COMP-424: Artificial intelligence

Tutorial 5

Question 1: Utility theory (from Russell & Norvig)

A used-car buyer can decide to carry out various tests with various costs (e.g., kick the tires, take the car to a qualified mechanic, etc.) and then, depending on the outcome of the tests, decide which car to buy. A car can be in good shape (denoted $q+$) or in bad shape (denoted $q-$) and the tests might help indicate what shape the car is in.
Consider the case where a buyer is considering a car (denoted c_1) that costs \$1500.
Its market value is \$2000 if it is in good shape; if not, \$700 in repairs will be needed to make it in good shape.
The buyer's prior estimate is that c_1 has a 70% chance of being in good shape.
There is time to carry out at most one test (denoted t_1), that test costs \$50.

a) Draw a decision network that represents this problem.
b) Calculate the expected net gain from buying c_1, given no test.
c) Tests can be described by the probability that the car will pass or fail the test given that the car is in good or bad shape. We have the following information: $P(pass(c_1,t_1)|q+)=0.8;\ P(pass(c_1,t_1)|q-)=0.35$.
Use Bayes' theorem to calculate the probability that the car will pass (or fail) its test and hence the probability that it is in good (or bad) shape given each possible test outcome.
d) Calculate the optimal decisions given either a pass or a fail, and their expected utilities.
e) Calculate the value of information of the test, and derive an optimal conditional plan for the buyer.

Question 2: Markov decision processes and reinforcement learning

Every day, Fred wakes up and plays the lottery. In Fred's special world, there are two distinct lotteries:
L1: with probability p Fred will be alive the next day, with probability $1-p$ he will be dead.
L2: with probability q Fred will be alive the next day, with probability $1-q$ he will be dead.
But L2 can only be played once, after which, if played again, it results in death with probability 1.
Regardless of which lottery is played, every day Fred is alive, he receives n units of happiness. Once dead, he experiences no happiness.

(a) Draw the Markov Decision Process which best represents Fred's situation.

(b) Fortunately, Fred has many lives during which he can learn how to choose between lotteries. You observe his sequence of actions during his first five lives (presume that the sequence stops when he dies):
Life 1: L1, L2, L1, L1, L1
Life 2: L2, L1, L1
Life 3: L1, L1, L1, L2, L1, L1
Life 4: L1, L2
Life 5: L2, L1, L1, L1, L1
Using the Monte-Carlo method, estimate p and q.

(c) What is the optimal (deterministic) policy for this MDP, given your estimates in part (b)? Assume a discount factor, $\gamma = 0.9$.

(d) Now estimate the value function at each state using the TD-method, based on the sequence of actions from Fred's first life. Assume he starts out alive, and having played neither lottery. Also assume $n=1$, learning rate $\alpha = 0.1$, and discount factor $\gamma = 0.9$. Give the TD-update rule and fill in a table (similar to the one shown below) with state values after every action is taken.

State/action pair	V(s1)	V(s2)	V(s3)
Initial value	0	0	0
$t = 1 : s =$ $, r =$ $, a = L1$			
$t = 2 : s =$ $, r =$ $, a = L2$			
$t = 3 : s =$ $, r =$ $, a = L1$			
$t = 4 : s =$ $, r =$ $, a = L1$			
$t = 5 : s =$ $, r =$ $, a = L1$			
$t = 6 : s =$ $, r =$			