Boğaziçi University, Dept. of Computer Engineering

CMPE 547, BAYESIAN STATISTICS AND MACHINE LEARNING

Fall 2015, Midterm

Name: ______________________________

Student ID: ___________________________

Signature: ____________________________

- Please print your name and student ID number and write your signature to indicate that you accept the University honour code.
- During this examination, you may not use any notes, books or calculators. You may use one A4 cheat sheet prepared with your own hand writing.
- Read each question carefully and show all your work. Underline your final answer to each question.
- There are **6** questions. Point values are given in parentheses.
- You have **180 minutes** to do all the problems.

Q	1	2	3	4	5	6	Total
Score							
Max	10	10	20	20	20	20	100

Q.1 By using Jensens equality, show that the variance of a random variable is always positive. *(10 points)*

Q.2 The KL (Kullback-Leibler) divergence is defined as

$$KL(P||Q) = \int p(x) \log p(x)/q(x)$$

Derive an expression for $KL(p||q)$ and $KL(q||p)$ in terms of V and Σ when $p(x) = \mathcal{N}(x; 0, V)$ and $q(x) = \mathcal{N}(x; 0, \Sigma)$. $\mathcal{N}(x; m, S)$ denotes a multivariate Gaussian with mean vector m and covariance matrix S.

(10 points)

Q.3 (**Bayesian Networks**) A distribution admits the following factorisation

$$p(A,B,D,F,T,L,M,X) = p(F|T,L)p(M)p(T|A)p(B|M)p(X|F)p(L|M)p(D|F,B)p(A)$$

where A,B,D,F,T,L,M,X are are discrete random variables with N states.

(a) Draw the corresponding directed graphical model.

(b) How many parameters in total need to be specified ?

(c) Verify the following conditional independence statements using d-separation. State if they are true or false and explain why.

i. $A \perp\!\!\!\perp M|\emptyset$

ii. $A \perp\!\!\!\perp M|X$

iii. $T \perp\!\!\!\perp L|X$

iv. $X \perp\!\!\!\perp L|F$

v. $X \perp\!\!\!\perp L|D$

Q.4 (**Undirected Graphical Models**) Consider the following probability model:

$$p(x) \;\propto\; 2^{x^\top W x + h^\top x}$$

where $x \in \{-1, 1\}^4$ and W and h are known parameters as given below

$$W \;=\; \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

$$h^\top \;=\; \begin{pmatrix} -1 & 1 & 0 & -1 \end{pmatrix}$$

(a) Draw the associated undirected graphical model

(b) Draw the associated factor graph and write the expression for each factor node

(c) Find the marginals $p(x_i)$ for $i = 1 \ldots 4$

(20 points)

Q.5 (AR(1) Model) Consider the following model very similar to the one discussed in detail during the lectures.

$$\begin{aligned} A &\sim \mathcal{N}(A; 0, P) \\ \Lambda &\sim \mathcal{G}(\Lambda; \nu, b) \\ x_k | x_{k-1}, A, \Lambda &\sim \mathcal{N}(x_k; A x_{k-1}, 1/\Lambda) \end{aligned}$$

where $\mathcal{N}$ is a Gaussian and

$$\mathcal{G}(x; a, b) = \exp\left((a-1)\log x - bx - \log\Gamma(a) + a \log b\right)$$

is the gamma distribution. Assume that you know the hyperparameters $\theta = (\nu, b, P)$. Suppose we observe $x_0, x_1, \dots x_K \equiv x_{0:K}$.

(a) Derive a variational Bayes algorithm to approximate

$$p(A, \Lambda | x_{0:K}, \theta)$$

with an approximating distribution of form $q(A)q(\Lambda)$ and give the update rules.

(b) Derive an EM algorithm for finding the MAP estimate

$$\Lambda^* = \operatorname*{argmax}_{\Lambda} p(\Lambda | x_{0:K}, \theta)$$

(20 points)

Q.6 **(Model Selection)** A possible generative model for clustering N data points x_i for $i = 1 \dots N$ is given as follows:

$$\begin{aligned} r_{i,1:K} &\sim \prod_{k=1}^{K} (1/K)^{r_{i,k}} \\ x_i | r_{i,1:K} &\sim \prod_{k=1}^{K} \mathcal{N}(x_i; \mu_k, 1)^{r_{i,k}} \end{aligned}$$

(a) Draw the directed graphical model for this problem.

(b) How would you determine K?

(20 points)