COMP-424: Artificial intelligence

Homework 4

Due on *myCourses* Saturday Apr 14, 11:59pm.

General instructions.

- This is an individual assignment. You can discuss solutions with your classmates, but should only exchange information orally, or else if in writing through the discussion board on *myCourses*. All other forms of written exchange are prohibited.
- Unless otherwise mentioned, the only sources you should need to answer these questions are your course notes, the textbook, and the links provided. Any other source used should be acknowledged with proper referencing style in your submitted solution.
- Submit a single pdf document containing all your pages of your written solution on your McGill's *myCourses* account. You can scan-in hand-written pages. If necessary, learn how to combine many pdf files into one.

Question 1: Hidden Markov Models

You need to design a HMM which can be used to make decisions regarding buying and selling stocks. Your HMM should have two states {High, Low} and three observations {Buy, Sell, Keep}.

Starting probability of stock market being in a low state is 0.4. At each time step if the market is in a low state, then there is 0.4 probability that market will remain in low state during the next time step. If the market is in high state then there is 0.3 probability that it will remain in high state. Given that market is in high state probability of observing Sell is 0.6 and probability of observing Keep is 0.3. If the market is in low state the probability of observing Buy is 0.5 and probability of observing Keep is 0.3.

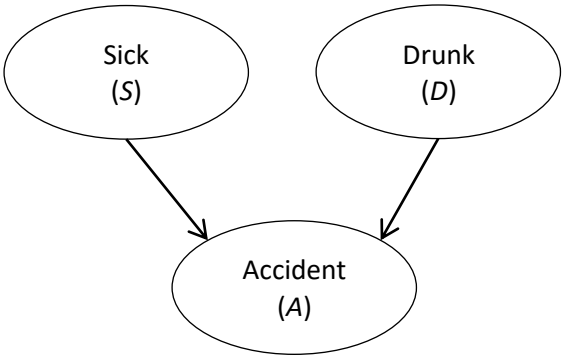
a) Draw the HMM that describes this scenario, clearly indicating the relevant parameters using conditional probability tables.
b) What is the probability of the observation sequence {*Buy*, *Keep*, *Sell*}?
c) What is the probability of the stock state being High after observing the sequence in b)?
d) What is the most likely sequence of three states explain the observation sequence in b)?

For each of these tasks, write down your calculations, or provide the code that you wrote to compute the answer.

S (F)	S (T)
0.8	0.2

D (F)	D (T)
0.9	0.1

Question 2: Utility



Consider the Bayes Net shown here, with all Bernoulli variables, which models driving accidents. Having an accident has a utility of -200 if the car was insured, but has a utility of -400 if the car was not insured (or the insurance claim is denied). Having insurance when there is no accident has a utility of -10, and not having insurance and an accident has a utility of 0.

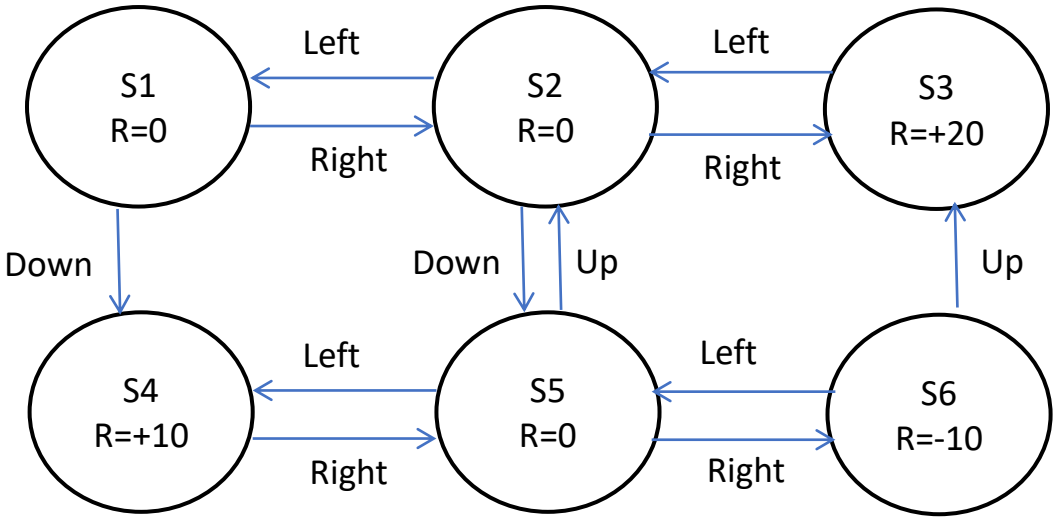
Use the principle of Maximum Expected Utility and Value of Information to answer the following questions. For parts a)-c) assume the insurance company pays for 100% of the cases.

S	D	A (T)
F	F	0.1
F	T	0.2
T	F	0.3
T	T	0.5

a) Given no information on whether the driver will be driving in a sick or drunk state, how much should they pay to insure the car?
b) How much should they pay for insurance if you know for certain that they will drive both sick *(S=True)* and **not drunk** *(D=False)*?
c) How much should they pay for insurance if you know that they will drive drunk *(D=True)*?
d) A company is offering cheaper insurance but has a reputation of rejecting 20% of insurance claims. How much should they charge for this insurance, to make it competitive with the insurance offered by the more reliable company? (*Hint*: Set the cost of the new insurance to have the same MEU as the other insurance.)

Question 3: Markov Decision Processes

Consider the MDP shown below. It has 6 states and 4 actions. As shown on the figure, the transitions for all actions have a Pr=0.8 of succeeding (and leading to the state shown by the arrow) and Pr=0.2 of failing (in which case the agent stays in place). For other transitions that are not shown, assume that they cause the state to stay the same (e.g. *T(S1,Left,S1)=1*). The rewards depend on state only and are shown in each node (state); rewards are the same for all actions (e.g. *R(S4,a)=+10, ∀a*). Assume a discount factor of $\gamma=0.9$.



a) Describe the space of all possible policies for this MDP. How many are there?
b) Assuming an initial policy $\pi^0(s)=Right, \forall s$, perform policy evaluation to get the initial value function for each state, $V^0(s)$, $\forall s$.
c) Given the initial estimate, V^0, if you run an iteration of policy improvement, what will be the new policy at each state? If necessary, break ties alphabetically, e.g. "Down" before "Left", etc.)
d) What is the optimal value function at each state for this domain?
e) Is the optimal value function unique? Explain.
f) What is the optimal policy at each state for this domain?
g) Is the optimal policy unique? Explain.
h) Suggest a change to the reward function that changes the value function but does not change the optimal policy.

Question 4: Bandits

Consider the following 5-armed bandit problem. The initial value estimates of the arms are given by Q = {1, 1, 3, 1, 1}, and the actions are represented by A={1, 2, 3, 4, 5}. We observe a trajectory consisting of plays and rewards: T={A_1=3, R_1=2, A_2=5, R_2=0, A_3=3, R_3=1, A_4=1, R_4=0, A_5=3, R_5=0}.

a) Show the estimated Q values at each time step of the trajectory using the average of the observed rewards, where available. Do not consider the initial estimates as samples.
b) It turns out the player was following an epsilon-greedy strategy. For each time step, report whether it can be concluded with certainty that a random action was selected.