UBC-CS CPSC 422, 2017

Homework 2: Due, Fri Oct 20, noon
(tot: 170 points)

Question 1: Reinforcement Learning : Q-learning (40 points)

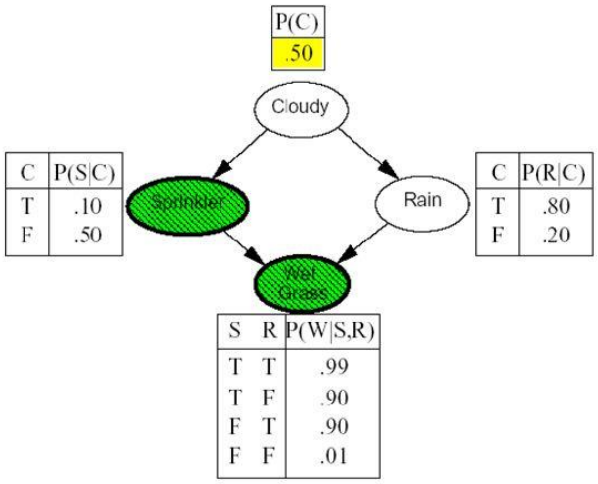
Suppose an agent is doing Q-learning with discount factor γ = 0:9. The agent has 4 actions available to it: up, right, left and down. Initially all Q-values are zero. Assume that it is using $\alpha_k = 1/k$.
(a) [15 points] Suppose initially the agent had the following sequence of state-action-reward experiences:
s17; right; 2; s18; up; 8; s14; right;-6; s15
Show what Q values are assigned due to this sequence of experiences. (show your work)

(b) [15 points] Suppose that later the agent had the following sequence of state-action-reward experiences (and it had not previously visited these states except for the experiences in the previous part):
s23; up; 0; s18; up; 0; s14; right; 10; s15
Show what Q values are assigned due to this sequence of experiences. (show your work)

(c) [10 points] Which update from the previous part would be different if you were running SARSA? why?

Question 2: Approximate Reasoning in Belief Networks (75 points)

For this exercise, you will work with the network represented below, and assume you have observed that the *sprinkler is on* and the *grass is wet*. We will label the first event by *s* and the second by *w*. We will label the event corresponding to rain as *r* and estimate the probability of the event *P(r|s,w)* using sampling methods.



We will be using existing implementations of rejection sampling and likelihood weighting and focus on the interpretation of their results. You can use any programming language and/or plotting tool of your choice for this analysis (note that there are many data points, so MS Excel might not scale

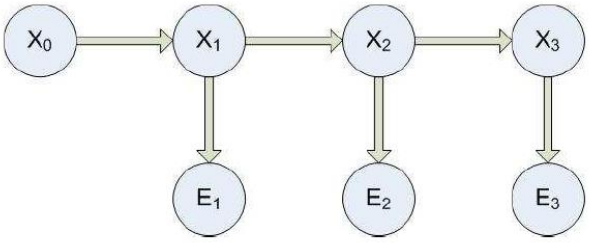
handle them well). (One possibility is Matlab, which can read the provided files with the function csvread and plot results on a logarithmic x-axis with the function semilogx.)
To get the files needed for this assignment, download and unzip file *hmw1.zip* from the course webpage (or Connect).
Rejection sampling and likelihood weighting are anytime algorithms: they already yield results with a small number of samples, but results continuously improve as the number of samples grows. We will study this characteristic by plotting *P(r|s,w)* based on a different number of samples.

a. **(points 25)** File *rs_1.csv* contains the result of running rejection sampling for 100000 samples: each line contains one sample; -1 denotes a rejected sample, 1 denotes an accepted sample with R=T, 2 denotes an accepted sample with R=F. Generate a graph whose x-axis is the number of used samples N, and whose y-axis is *P(r|s,w)* as computed based on the first N samples. Start the graph with the first accepted sample (for this particular file, the 2nd one) in order to avoid divisions by zero; make the x-axis logarithmic so you can see the algorithm's behavior for small sample sizes. What is the algorithm's approximation of *P(r|s,w)* using 100000 samples?

b. **(points 25)** Let *p* denote the true probability *P(r|s,w)* and let *s* denote the approximation of *p* using rejection sampling based on *n* accepted samples. Theoretically, using Hoeffding's inequality, derive the tightest bound ε such that $P(|s - p| > \varepsilon) < 0.05$. How many accepted samples *n* are there at N=100000? What is the value of ε for that *n*? Augment your plot from the question above with upper and lower confidence bounds $s + \varepsilon$ and $s - \varepsilon$ where both *s* and ε are computed from the accepted samples up to that point. (Remember that some samples are rejected, so N is not equal to *n*.)

c. **(points 25)** File *lw_1.csv* contains the result of running likelihood weighting for 10000 samples; each line contains two numbers: the first number is the sample (1 denotes R=T, 2 denotes R=F) and the second number is the weight. What is the algorithm's approximation of *P(r|s,w)* using 100000 samples? Generate a graph equivalent to your graph in the question above; which algorithm seems to converge faster?

Question3: Temporal Reasoning in Belief Networks (45 points)
You are given the temporal model in the figure below, where all variables are Boolean.



The transition model $P(X_{t+1} | X_t)$ is given by the CPT

X_t	$P(X_{t+1} = t)$
t	0.7
f	0.4

While the sensor model $P(E_t| X_t)$ is given by the CPT

X_t	$P(E_t = t)$
t	0.8
f	0.3

Suppose to observe the sequence $\boldsymbol{e} = (t, f, t)$.
Let's also suppose that $P(X_0 = true) = P(X_0 = false) = 0.5$

(a) **(20 points)** Find the probability estimates $p(X_i = t| \boldsymbol{e})$ for i=1,2,3 using the appropriate algorithm covered in class. Make sure to show the steps of your solution, not just the final number.
(b) **(25 points)** Find the most likely sequence of states (s_1, s_2, s_3) of variables X_1, X_2, X_3 using the appropriate algorithm covered in class. Again, make sure to show the steps of your solution, not just the final numbers.

Question4 - (10 points)

[Note that this question is worth marks, so don't forget to do it.]
Fill out the following google form. Thanxs!
https://goo.gl/forms/byZEFAWOOiTFpM982