CS 188
Fall 2024

Introduction to
Artificial Intelligence

Written HW9

Due: Tuesday 11/12 at 11:59pm.

Policy: Can be solved in groups (acknowledge collaborators) but must be submitted individually.

Make sure to show all your work and justify your answers.

Note: This is a typical exam-level question. On the exam, you would be under time pressure, and have to complete this question on your own. We strongly encourage you to first try this on your own to help you understand where you currently stand. Then feel free to have some discussion about the question with other students and/or staff, before independently writing up your solution.

Note: Leave the self-assessment sections blank for the original submission of your homework. After the homework deadline passes, we will release the solutions. At that time, you will review the solutions, self-assess your initial response, and complete the self-assessment sections below. The deadline for the self-assessment is 1 week after the original submission deadline.

Your submission on Gradescope should be a PDF that matches this template. Each page of the PDF should align with the corresponding page of the template (page 1 has name/collaborators, question begins on page 2.). **Do not reorder, split, combine, or add extra pages**. The intention is that you print out the template, write on the page in pen/pencil, and then scan or take pictures of the pages to make your submission. You may also fill out this template digitally (e.g. using a tablet.)

First name	
Last name	
SID	
Collaborators	

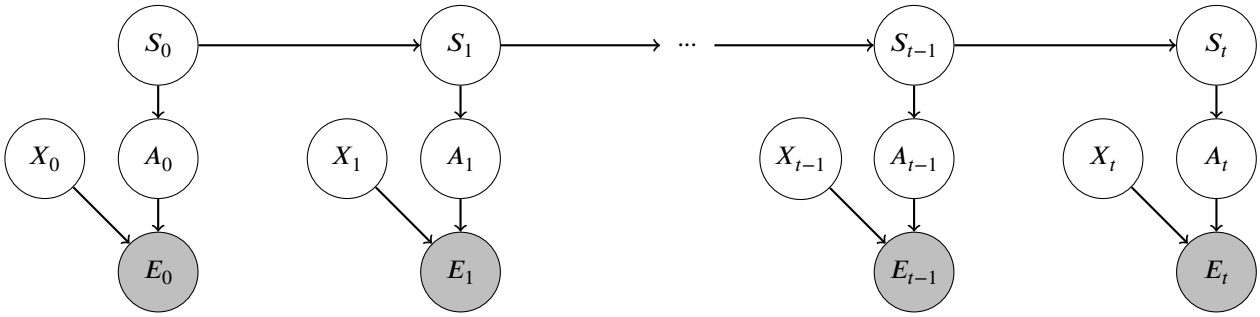
CS 188 Fall 2024 | Introduction to Artificial Intelligence | Written HW9

Q1. [12 pts] Course Evaluations

Every semester we try to make CS 188 a little better. Let S_t represent the quality of the CS 188 offering at semester t, where $S_t \in \{1, 2, 3, 4, 5\}$. Changes to the course are incremental so between semester t and $t+1$, the value of S can only change by at most 1. Each possible transition occurs with equal probability. As examples: if $S_t = 1$, then $S_{t+1} \in \{1, 2\}$ each with probability $1/2$. If $S_t = 2$, then $S_{t+1} \in \{1, 2, 3\}$ each with probability $1/3$.

Let $E_t \in \{+e, -e\}$ represent the feedback we receive from student evaluations for the semester t, where $+e$ is generally positive feedback and $-e$ is negative feedback. Student feedback is dependent on whether released assignments were helpful for the given semester ($A_t \in \{+a, -a\}$) which is dependent on the quality of the semester's course offering (S_t). Additionally, student evaluations are also dependent on events external to the class ($X_t = \{+x, -x\}$).

The following HMM depicts the described scenario:



(a) Consider the above dynamic bayes net which ends at some finite timestep t. In this problem, we are trying to approximate the most likely value of S_t given all the evidence variables up to and including t. For each of the following subparts, first decide whether the given method can be used to solve this problem. Then, if yes, select all CPTs which must be known to run the algorithm.

(i) [1 pt] Variable elimination

○ No ● Yes: ■ $P(S_0)$, $P(S_t|S_{t-1}), t > 0$ ■ $P(E_t|X_t, A_t)\forall t$ ■ $P(A_t|S_t)\forall t$ ■ $P(X_t)\forall t$

(ii) [1 pt] Value iteration

● No ○ Yes: □ $P(S_0)$, $P(S_t|S_{t-1}), t > 0$ □ $P(E_t|X_t, A_t)\forall t$ □ $P(A_t|S_t)\forall t$ □ $P(X_t)\forall t$

(iii) [1 pt] Gibbs sampling

○ No ● Yes: ■ $P(S_0)$, $P(S_t|S_{t-1}), t > 0$ ■ $P(E_t|X_t, A_t)\forall t$ ■ $P(A_t|S_t)\forall t$ ■ $P(X_t)\forall t$

(iv) [1 pt] Prior sampling

○ No ● Yes: ■ $P(S_0)$, $P(S_t|S_{t-1}), t > 0$ ■ $P(E_t|X_t, A_t)\forall t$ ■ $P(A_t|S_t)\forall t$ ■ $P(X_t)\forall t$

(v) [1 pt] Particle Filtering

○ No ● Yes: ■ $P(S_0)$, $P(S_t|S_{t-1}), t > 0$ ■ $P(E_t|X_t, A_t)\forall t$ ■ $P(A_t|S_t)\forall t$ ■ $P(X_t)\forall t$

First option is true because you can use variable elimination to solve for $P(S_t|E_{0:t})$ and then take the argmax over all possible values of S_t.
Second option is false because you are not solving for the max of S_t given the reward, actions, and previous value. Value iteration also does not consider evidence variables.
Third option is true because you can approximate the probability using gibbs sampling by sampling each variable over many iterations and using the final probability distribution as the result.
Fourth option is true because you can sample from all the CPTs to get $P(S_t|E_{0:t})$ and then take the argmax.
Fifth option is true because you can use particle filtering to approximate HMMs and determine the most likely S_t from the value

with the most particles.

All methods which are valid will require using all of the CPTs in order to calculate the probability since the CPTs of $E_{0:t}$ and S_t depend on all the other variables in the DBN.

(b) For the HMM shown above, determine the correct recursive formula for the belief distribution update from $B(S_{t-1})$ to $B(S_t)$. Recall that the belief distribution $B(S_t)$ represents the probability $P(S_t|E_{0:t})$ and involves two steps: **(i)** Time elapse and **(ii)** Observation update.

$$B(S_t) \propto \underline{\quad \textbf{(ii)} \quad} \cdot \underline{\quad \textbf{(i)} \quad}$$

(i) [1 pt] Time elapse

- ● $\sum_{S_{t-1}} P(S_t|S_{t-1})B(S_{t-1})$
- ○ $\sum_{S_{t-1}} \sum_{A_{t-1}} P(S_t|S_{t-1})P(A_{t-1}|S_{t-1})B(S_{t-1})$
- ○ $\sum_{S_{t-1}} P(S_t|S_{t-1})P(A_t|S_t)B(S_{t-1})$
- ○ $\sum_{S_{t-1}} \sum_{A_{t-1}} P(S_t|S_{t-1})P(A_{t-1}|S_{t-1})P(A_t|S_t)B(S_{t-1})$

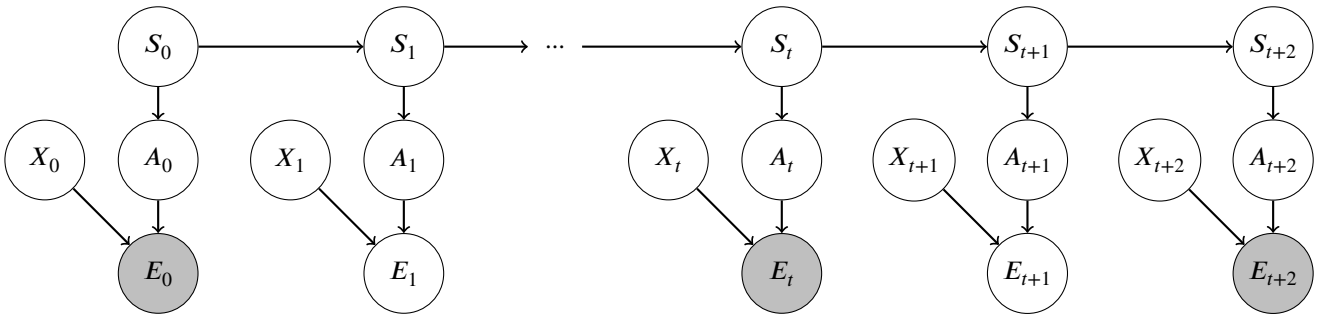
This is the standard time elapse update (see 8.2 of the textbook). Note that the extra variables do not affect that transition from S_t to its next state S_{t+1} so the time elapse expression is the same.

(ii) [1 pt] Observation update

- ○ $P(E_t|X_t, A_t)$
- ○ $P(E_t|X_t, A_t)P(X_t)P(A_t|S_t)$
- ○ $\sum_{x\in X_t} \sum_{a\in A_t} P(E_t|x,a)$
- ● $\sum_{x\in X_t} \sum_{a\in A_t} P(E_t|x,a)P(x)P(a|S_t)$
- ○ $\prod_{x\in X_t} \prod_{a\in A_t} P(E_t|x,a)$
- ○ $\prod_{x\in X_t} \prod_{a\in A_t} P(E_t|x,a)P(x)P(a|S_t)$

The observation update is the probability of evidence (E_t) given the state (S_t). Since there are extra hidden variables X_t and A_t at each timestep, these need to be summed out and eliminated to get the correct expression for $P(E_t|S_t)$ at each timestep.

Due to the differences between CS188 offerings in the fall and spring semesters, we realize that only student evaluations from past fall semesters are accurate enough to be incorporated into our model. Assume that a fall semester occurs during an even timestep and that t is even in the diagram below. The new HMM can be repesented as follows:



(c) [2 pts] In this question, we are trying to derive a recursive formula for the two-step belief distribution update from $B(S_t)$ to $B(S_{t+2})$ for the new problem described above. Which of the following steps represent the **correct and most efficient** method of performing HMM updates to get the belief distribution at S_{t+2} from the current belief at S_t?
For the following notation, let $B(S_t) = P(S_t|E_{0:t:2})$ and $B'(S_t) = P(S_t|E_{0:t-1:2})$ where $E_{0:i:2}$ represents the set of all evidence variables at even timesteps up to i. Further, let $O(E_t)$ represent the value of the observation update expression from the previous part **(ii)**. (Note that $O(E_{t+1})$ and $O(E_{t+2})$ would represent the appropriate observation update expressions for timestep $t+1$ and $t+2$ respectively.)

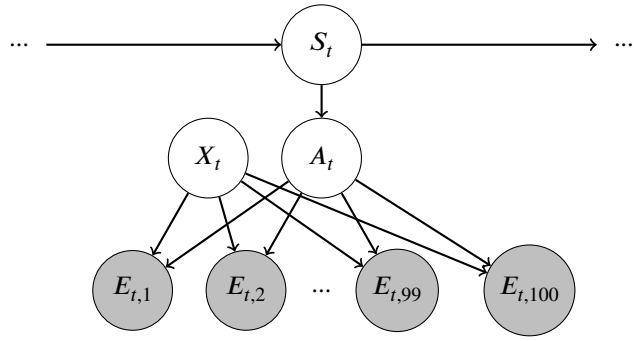
- ● $B'(S_{t+1}) = \sum_{S_t} P(S_{t+1}|S_t)B(S_t)$
  $B'(S_{t+2}) = \sum_{S_{t+1}} P(S_{t+2}|S_{t+1})B'(S_{t+1})$
  $B(S_{t+2}) \propto O(E_{t+2})B'(S_{t+2})$
- ○ $B'(S_{t+1}) = \sum_{S_t} P(S_{t+1}|S_t)B(S_t)$
  $B(S_{t+1}) = \sum_{E_{t+1}} O(E_{t+1})B'(S_{t+1})$
  $B'(S_{t+2}) = \sum_{S_{t+1}} P(S_{t+2}|S_{t+1})B(S_{t+1})$
  $B(S_{t+2}) \propto O(E_{t+2})B'(S_{t+2})$
- ○ $B'(S_{t+2}) = \sum_{S_{t+1}} \sum_{S_t} P(S_{t+2}|S_{t+1})P(S_{t+1}|S_t)B(S_t)$
  $B(S_{t+2}) \propto O(E_{t+2})B'(S_{t+2})$
- ○ $B'(S_{t+1}) = \sum_{S_t} P(S_{t+1}|S_t)B(S_t)$
  $B(S_{t+1}) \propto O(E_{t+1})B'(S_{t+1})$

$B'(S_{t+2}) = \sum_{S_{t+1}} P(S_{t+2}|S_{t+1})B(S_{t+1})$
$B(S_{t+2}) \propto O(E_{t+2})B'(S_{t+2})$

◯ None of the above.

Top left: $B'(S_{t+1})$ is a normal time elapse update. Since there is no evidence at timestep $t+1$, we do not perform an observation update and directly use the belief from $B'(S_{t+1})$ to calculate the new belief at $B'(S_{t+2})$. Then given evidence at timestep $t+2$, we update our belief at $B'(S_{t+2})$ with evidence to get $B(S_{t+2})$ in a normal observation update step.
Top right: This is the same as the first choice except the two time elapse steps are done in one equation to directly get $B'(S_{t+2})$. This is less efficient because you have to recompute $\sum_{S_t} P(S_{t+1}|S_t)B(S_t)$ at every timestep.
Bottom left: This is also equivalent to the first choice because $B(S_{t+2}) = B'(S_{t+2})$. Since we are summing over all possible values of evidence at E_{t+1}, this term becomes 1 and we are only left with $B'(S_{t+2})$. This is inefficient because extra calculation is done to sum out the unknown variables at odd timesteps.
Bottom right: This is the normal HMM update process, but since we don't have evidence for E_{t+1}, it doesn't make sense to incorporate evidence to get $B(S_{t+1})$ since the value of E_{t+1} is unknown.

(d) Now consider a scenario where instead of getting one general student feedback as evidence (E_t), we instead get individual student feedback from 100 students. Let the variable $E_{t,n}$ represent the evidence from student n at timestep t. Assume that the new evidence variables ($E_{t,n}$) can each take on the value of $+e$ or $-e$ with the same probability distribution as the single variable case (E_t).



(i) [2 pts] Which of the following statements are true regarding this new setup?

☐ The evidence variables **within the same timestep** are independent of each other ($E_{t,j} \perp\!\!\!\perp E_{t,k}\ \forall j \neq k$).

☐ The evidence variables **between any two different timesteps** are independent of each other ($E_{t_1,j} \perp\!\!\!\perp E_{t_2,k}\ \forall t_1 \neq t_2$).

■ The expression to calculate the time elapse step from S_t to S_{t+1} for this new setup will be the same as the time elapse expression in the case of one evidence from Q1(b)(i).

○ None of the above.

The first option is false because the variables are dependent through any connecting path (E-X-E, E-A-E, etc.).
The second option is false because evidence variables are still connected in an active path by common cause.
The third option is true because the change in evidence variables will only affect the observation update step.

(ii) [1 pt] In the observation update at timestep t, we receive as evidence 60 positive evaluations ($+e$) and 40 negative evaluations ($-e$). Let x be the observation update probability at timestep t of observing one positive evaluation ($x = O(E_t = +e)$). Now, let $f(x)$ represent the new observation update expression for the case with the observed 100 evidence variables. Which of the following functions f gives the correct observation update for the new scenario? For this part only, regardless of your previous answer, please assume that each students' feedback is **independent** of each other.

○ $f(x) = x^{100}$

○ $f(x) = 60x \cdot 40(1-x)$

● $f(x) = x^{60} \cdot (1-x)^{40}$

○ None of the above.

The new observation update would be $P(E_1, ..., E_{100}|S)$. Since the evidence is independent, $P(E_1, ..., E_{100}|S) = \prod_t P(E_t|S) = (P(+e|S))^{60} \cdot (P(-e|S))^{40}$. Replacing $P(+e|S)$ with x since that was the probability solved before for one evidence variable, we would get $x^{60} \cdot (1-x)^{40}$.

Q1 Self-Assessment - leave this section blank for your original submission. We will release the solutions to this problem after the deadline for this assignment has passed. After reviewing the solutions for this problem, assess your initial response by checking one of the following options:

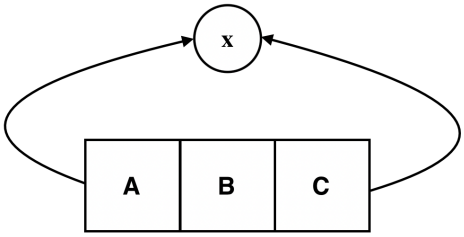
- ◯ I fully solved the problem correctly, including fully correct logic and sufficient work (if applicable).
- ◯ I got part or all of the question incorrect.

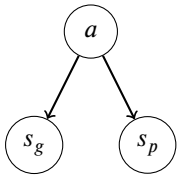
If you selected the second option, explain the mistake(s) you made and why your initial reasoning was incorrect (do not re-iterate the solution. Instead, reflect on the errors in your original submission). Approximately 2-3 sentences for *each* incorrect sub-question.

Q2. [5 pts] Learning to Act

In lecture and discussion, we have mainly used the Naive Bayes algorithm to do binary classification, such as classifying whether an email is spam. However, we can also use Naive Bayes to learn how to act in an environment. This problem will explore learning good policies with Naive Bayes and comparing them to policies learned with RL.

We consider the following one-dimensional grid world environment with three squares, named A, B, and C from left to right. Pacman has two possible actions at each square: left and right. Taking the left action at square A or the right action at square C will transition to a terminal state x where no further action can be taken. At each timestep, Pacman observes his own position (s_p) as well as the ghost's position (s_g), and he uses these observations to decide on an action.





(a) In this part, Pacman has no idea about the transition probabilities of the ghost or the rewards it gets. However, Pacman has access to an expert demonstration dataset, which gives reasonably good actions to take in a number of scenarios. The dataset is divided into **training**, **validation**, and **test** datasets. The following is the **training set** of the dataset.

s_p	s_g	a
B	C	left
C	A	left
A	B	left
C	C	right
B	A	right
C	B	right

(i) [1 pt] Using the standard Naive Bayes algorithm, what are the maximum likelihood estimates for the following conditional probabilities (encoded in the Bayes Net)?

$P(s_p = C \mid a = \text{left}) =$ 1/3

There are three left actions. Of these three, one entry has $s_p = C$.

$P(s_p = A \mid a = \text{right}) =$ 0

There are three right actions and no entry among these where $s_p = A$ so the probability is 0.

$P(a = \text{left}) =$ 0.5

There are 3 left actions out of 6 actions total.

(ii) [2 pts] Using the standard Naive Bayes algorithm, which action should we choose in the following new scenarios?

$s_p = A, s_g = C$ ● Left ○ Right

Left: $P(s_p = A, s_g = C|a = left) = P(s_p = A|a = left) \cdot P(s_g = C|a = left) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

Right: $P(s_p = A, s_g = C|a = right) = P(s_p = A|a = right) \cdot P(s_g = C|a = right) = 0 \cdot \frac{1}{3} = 0$

Left action has a higher probability so agent would choose left.

$s_p = C, s_g = B$ ○ Left ● Right

Left: $P(s_p = C, s_g = B|a = left) = P(s_p = C|a = left) \cdot P(s_g = B|a = left) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

Right: $P(s_p = C, s_g = B|a = right) = P(s_p = C|a = right) \cdot P(s_g = B|a = right) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$

Right action has a higher probability so agent would choose right.

(iii) [2 pts] Suppose we want to add Laplace smoothing with strength k in the Naive Bayes algorithm. (There is no smoothing when $k = 0$.) Which of the following are true?

☐ To find the optimal value of k, we pick the value of k which gives the highest accuracy on the training set.

■ To find the optimal value of k, we pick the value of k which gives the highest accuracy on the validation set.

☐ To find the optimal value of k, we pick the value of k which gives the highest accuracy on the test set.

■ If $k = 0$, we may observe low accuracy on the test set due to overfitting.

■ If k is a very large integer, the posterior probability for each action will be close to 0.5.

○ None of the above.

k is a hyperparameter and we should choose its optimal value based on performance on the validation set. Note that we cannot test hyperparameters on the test set and k is already being trained on the training set.

$k = 0$ means that there is no laplace smoothing so it is possible to overfit on the training data.

As k approaches ∞, the probability will approach uniform over all the possible actions which would be $1/2$.

Q2 Self-Assessment - leave this section blank for your original submission. We will release the solutions to this problem after the deadline for this assignment has passed. After reviewing the solutions for this problem, assess your initial response by checking one of the following options:

○ I fully solved the problem correctly, including fully correct logic and sufficient work (if applicable).

○ I got part or all of the question incorrect.

If you selected the second option, explain the mistake(s) you made and why your initial reasoning was incorrect (do not re-iterate the solution. Instead, reflect on the errors in your original submission). Approximately 2-3 sentences for *each* incorrect sub-question.