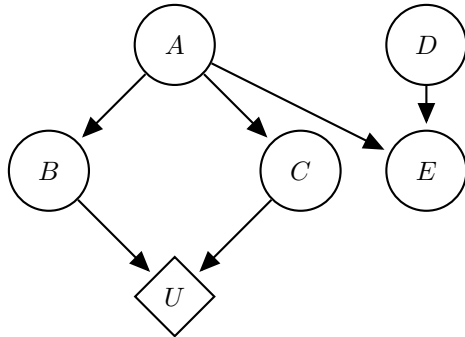# CS188: Exam Practice Session 9 Solutions

Q1. VPI

(a) Consider a decision network with the following structure, where node U is the utility:

(i) For each of the following, choose the most specific option that is guaranteed to be true:

○ $VPI(B) = 0$	● $VPI(B) \geq 0$	○ $VPI(B) > 0$
● $VPI(D) = 0$	○ $VPI(D) \geq 0$	○ $VPI(D) > 0$
○ $VPI(E) = 0$	● $VPI(E) \geq 0$	○ $VPI(E) > 0$
○ $VPI(A|E) = 0$	● $VPI(A|E) \geq 0$	○ $VPI(A|E) > 0$
● $VPI(E|A) = 0$	○ $VPI(E|A) \geq 0$	○ $VPI(E|A) > 0$
● $VPI(A|B,C) = 0$	○ $VPI(A|B,C) \geq 0$	○ $VPI(A|B,C) > 0$

Any node which is d-separated from the parents of the utility node is guaranteed to have 0 VPI. We always have the guarantee that VPI ≥ 0, but since here we have no assumptions about the utility function, we could have $U(B,C) = 0$, in which case MEU will always be 0 regardless of the information we have and thus VPI is 0.

(ii) For each of the following, fill in the blank with the most specific of $>, \geq, <, \leq, =$ to guarantee that the comparison is true, or write ? if there is no possible guarantee.

$VPI(B)$ ____?____ $VPI(A)$
If U depends only on B, then $VPI(B) > VPI(A)$. If U depends only on C, then $VPI(A) > VPI(B)$. There exist scenarios in which both inequalities are true, so we can't guarantee either.

$VPI(B,C)$ ____$\geq$____ $VPI(A)$
B and C together are more informative because the only influence A has on the utility goes through those nodes. However, since we could again have $U = 0$, the inequality is not strict.

$VPI(B,C)$ ____?____ $VPI(B) + VPI(C)$
If B and C are two coin flips and the utility function is whether we can guess their XOR, then VPI of either of them individually is 0, but their joint VPI is positive. On the other hand, if B and C are deterministically equivalent, (for example, if the conditional distribution of both is such that they are both identical to A and thus to each other), then $VPI(B) = VPI(C) = VPI(B,C)$, so $VPI(B) + VPI(C) = 2 \cdot VPI(B,C) > VPI(B,C)$.

$VPI(B|C)$ ____$\geq$____ $VPI(A|C)$
Conditioned on C, A only influences U through B, so B gives at least as much information.

1

Q2. HMMs - Forward Algorithm

Below is the forward algorithm update equation for Hidden Markov Models. As seen in lecture, we used $e_{1:t}$ to denote all the evidence variables $e_1, e_2, \ldots, e_t$. Similarly, $e_{1:t-1}$ denotes $e_1, e_2, \ldots, e_{t-1}$. For reference, the Bayes net corresponding to the usual Hidden Markov Model is shown on the right side of the equation below.
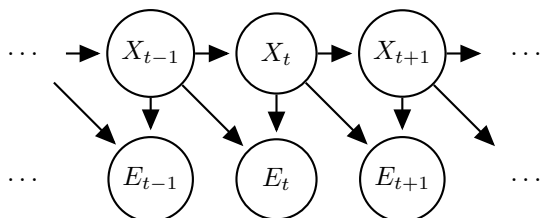
$$P(x_t|e_{1:t}) \propto P(x_t, e_{1:t}) \tag{1}$$
$$= P(e_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1}, e_{1:t-1}) \tag{2}$$
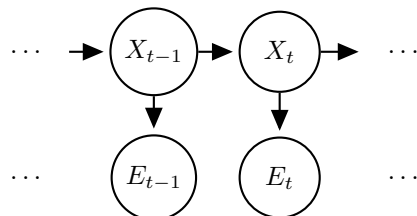
Hidden Markov Models can be extended in a number of ways to incorporate additional relations. Since the independence assumptions are different in these extended Hidden Markov Models, the forward algorithm updates will also be different.
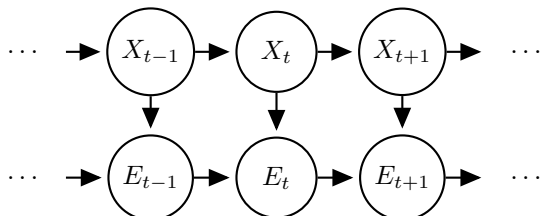
Complete the forward algorithm updates for the extended HMMs specified by the following Bayes nets:

(a) $P(x_t|e_{1:t}) \propto \sum_{x_{t-1}} P(x_{t-1}, e_{1:t-1}) \cdot \underline{P(e_t \mid x_t, x_{t-1}) P(x_t \mid x_{t-1})}$
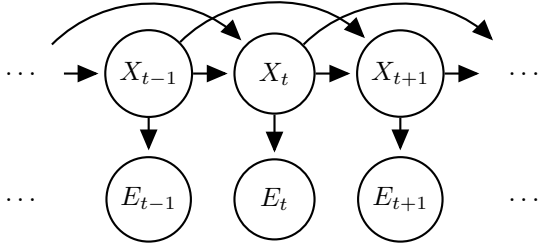
$$P(x_t|e_{1:t}) \propto P(x_t, e_{1:t}) = \sum_{x_{t-1}} P(x_{t-1}, x_t, e_{1:t-1}, e_t)$$
$$= \sum_{x_{t-1}} P(e_t \mid x_{t-1}, x_t, e_{1:t-1}) P(x_{t-1}, x_t, e_{1:t-1})$$
$$= \sum_{x_{t-1}} P(e_t \mid x_t, x_{t-1}) P(x_{t-1}, x_t, e_{1:t-1})$$
$$= \sum_{x_{t-1}} P(e_t \mid x_t, x_{t-1}) P(x_t \mid x_{t-1}, e_{1:t-1}) P(x_{t-1}, e_{1:t-1})$$
$$= \sum_{x_{t-1}} P(e_t \mid x_t, x_{t-1}) P(x_t \mid x_{t-1}) P(x_{t-1}, e_{1:t-1})$$

(b) $P(x_t|e_{1:t}) \propto \sum_{x_{t-1}} P(x_{t-1}, e_{1:t-1}) \cdot \underline{P(e_t \mid x_t, e_{t-1}) P(x_t \mid x_{t-1})}$

$$P(x_t|e_{1:t}) \propto P(x_t, e_{1:t}) = \sum_{x_{t-1}} P(x_{t-1}, x_t, e_{1:t-1}, e_t)$$

$$= \sum_{x_{t-1}} P(e_t \mid x_{t-1}, x_t, e_{1:t-1}) P(x_{t-1}, x_t, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t, e_{t-1}) P(x_{t-1}, x_t, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t, e_{t-1}) P(x_t \mid x_{t-1}, e_{1:t-1}) P(x_{t-1}, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t, e_{t-1}) P(x_t \mid x_{t-1}) P(x_{t-1}, e_{1:t-1})$$

(c) $P(x_t, x_{t+1}|e_{1:t}) \propto \sum_{x_{t-1}} P(x_{t-1}, x_t, e_{1:t-1}) \cdot \underline{P(e_t \mid x_t) P(x_{t+1} \mid x_{t-1}, x_t)}$


$$P(x_t, x_{t+1}|e_{1:t}) \propto P(x_t, x_{t+1}, e_{1:t}) = \sum_{x_{t-1}} P(x_{t-1}, x_t, x_{t+1}, e_{1:t-1}, e_t)$$

$$= \sum_{x_{t-1}} P(e_t \mid x_{t-1}, x_t, x_{t+1}, e_{1:t-1}) P(x_{t-1}, x_t, x_{t+1}, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t) P(x_{t-1}, x_t, x_{t+1}, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t) P(x_{t+1} \mid x_t, x_{t-1}, e_{1:t-1}) P(x_t, x_{t-1}, e_{1:t-1})$$

$$= \sum_{x_{t-1}} P(e_t \mid x_t) P(x_{t+1} \mid x_t, x_{t-1}) P(x_t, x_{t-1}, e_{1:t-1})$$

Q3. DNA Sequencing

Suppose you want to model the problem of DNA sequencing using the following set-up:

- X_i, $Y_i \in \{A, T, C, G\}$

- X_i : ith base of an individual
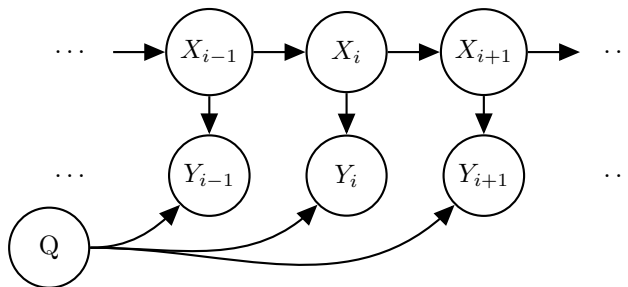
- Y_i : ith base output by DNA sequencer

(a) First, you start by using a standard HMM model, shown below.
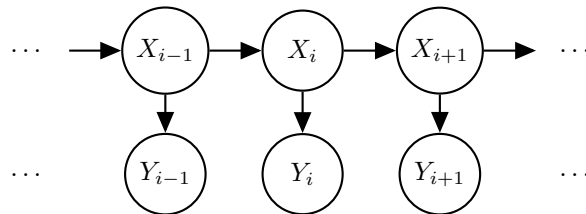
    (i) Which of the following assumptions are made by the above HMM model

        ☐ $X_i \perp\!\!\!\perp X_j \;\; \forall \, i \neq j$
                                        ■ $X_i \perp\!\!\!\perp Y_{i+1} \mid X_{i+1} \;\; \forall \, i$

        ☐ $Y_i \perp\!\!\!\perp Y_j \;\; \forall \, i \neq j$
                                          ☐   None of the provided options.

        ☐ $X_i \perp\!\!\!\perp Y_j \;\; \forall \, i \neq j$

        ■ $X_{i-1} \perp\!\!\!\perp X_{i+1} \mid X_i \;\; \forall \, i$

(b) Now you want to model the quality of your sequencer with a random variable Q, and decide to use the following modified HMM:

    (i) Which of the following assumptions are made by the above modified HMM model?

        ☐ $X_i \perp\!\!\!\perp X_j \;\; \forall \, i \neq j$
                                          ■ $Q \perp\!\!\!\perp X_i \;\; \forall \, i$

        ☐ $Y_i \perp\!\!\!\perp Y_j \;\; \forall \, i \neq j$
                                          ☐ $Q \perp\!\!\!\perp X_i \mid Y_i \;\; \forall \, i$

        ☐ $X_i \perp\!\!\!\perp Y_j \;\; \forall \, i \neq j$
                                          ☐ $Q \perp\!\!\!\perp X_i \mid Y_1, ... Y_N \;\; \forall \, i$

        ■ $X_{i-1} \perp\!\!\!\perp X_{i+1} \mid X_i \;\; \forall \, i$
                                          ☐   None of the provided options.

        ■ $X_i \perp\!\!\!\perp Y_{i+1} \mid X_{i+1} \;\; \forall \, i$

    (ii) You observe the sequencer output $y_1, \dots, y_N$ and want to estimate probability distribution of the particular sequence of length c starting at base k: $P(X_k \dots X_{k+c-1} \mid y_1, \dots y_N)$.
         Select all elimination orderings which are maximally efficient with respect to the sum of the generated factors' sizes.

        ☐ $X_1, \dots, X_{k-1}, X_{k+c}, \dots, X_N, Q$
                                   ☐ $X_1, \dots, X_{k-1}, Q, X_{k+c}, \dots, X_N$

        ■ $X_1, \dots, X_{k-1}, X_N, \dots, X_{k+c}, Q$
                                   ☐ $X_1, \dots, X_{k-1}, Q, X_N, \dots, X_{k+c}$

        ☐ $Q, X_1, \dots, X_{k-1}, X_{k+c}, \dots, X_N$
                                   ☐   None of the provided options: _____

        ☐ $Q, X_1, \dots, X_{k-1}, X_N, \dots, X_{k+c}$

    The most efficient ordering above eliminates Q last because eliminating Q generates a factor that includes all of the remaining variables. It is more efficient to eliminate from the outside, in because they generate factors with only 1 variable rather than 2.

    (iii) How many entries are in the final conditional probability table $P(X_k, \dots, X_{k+c-1} \mid y_1, \dots, y_N)$? The answer takes the form a^b – what are a and b?

$a =$ _____4_____ $b =$ _____c_____

Each X_i can take on 4 values and there are c variables (the Y's are fixed). Therefore, there are 4^c entries in the CPT.