CIS 4190/5190: Applied Machine Learning	Fall 2022
Homework 6	
Handed Out: November 16	*Due: November 30, 8:00 p.m.*

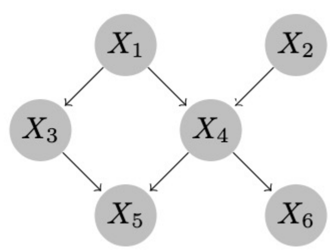
- You are encouraged to format your solutions using LaTeX. You'll find some pointers to resources for learning LaTeXamong the Canvas primers. Handwritten solutions are permitted, but remember that you bear the risk that we may not be able to read your work and grade it properly — do not count on providing post hoc explanations for illegible work. You will submit your solution manuscript for written HW1 as a single PDF file.

- The homework is **due at 8:00 PM** on the due date. We will be using Gradescope for collecting the homework assignments. Please submit your solution manuscript as a PDF file via Gradescope. Post on Piazza and contact the TAs if you are having technical difficulties in submitting the assignment.

1 Multiple Choice & Written Questions

Note: You do not need to show work for multiple choice questions. If formatting your answer in LaTeX, use our LaTeX template hw6_template.tex (This is a read-only link. You'll need to make a copy before you can edit. Make sure you make only private copies.).

1. [Bayesian Network] (10 pts) Consider the following Bayesian network over six random variables $X_1, X_2, X_3, X_4, X_5, X_6$ shown below (assume for simplicity that each random variable takes two possible values):



   (a) (1pts) Write the joint probability distribution function $p(X_1, X_2, X_3, X_4, X_5, X_6)$ as a product of factors (one for each variable).

   (b) (3 pts) If you remove the edge $X_1 \to X_3$ from the above network, will the class of joint probability distributions represented by the resulting network be smaller or larger than the class associated with the original network? Briefly explain your answer.

   (c) (3 pts) Consider the following probability distribution:

$$p(X_1)p(X_2)p(X_3 \mid X_2)p(X_4 \mid X_1, X_2)p(X_5)p(X_6).$$

What edge(s), if any, would we have to add for it to be represented by the Bayesian network above?

(d) (8 pts) Given the above figure, use d-separation to determine whether each of the following is true or false. Briefly justify your answer.

i. $p(X_1, X_2) = p(X_1)p(X_2)$

ii. $p(X_3, X_6 \mid X_4) = p(X_3 \mid X_4)p(X_6 \mid X_4)$

iii. $p(X_1, X_2 \mid X_6) = p(X_1 \mid X_6)p(X_2 \mid X_6)$

iv. $p(X_2, X_5 \mid X_4) = p(X_2 \mid X_4)p(X_5 \mid X_4)$

2. [Attention Mechanism] (6 pts) In this problem, we will walk through how self-attention layers we introduced in class are calculated. Suppose we have a transformer model, where the hidden state. We input the sentence "Bob likes cookie" into the model, and below are the values of the query, value, and key vectors we get from the model.

Index	Input Word	Key k	Query q	Value v
1	Bob	$[0.8, 0.2, 0.2]$	$[0.2, 0.6, 0.2]$	$[0.9, 0.1, 0.0]$
2	likes	$[0.1, 0.7, 0.3]$	$[0.3, 0.3, 0.1]$	$[0.2, 0.9, 0.5]$
3	cookie	$[0.0, 0.5, 0.4]$	$[0.4, 0.1, 0.8]$	$[0.6, 0.1, 0.8]$

We will focus on computing the next vector for "cookie". You are welcome to use electronic devices to help with calculations in this question.

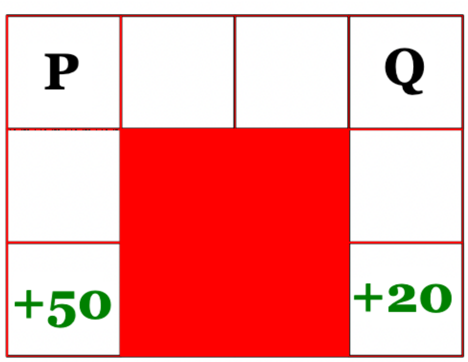
(a) (3 pts) What is the attention weight of "cookie" for each of the 3 words in the sequence? Recall that the attention weight of cookie for word i is

$$a_3 = \text{softmax}([q_3^\top k_i]_{i=1}^3).$$

You can round your answer to two decimal places.

(b) (3 pts) What is the output of the attention layer for "cookie"?

3. [Value Iteration] (12 pts) Consider the following gridworld environment:



The agent can move to adjacent squares (North, East, South, West), and does not move if it uses an invalid action. Unless otherwise specified, assume that the agent always moves successfully if its target is valid. The episode ends when the agent moves to a state with a green number, and collects the reward shown. All other rewards are zero. Assume that the discount factor is $\gamma = 1$. Suppose that we use value iteration to compute V^*. In particular, we start with the initial value function $V_1(s) = 0$ for all $s \in S$, and $V_i(s)$ represents the value function at the end of the ith iteration. For example, we have $V_2(s) = 0$ except for states with rewards, in which case $V_2(s)$ equals the reward (either 50 or 20).

(a) (4 pts) For what i is $V_i(P) > 0$? For what i is $V_i(P) = V^*(P)$?

(b) (4 pts) For what i is $V_i(Q) > 0$? For what i is $V_i(Q) = V^*(Q)$?

(c) (4 pts) Suppose that the actions have a failure probability: with probability 0.8, the agent reaches its target, but with probability 0.2, it remains in the same square. How does this change your answer to (a)?

4. [Reinforcement Learning] (5 pts) Imagine that you are designing a robot to run a maze. You decide to give it a reward of +1 for escaping the maze and a reward of 0 at all other times. The task breaks down naturally into "episodes"—the successive runs through the maze that terminate when you reach the goal (as in the gridworld example), so you decide to treat it as an episodic task, where the goal is to maximize expected total reward in the episode $R = r_1 + r_2 + ... + r_T$, where T is the number of steps in an episode. Suppose at some point during training, the policy has learned to solve the maze but does so inefficiently, missing several shortcuts.

(a) (1 pts) Would further training help the policy learn a more efficient route?

(b) (2 pts) Would introducing a discount factor $\gamma < 1$ fix the issue?

(c) (2 pts) Would introducing a small negative reward on each step fix the issue?

2 Python Programming Questions

A Google Colab notebook is linked in the "HW6 Coding" assignment on Canvas. This will tell you everything you need to do, and provide starter code.