NAME: ______________________________ STUDENT NO.: ____________

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE AND ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Midterm Examination, April 28, 2013
Time allowed: 150 minutes

ECE521H1S — Inference Algorithms

Exam Type A: No additional notes, books or data permitted
Calculator Type 2: All non-programmable electronic calculators allowed
Examiner: Brendan J. Frey

Instructions

- Make sure you have a complete exam paper, with 12pages, including this one.
- Write your name and student number above, and enter the first letter of your last name in the box below.
- Answer **all** questions, and note the value of each question. A total of **x marks** is available.
- Answer each question directly on the examination paper, using the back of each page if necessary. Indicate clearly where your work can be found.
- Show your work! State assumptions, show all steps, and present all results clearly.

EXAMINER'S REPORT

1.		/10
2.		/10
3.		/10
4.		/10
5.		/10
6.		/10
Total:		/60

1. (10 marks) Short answer questions.

a) Consider a 2D dataset with 3 examples: $(-1, -1), (0, 0), (1, 1)$. If you apply PCA, what will be the first principal component? Provide it as a 2D vector. (2 marks)

b) Consider a 1D dataset with 4 examples: $-3, -1, 2, 4$. By hand, apply k-means clustering until convergence, assuming that the initial prototypes are -4 and 0. For each iteration, report the assignments of examples to prototypes and the new values of the prototypes. (3 marks)

c) Write an expression for the conjugate prior of the geometric distribution, up to a constant of proportionality. The conjugate prior should have non-zero probability over all values of θ for $0 < \theta < 1$. How many parameters does it have? Recall that the geometric random variable (RV) k is the number of Bernoulli trials until a success is observed, where the probability of success in every trial is θ. The PMF is $P(k) = (1-\theta)^{k-1}\theta$. (2 marks)

d) Suppose you use naive Bayes to classify a 3D test case as being from class 0 or class 1. For class 0, the likelihoods of the 3 inputs are 2, 5 and 4, whereas for class 1, the likelihoods are 2, 3 and 20. The prior probabilities are 0.6 and 0.4 for class 0 and class 1. What are the posterior probabilities and what is the most probable class? (3 marks)

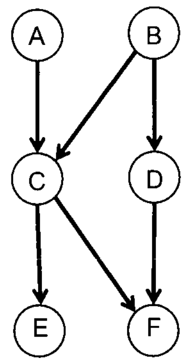
2. (10 marks) Bayesian inference. In a simplified model of driving tests, there are two types of drivers: bad drivers and good drivers. For a single driving test, a good driver has a 80% probability of passing and a bad driver has a 25% probability of passing. The prior probability that a driver is good is 0.4.

a) If a driver passes the test on the k-th attempt, what is the probability that he or she is a good driver? (3 marks)

b) Suppose we know that a driver passed the test on either the first or second attempt, $k \leq 2$. What is the probability that he or she is a good driver? (3 marks)

c) Suppose we know that a driver passed the test on either the first or second attempt, $k \leq 2$. What is the probability that he or she succeed on the first try, $k = 1$? (4 marks)

3. (10 marks) Graphical models: Properties and conditional independence. Consider the following BN.



a) Write down the formula of the joint probability distribution represented by the BN in terms of conditional and marginal probabilities. (2 marks)

b) Circle T or F for each of the following conditional independence statements of the random variables in the Bayesian network above. (3 marks)

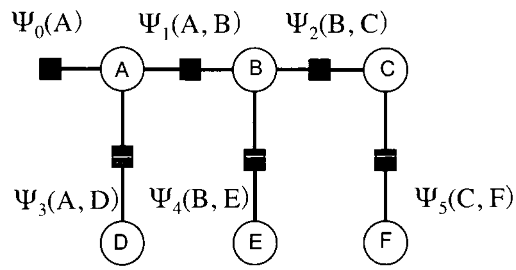
$C \perp\!\!\!\perp D \mid B$	T / F
$C \perp\!\!\!\perp D \mid B, F$	T / F
$B \perp\!\!\!\perp F \mid D$	T / F
$A \perp\!\!\!\perp D \mid C$	T / F
$A \perp\!\!\!\perp F \mid C$	T / F
$A \perp\!\!\!\perp F \mid B, C$	T / F

c) Convert the above Bayesian network to a Markov random field. Indicate the maximal cliques and for each maximal clique, specify the potential in terms of the probability distributions from part (a). (3 marks)

d) Convert the Bayesian network in part (a) to a factor graph with 6 function nodes. Specify each function in terms of the probability distributions from part (a). (2 marks)

4. (10 marks) The sum-product algorithm.

In the following factor graph, all of the random variables {A,B, ..., F} are binary. You are given the tables for each factor $\{\Psi_0, \Psi_1, ..., \Psi_5\}$. Also, messages $\mu_{A\to\Psi_1}$ and $\mu_{\Psi_2\to B}$ are pre-computed, as shown below.



A	Ψ_0
0	0.4
1	0.6

A	B	Ψ_1
0	0	0.1
0	1	0.9
1	0	0.1
1	1	0.9

B	C	Ψ_2
0	0	0.125
0	1	0.875
1	0	0.5
1	1	0.5

A B C	D E F	Ψ_3 Ψ_4 Ψ_5
0	0	0.5
0	1	0.5
1	0	0.25
1	1	0.75

A	$\mu_{A\to\Psi_1}$
0	0.4
1	0.6

B	$\mu_{\Psi_2\to B}$
0	1.0
1	1.0

(a) Use the sum-product algorithm to compute $\mu_{B\to\Psi_2}$, making sure to write down all intermediate steps. (2 marks)

$\mu_{B\to\Psi_2}(0) =$

$\mu_{B\to\Psi_2}(1) =$

(b) Now, given that we observe $E = 0$, use the sum-product algorithm to re-compute $\mu_{B \to \Psi_2}$, again writing down all intermediate steps. (3 marks)

$\mu_{B \to \Psi_2}(0) =$

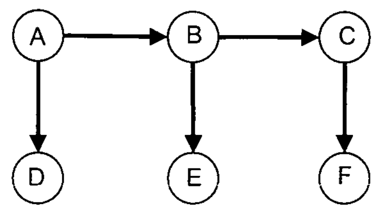
$\mu_{B \to \Psi_2}(1) =$

(c) Compute the posterior marginal probability $P(B|E = 0)$, again writing down all intermediate steps. (2 marks)

$P(B = 0|E = 0) =$

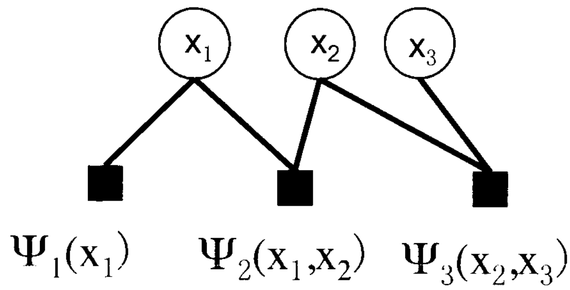
$P(B = 1|E = 0) =$

(d) Does the factor graph in part (a) correspond to the following BN? Why or why not? (2 marks)



(e) To compute $P(C)$, messages need to be propagated using a correct schedule in the factor graph. List an order in which the messages should be computed so as to correctly compute $P(C)$, using the fewest number of message updates. Indicate the order using a list such as $B \to \Psi_4, C \to \Psi_2, \ldots$. (1 mark)

5. (10 marks) Approximate inference. Consider the following factor graph with three variables and three factors.



x_1	Ψ_1
0	2
1	1

x_1	x_2	Ψ_2
0	0	2
0	1	1
1	0	3
1	1	3

x_2	x_3	Ψ_3
0	0	2
0	1	1
1	0	1
1	1	3

a) Suppose we run iterative conditional modes (ICM) by updating x_1, x_2, x_3 in that order, repeatedly. What will ICM converge to if we start with $x_1 = 0, x_2 = 1, x_3 = 1$? Show your work and write down the result of each update. (3 marks)

b) What will the ICM converge to if we start with $x_1 = 0, x_2 = 1, x_3 = 0$? If the solution is different from the solution in (a), which one is better? Show your work and write down the result of each update. (3 marks)

c) Suppose we run Gibbs sampling by updating x_1, x_2, x_3 in that order. What is the probability that we go from $x_1 = 0, x_2 = 1, x_3 = 1$ to $x_1 = 1, x_2 = 1, x_3 = 0$ in one iteration? Show your work. (4 marks)

6. (10 marks) Learning fully and partially observed BNs. Based on five training cases ($t = 1 \dots 5$), you train the parameters of a very large BN that contains a binary variable y with a binary parent x.

a) Suppose both x and y are observed and the five observations are as given below. For each training case, the column indicates the configuration of x and y that was observed. Find the maximum likelihood estimate of $P(y|x)$. (2 marks)

x	y	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
0	0	1	0	0	0	0
0	1	0	0	1	0	0
1	0	0	1	0	0	1
1	1	0	0	0	1	0

x	y	$P(y\|x)$
0	0	
0	1	
1	0	
1	1	

b) Suppose x is observed, as indicated below by the non-zero entries. y is unobserved, but probabilistic inference was used to fill in y as shown below. Find the updated distribution $P(y|x)$. (2 marks)

x	y	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
0	0	.8	0	.1	0	0
0	1	.2	0	.9	0	0
1	0	0	.7	0	.2	.9
1	1	0	.3	0	.8	.1

x	y	$P(y\|x)$
0	0	
0	1	
1	0	
1	1	

c) Suppose x and y are both unobserved, but probabilistic inference was used to fill them in as shown below. Find the updated distribution $P(y|x)$. (3 marks)

x	y	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
0	0	.6	.1	.1	.1	.1
0	1	.2	.3	.7	.1	.2
1	0	.1	.5	.1	.7	.6
1	1	.1	.1	.1	.1	.1

x	y	$P(y\|x)$
0	0	
0	1	
1	0	
1	1	

d) Suppose both x and y are unobserved and 400 steps of Gibbs sampling are used to fill them in. Samples from steps 100, 200, 300 and 400 are used. For each training case, the 4 samples are as given below. Find the maximum likelihood estimate of $P(y|x)$. (3 marks)

x	y	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
0	0	1 0 1 0	0 1 0 0	0 0 0 0	0 0 0 0	1 0 0 0
0	1	0 1 0 0	0 0 0 0	1 1 0 1	0 0 0 0	0 0 0 0
1	0	0 0 0 1	1 0 1 1	0 0 0 0	0 1 0 0	0 1 1 1
1	1	0 0 0 0	0 0 0 0	0 0 1 0	1 0 1 1	0 0 0 0

x	y	$P(y\|x)$
0	0	
0	1	
1	0	
1	1	