dfrac{G_0 + G_1 + G_3 + G_4}{4}$

$$= \frac{9.5 + (-4.4) + 9 + (-10)}{4}$$

$$= \frac{4.1}{4} = 1.025$$

$$v_\pi(\text{unwell}) = E[G_t \mid S_t = \text{unwell}]$$

$$= \text{Avg}(G_2)$$

$$= G_2 = 5.9$$

Q5)

(a) State Space

$S = \{\text{high}, \text{low}, \text{dead}\}$

(b) $A = \{\text{charge}, \text{work}\}$

(c)

$P(\text{high} \mid \text{high}, \text{charge}) = 1$

$P(\text{high} \mid \text{low}, \text{charge}) = 1$

$P(\text{high} \mid \text{dead}, \text{charge}) = 1$

$P(\text{high} \mid \text{high}, \text{work}) = 0.9$

$P(\text{low} \mid \text{high}, \text{work}) = 0.1$

$P(\text{low} \mid \text{low}, \text{work}) = 0.9$

$P(\text{dead} \mid \text{low}, \text{work}) = 0.1$

(d)

$r(\text{high}, \text{work}, \text{high}) = 10$

$r(\text{high}, \text{charge}, \text{high}) = 0$

$r(\text{low}, \text{charge}, \text{high}) = 0$

$r(\text{low}, \text{work}, \text{dead}) =$

$10 - 20 = -10$