[AI18] Assignment 5

Due: Nov 19 (before class)

[1] Let X be a random variable of the weather on Nov 28 2018, and Y be a random variable of the weather on Nov 30 2018. Assume $X, Y \in \{\text{Sun}, \text{Rain}\}$. Their probability distributions are given in Table 1.

Table 1. Joint Probability of Two Day's Weathers

P(X=Sun, Y=Sun)	P(X=Sun, Y=Rain)	P(X=Rain, Y=Sun)	P(X=Rain, Y=Rain)
0.2	0.4	0.1	0.3

[1.1] What is the probability that Nov 30 is sunny? **0.3**

[1.2] What is the probability that Nov 28 is rainy? **0.4**

[1.3] Are X and Y independent? Justify your answer.
X and Y are not independent because P(X and Y) is not equal to P(X)*P(Y)

[2] Fig 1 is a Markov model of weather, where x_t is the weather of the t_{th} day in 2018 and $t = 1, 2, \ldots, 365$.

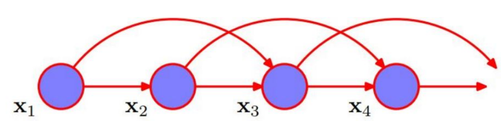

Fig. 1. A Markov Model

[2.1] The model suggests the weather of Nov 30 depends on the weather of some previous days. What are the dates of these days? **Nov 29 and 28**

[2.2] Simplify the following probability based on the given model, so that (1) we have separate probabilities of X_6 and X_7, and (2) each separated probability has as few conditional variables as possible.[1]

$$P(X_7, X_6 \mid X_5, X_4, X_3, X_2, X_1) = P(X_7 \mid X_5) * P(X_6 \mid X_5, X_4) \tag{1}$$

[2.3] Suppose someone designs the above model and have the following statistics: P(X_3=Sun | X_2=Rain, X_1=Sun)=0.5, P(X_8=Sun | X_7=Rain, X_6=Rain)=0.5 and P(X_{25}=Sun | X_{23}=Sun, X_{24}=Rain) = 0.4. Can you conclude if the designer assumes stationary or non-stationary environment? Briefly justify your answer.
The model assumes non-stationary environment because P(X_t=Sun | X_{t-1}=Rain, X_{t-2}=Sun) varies across t. For t=3 it is 0.5 and for t=25 it is 0.4

[1] For example, $P(A \mid B)$ has one conditional variable B, and $P(A \mid B, C)$ has two B and C.

[3] Answer the following questions based on the transition diagram in Fig 2.

[3.1] What are the unknown probabilities P_1 and P_2? $P_1 = 0.7$ **and** $P_2 = 0.2$

[3.2] Give the transition matrix A, where A_{ij} is the probability of transiting from state X_i to X_j.

	X1	X2	X3
X1	0.3	0.7	0
X2	0.6	0.4	0
X3	0.8	0	0.2

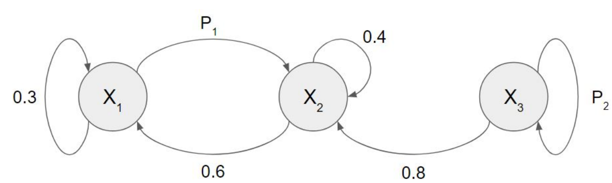

Fig. 2. A Transition Diagram

[4] Fig 3 is a hidden Markov model, and Fig 4 gives its conditional probability distributions. Assume the environment is stationary. Answer the following questions.

[4.1] What are the unknown probabilities P_1 and P_2 in Fig 4?
$P_1 = 0.9$ **and** $P_2 = 0.2$

[4.2] Compute the joint probability $P(X_1 = \text{low}, X_2 = \text{high}, Z_1 = \text{walk}, Z_2 = \text{walk-upstair})$. You cannot just give a number; you need to elaborate the arguments, including how to decompose the joint probability and what are the plugged-in values.
$P(walk) * P(walk - upstairs|walk) * P(low|walk) * P(high|walk - upstairs) = 1 * 0.2 * 0.7 * 0.9 = 0.098$

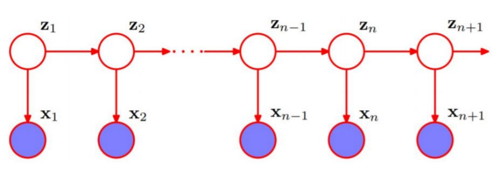

Fig. 3. A Hidden Markov Model

	X=high	X=low
Z=walk	0.3	0.7
Z=walk-upstairs	P_1	0.1

	Z=walk	Z=walk-upstairs
Z=walk	0.8	P_2
Z=walk-upstairs	0.4	0.6

Fig. 4. Conditional Probabilities of the HMM in Fig 4. The left table shows the emission probability $P(X_t \mid Z_t)$, and the right table shows the transition probability $P(Z_t \mid Z_{t-1})$.

[5] Fig 5 shows a sequence of observed pairs of random variables (x, z), where x is a binary variable taking value in $\{1, 2\}$ and z is a binary variable taking value in $\{a, b\}$. Please estimate $P(z = a \mid x = 2)$ and $P(x = 1 \mid z = b)$ based on the observations.
$\mathbf{P(z = a \mid x = 2) = 2/3}$ **and** $\mathbf{P(x = 1 \mid z = b) = 1/2}$

z	a	a	b	a	b
x	1	2	1	2	2

Fig. 5. A sequence of observed random variable pairs. Each column is an observed pair (at one time stamp).

[6] Describe one application of the Markov model. In your answer, please brief describe the application (e.g., forecast weather) and the random variable (e.g., X_t is the weather of day t). You cannot give an example already covered in the lecture slides (weather, text, page ranking, stock price, speech recognition, hand-written digit, part-of-speech, activity recognition).

Markov Models are extensively used for Biological sequence analysis(Protein, DNA or RNA). If we consider protein for example multiple protein sequences can be found as part of the evolution from the same ancestor. Markov models are used to find perfect alignment between 2 sequences. The random variable in this case is a protein belonging to one family in a particular period of time.

[7] What is the technique we can apply to learn a Hidden Markov Model? (just give the name)
Expectation Maximization Algorithm

[8] Bayesian network is a graphical representation of the conditional dependency between multiple variables in a directed acyclic graph. Fig 6 is a Bayesian network. Based on it, decompose the following joint probability

$$P(A, B, C, D, E) = P(A) * P(D) * P(C|A) * P(E|A, D) * P(B|C, A) \tag{2}$$

[9] What is the D-separation criterion used for?
D-separation is used to find out if 2 variables are conditionally independent

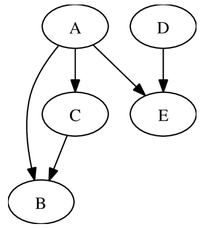

Fig. 6. A Bayesian Network

[10] We are given a list of probabilities of three binary random variables A, B and C.

$P(C = 0, A = 0 \mid B = 0) = 0.2$

$P(C = 0, A = 1 \mid B = 0) = 0.1$

$P(C = 1, A = 0 \mid B = 0) = 0.4$

$P(C = 1, A = 1 \mid B = 0) = 0.3$

$P(C = 0, A = 0 \mid B = 1) = 0.5$

$P(C = 0, A = 1 \mid B = 1) = 0.1$

$P(C = 1, A = 0 \mid B = 1) = 0.1$

$P(C = 1, A = 1 \mid B = 1) = 0.2$

$P(B = 0) = 0.4$

$P(B = 1) = 0.6$

[10.1] Infer P(A, B, C) from the above information. You need to give the probability of every event, i.e.,

$P(C = 0, A = 0, B = 0) = 0.08$

$P(C = 0, A = 1, B = 0) = 0.04$

$P(C = 1, A = 0, B = 0) = 0.16$

$P(C = 1, A = 1, B = 0) = 0.12$

$P(C = 0, A = 0, B = 1) = 0.30$

$P(C = 0, A = 1, B = 1) = 0.06$

$P(C = 1, A = 0, B = 1) = 0.06$

$P(C = 1, A = 1, B = 1) = 0.12$

[10.2] Infer probability P(B | C = 1) using the enumeration technique. You cannot just give a number; you need to elaborate the process, including

(1) which joint probabilities would you pick up from the solutions in [10.1] when fixing evidence C = 1?

Take following probabilities:

$P(C = 1, A = 0, B = 0) = 0.16$

$P(C = 1, A = 1, B = 0) = 0.12$

$P(C = 1, A = 0, B = 1) = 0.06$

$P(C = 1, A = 1, B = 1) = 0.12$

(2) how do you get the joint probability of B and C (by summing out the untargeted variable A)?

P(B=0,C=1) = 0.16+0.12=0.28
P(B=1,C=1) = 0.06+0.12 = 0.18

(3) how do you get P(B | C = 1) by normalizing the joint probability of B and C?

P(C=1) = 0.16+0.12+0.06+0.12=0.46
After normalizing:
P(B=0,C=1) = 0.28/0.46 = 0.60869565217
P(B=1,C-1) = 0.18/0.46 = 0.39130434782

[10.3] Variable enumeration is a basic inference technique but is often computationally slow. Name another inference technique that is (generally) more efficient that variable enumeration.

Sampling is more computationally faster than variable enumeration