School of Computer Science
The University of Adelaide

Introduction to Statistical Machine Learning
Assignment 3

Semester 2 2015
Due 11:55pm 30 October 2015

Instructions and submission guidelines:

- Answer all questions in a report.
- Make sure that your writing is legible and clear, and the mathematical symbols are consistent.
- You must sign an assessment declaration coversheet to submit with your assignment. The assessment declaration coversheet is included in the zip file.
- Submit your report via the MoodleForum on the course web page.

Question 1: The joint distribution for three boolean variables A, B, C is given in Table 1. Please compute the following probabilities. (20 marks, 5 each)

a	b	c	0.01
a	b	$\neg c$	0.01
a	$\neg b$	c	0.06
a	$\neg b$	$\neg c$	0.02
$\neg a$	b	c	0.04
$\neg a$	b	$\neg c$	0.04
$\neg a$	$\neg b$	c	0.80
$\neg a$	$\neg b$	$\neg c$	0.02

Table 1: $P(A, B, C)$

1. What is $P(A = a, B = b)$?
2. What is $P(B = b)$?
3. What is $P(A = a|B = b)$?
4. What is $P(A = \neg a|B = b)$?

Question 2: You have three baskets of fruit: the first one contains two apples, the second one contains two oranges, and the third one contains one apple and one orange. Assume that a basket is selected randomly and that a piece of fruit is picked randomly from that basket. Let B be the random variable corresponding to the basket number selected (B can have as value 1, 2 or 3) and let F be the random variable corresponding to the type of fruit picked (F can have as value $apple$ or $orange$). Please answer the following questions. (15 marks, 5 each)

1. What is the distribution $P(B)$ and what are the conditional distributions $P(F|B)$? Write your answers in tabular form.
2. What is the joint probability of selecting the first basket and picking an apple, i.e. what is $P(B = 1, F = apple)$?
3. If we observe that the picked fruit is an apple, what is the conditional probability that the chosen basket is the basket containing two apples, i.e. what is $P(B = 1|F = apple)$?

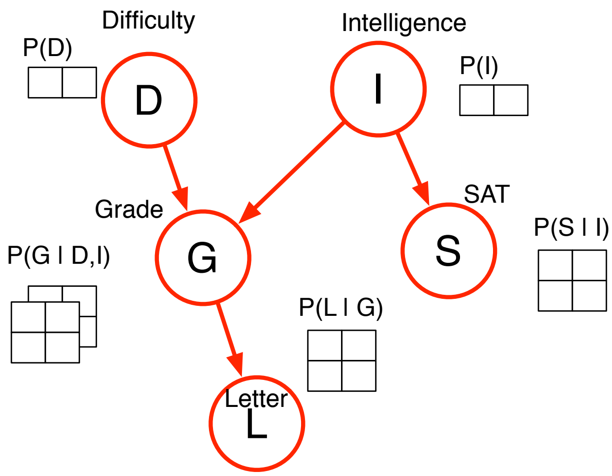

Figure 1: a Bayesian Network modelling student performance

Question 3: Given a Bayesian Network in Figure 1, please answer the following questions. (10 marks, 5 each)

1. Write down variable elimination for marginal inference to compute $P(G)$.
2. Write down the parameters.

Question 4: Let $A, B, C, D, \ldots$ be the variables. To estimate $P(A = 0|B = 0, C = 0)$, we can always set (no need to check if the denominator $= 0$ or not) the estimated one

$$\hat{P}(A = 0|B = 0, C = 0) = \frac{N_{(A=0,B=0,C=0)} + N_r}{N_{(B=0,C=0)} + (\#A) \times N_r},$$

where $N_r > 0$, and $\#A$ is the number of values of variable A can take. Please prove

$$\sum_A \hat{P}(A|B = 0, C = 0) = 1.$$

(10 marks)

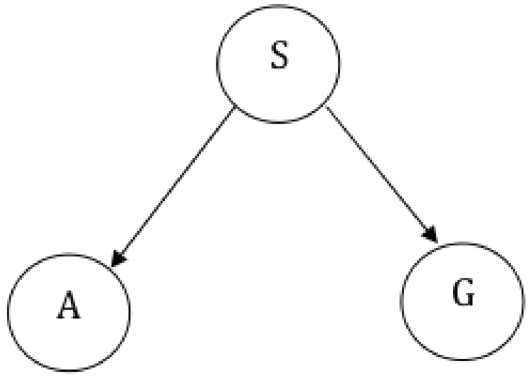

Figure 2: 4WD Bayesian network

Question 5: Please estimate $P(S), P(A|S), P(G|S)$ in Fig. 2 based on the data in Table 2 in the same way as Question 4 with $N_r = 1$. Assume all variables here take binary values 0 or 1. (10 marks)

S	A	G
1	0	0
0	0	1
1	1	0
1	1	0
0	0	0
1	0	1

Table 2: Data for 4WD model

Question 6: Independence $X \perp\!\!\!\perp Y$ means $P(X, Y) = P(X)P(Y)$, and Conditional Independence $X \perp\!\!\!\perp Y|Z$ means $P(X, Y|Z) = P(X|Z)P(Y|Z)$. Prove the following properties of independence (10 marks, 5 each):

1. Symmetry: $X \perp\!\!\!\perp Y|Z \Rightarrow Y \perp\!\!\!\perp X|Z$,
2. Decomposition: $X \perp\!\!\!\perp Y, W|Z \Rightarrow X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp W|Z$,

Question 7: Let $P(\mathbf{X})$ be the joint distribution of n discrete variables $X_1, X_2, ..., X_n$, where $\mathbf{X}$ denotes $(X_1, X_2, ..., X_n)$. Let $\mathbf{x} = (x_1, x_2, ..., x_n)$ be a realisation of $\mathbf{X}$, and $P(\mathbf{x})$ be the probability of $\mathbf{x}$. Mutual information between variables X_i and X_j are defined as

$$I(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \log \Big(\frac{P(x_i, x_j)}{P(x_i)P(x_j)} \Big).$$

For any distributions $P(\mathbf{X})$ and $P'(\mathbf{X})$, KL divergence is defined as

$$KL(P(\mathbf{X})||P'(\mathbf{X})) = \sum_{\mathbf{x}} P(\mathbf{x}) \log \frac{P(\mathbf{x})}{P'(\mathbf{x})}.$$

Please prove the following statements: (10 marks, 5 each)

1. $I(X_i, X_j) = 0$ if and only if X_i, X_j are independent.
2. $KL(P(\mathbf{X})||P'(\mathbf{X})) = 0$ if and only if $P(\mathbf{X}), P'(\mathbf{X})$ are the same.

Question 8: Please answer the following questions for kernels. (15 marks, 5 each)

1. Explain why a Kernel matrix needs to be positive semidefinite (PSD)?
2. Write down the dual form of binary support vector machines using kernel.
3. Explain advantages of Kernels.

$\sim\sim\sim$ Good luck $\sim\sim\sim$
by Javen Qinfeng Shi, 2015