CS 188 Fall 2006 — Introduction to Artificial Intelligence — Midterm 2

You have 80 minutes. The exam is closed book, one page cheat sheet allowed, non-programmable basic calculators allowed. 60 points total. Pace yourself, and don't panic!

Mark your answers ON THE EXAM ITSELF. Write your name, SID, login, and section number at the top of each page.

For multiple-choice questions, CIRCLE **ALL** CORRECT CHOICES (in some cases, there may be more than one).

If you are not sure of your answer you may wish to provide a *brief* explanation. All short answers can be successfully answered in a few sentences *at most*.

For staff use only

Q. 1	Q. 2	Q. 3	Q. 4	Total
/9	/20	/17	/14	/60

1. (9 points.) Probabilities

(a) **(6 pts)** For each statement about distributions over X, Y, and Z, if the statement is not always true, state a conditional independence assumption which makes it true. If the statement is always true for any distribution, simply state so.

i. $P(x|y) = P(x,y)/P(y)$ Always true

ii. $P(x,y) = P(x)P(y)$ True for $x \perp\!\!\!\perp y$

iii. $P(x,y,z) = P(x|z)P(y|z)P(z)$ True for $x \perp\!\!\!\perp y|z$

iv. $P(x,y,z) = P(x)P(y)P(z|x,y)$ True for $x \perp\!\!\!\perp y$

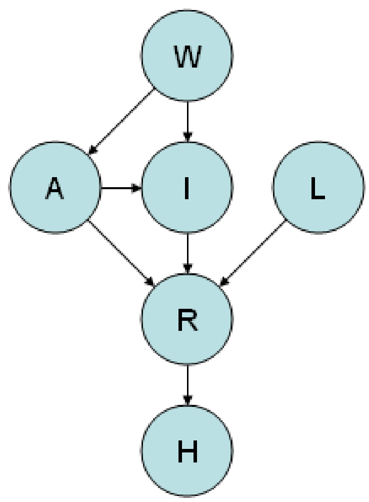
v. $P(x,y) = \sum_z P(x,y,z)$ Always true

(b) **(3 pts)** When designing a Bayes' net, why do we not make every variable depend on as many other variables as possible?

Adding dependencies increases the complexity of the model which makes inference more expensive. Estimation of parameters also requires more data.

2. (20 points.) Graphical Models
You are considering founding a startup to make AI based robots to do household chores, and you want to reason about your future. There are three ways you can possibly get rich (R), either your company can go public via an IPO (I), it can be acquired (A), or you can win the lottery (L). Your company cannot get acquired if it goes public. Of course, in order for your company to either go public or get acquired, your robot has to actually work (W). You decide that if you do strike it rich then you will probably retire to Hawaii (H) to live the good life.

(a) **(5 pts)** Draw a graphical model for this problem which reflects the causal structure of the problem as stated.



(b) **(5 pts)** Which of the following independence properties are true for your network?

- $A \perp\!\!\!\perp I$
- $L \perp\!\!\!\perp I$ True
- $L \perp\!\!\!\perp I | R$
- $L \perp\!\!\!\perp W | H$
- $W \perp\!\!\!\perp H | L$
- $W \perp\!\!\!\perp H | R$ True

(c) **(2 pts)** Write out an expression for an entry $P(a, h, i, \ell, r, w)$, of the joint distribution encoded by your network, $P(A, H, I, L, R, W)$ in terms of quantities provided by the network.

$P(a, h, i, l, r, w) = P(w)P(a|w)P(i|w, a)P(r|a, i, l)P(h|r)$

(d) **(4 pts)** Are there any nodes in your network whose conditional probability table (CPT) specifications *cannot* affect the value of $P(I|L)$? That is, were their CPTs to change, the value of $P(I|L)$ would not change. If so, list all such nodes.

H, L, R

Imagine you choose to compute $P(h|w)$, the probability that you will end up in Hawaii if the robot does work, by sampling with likelihood weighting. You draw the following samples from your network.

- $\langle \neg a, h, i, \neg \ell, r, w \rangle$
- $\langle \neg a, \neg h, \neg i, \neg \ell, \neg r, w \rangle$
- $\langle a, \neg h, \neg i, \neg \ell, \neg r, w \rangle$

(e) **(1 pts)** What weights do these samples have, in terms of conditional probabilities given by your network?

$P(w)$

(f) **(2 pts)** What is your sample-based approximation to $P(h|w)$?

All samples are weighted equally (despite having weights less than 1).
$\hat{P}(h|w) = \frac{1}{3}$

(g) **(1 pt)** What would your sample-based approximation to $P(h|w)$ have been if these samples had been generated by rejection sampling?

Samples would have weights of 1.
$\hat{P}(h|w) = \frac{1}{3}$

3. (17 points.) HMMs

Every day, your pet cat Markov is either in a good or bad mood, with one mood transition between days. Your cat's behavior on a day depends on its mood, but the true mood is always a bit of a mystery. On day one, Markov is in a good mood. The cat mood process can be modeled using an HMM with the following parameters:

$P(M_1)$

good	1
bad	0

$P(M_t \mid M_{t-1} = good)$

good	3/4
bad	1/4

$P(M_t \mid M_{t-1} = bad)$

good	1/4
bad	3/4

Transitions

$P(B_t \mid M_t = good)$

hiss	0/5
meow	3/5
purr	2/5

$P(B_t \mid M_t = bad)$

hiss	3/5
meow	1/5
purr	1/5

Emissions

Note that since you know that Markov is in a good mood on day one, the value of her behavior on day one (B_1) is irrelevant to all calculations below, and you can prune B_1 from the network.

(a) **(1 pts)** What is $P(M_2 = \text{good})$?

$P(M_2 = \text{good}) = P(M_2 = \text{good}|M_1 = \text{good}) = \frac{3}{4}$

(b) **(3 pts)** What is $P(B_2 = \text{meow})$?

$$\begin{aligned} P(B_2 = \text{meow}) &= P(B_2 = \text{meow}|M_2 = \text{good})P(M_2 = \text{good}) + P(B_2 = \text{meow}|M_2 = \text{bad})P(M_2 = \text{bad}) \\ &= (\frac{3}{5})(\frac{3}{4}) + (\frac{1}{5})(\frac{1}{4}) \\ &= \frac{1}{2} \end{aligned}$$

(c) **(3 pts)** What is $P(M_2 = \text{good}|B_2 = \text{meow})$

$$\begin{aligned} P(M_2 = \text{good}|B_2 = \text{meow}) &= P(B_2 = \text{meow}|M_2 = \text{good})P(M_2 = \text{good})/P(B_2 = \text{meow}) \\ &= (\frac{3}{5})(\frac{3}{4})/(\frac{1}{2}) \\ &= \frac{9}{10} \end{aligned}$$

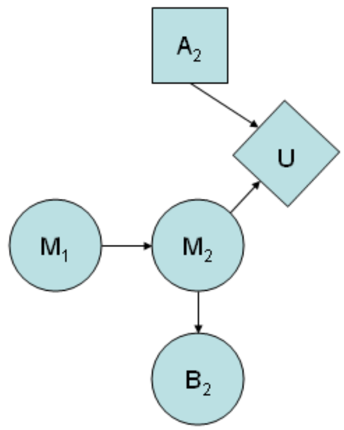
(d) **(3 pts)** What does $P(B_T = \text{meow})$ approach for very large T?

As T gets large, M_T tends towards the uniform distribution.

$$\begin{aligned} P(B_T = \text{meow}) &= P(B_T = \text{meow}|M_T = \text{good})P(M_T = \text{good}) + P(B_T = \text{meow}|M_T = \text{bad})P(M_T = \text{bad}) \\ &= (\frac{3}{5})(\frac{1}{2}) + (\frac{1}{5})(\frac{1}{2}) \\ &= \frac{2}{5} \end{aligned}$$

On any day, you have two actions in A_t available to you, *pick-up* or *leave-alone*. If you try to pick Markov up on a bad mood day, you are guaranteed to be scratched, while on a good mood day you are guaranteed a hug. Imagine you have a utility of $U_h = 2$ for hugs, $U_s = -1$ for scratches, and $U_n = 0$ for leaving the cat alone. Consider the problem of deciding whether or not to pick up your cat on day two.

(e) **(2 pt)** Draw a decision diagram over the variables M_1, M_2, and B_2, including nodes for the action A_2 and the resulting utility U.



(f) **(2 pts)** What is the expected utility of picking up the cat on day two given no observations of behavior?

$$\begin{aligned} U(M_2, A_2 = pick-up) &= P(M_2 = good)U(hug) + P(M_2 = bad)U(scratch) \\ &= (\frac{3}{4})(2) + (\frac{1}{4})(-1) \\ &= \frac{5}{4} \end{aligned}$$

(g) **(2 pts)** What is the expected utility of picking up the cat on day two given $B_2 = meow$?

$$\begin{aligned} U(M_2, A_2 = pick-up | B_2 = meow) &= P(M_2 = good | B_2 = meow)U(hug) + P(M_2 = bad | B_2 = meow)U(scratch) \\ &= (\frac{9}{10})(2) + (\frac{1}{10})(-1) \\ &= \frac{17}{10} \end{aligned}$$

(h) **(1 pts)** What is the maximum expected utility (MEU) for your cat-picking-up decision on day 2, given no observations of behavior?

$MEU = max_{A_2}(EU(pick-up), EU(\text{do nothing})) = \frac{5}{4}$

4. (14 points.) Short Answer

(a) **(6 pts)** Let $MEU(e)$ be the maximum expected utility over available actions A given evidence $E = e$, where utility depends on some variable S and the action $a \in A$:

$$MEU(e) = \max_a \sum_s P(s|e)U(s,a)$$

Assume $a(e)$ is a maximizing action. Consider the value of revealing another variable E'. Let $MEU(e, e')$ be the maximum expected utility given the evidence $E = e$ and $E' = e'$, defined analogously, with a maximizing action $a(e, e')$. Let $VPI(E'|e)$ be the value of perfect information about E' given the current evidence $E = e$. Circle all of the following which are true:

i. $\forall e', MEU(e) = MEU(e, e')$ False

ii. $\forall e', MEU(e) \leq MEU(e, e')$ False: e' might reveal that things aren't as good as you thought when you only knew about e.

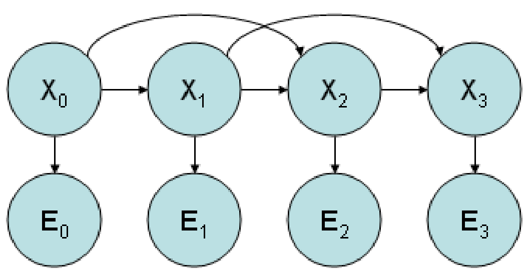
iii. $MEU(e) \leq \sum_{e'} P(e'|e)MEU(e, e')$ True

iv. $VPI(E'|e) = (\sum_{e'} P(e'|e)MEU(e, e')) - MEU(e)$ True

v. $(\forall e', a(e, e') = a(e)) \Rightarrow (VPI(E'|e) > 0)$ False

vi. $(\forall e', a(e, e') = a(e)) \Rightarrow (VPI(E'|e) = 0)$ True

(b) **(4 pts)** In a second-order HMM, each X_t depends on the previous *two* states, X_{t-1} and X_{t-2}, and we have that $X_t \perp\!\!\!\perp X_i \mid \{X_{t-1}, X_{t-2}\}$ for all $i < t - 2$. Moreover, each observation still depends only on the current state, so $E_t \perp\!\!\!\perp V \mid \{X_t\}$ for all other variables V. Draw the second-order HMM as a graphical model over the variables $\{X_0, X_1, X_2, X_3, E_0, E_1, E_2, E_3\}$. For time 0 and 1, assume that X_t depends only on the previous zero and one states, respectively.



(c) **(4 pts)** Write the second-order forward recurrence, which expresses $P(x_t, x_{t-1}, e_{1:t})$ in terms of $P(x_{t-1}, x_{t-2}, e_{1:t-1})$ and other second-order HMM local conditional probabilities.

$$\begin{aligned} P(x_t, x_{t-1}, e_{1:t}) &= \sum_{x_{t-2}} P(x_t, x_{t-1}, x_{t-2}, e_{1:t}) \\ &= P(x_t|e_t) \sum_{x_{t-2}} P(x_t|x_{t-1}, x_{t-2}) P(x_{t-1}, x_{t-2}, e_{1:t-1}) \end{aligned}$$

End of Exam