Name: ______________________________ Student No.: ____________

University of Toronto
Faculty of Applied Science and Engineering
Department of Electrical and Computer Engineering

Final Examination, April 17, 2015
Time allowed: 150 minutes

ECE521H1S — Inference Algorithms and Machine Learning

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 10pages, 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 **65 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.		/5
3.		/10
4.		/12
5.		/10
6.		/10
7.		/8
Total:		/65

1. (10 marks) Short answer questions.

a) Consider a 2D dataset with 3 examples: $(-1,0),(0,1),(1,2)$. 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: $1,3,6,8$. By hand, apply k-means clustering until convergence, assuming that the initial prototypes are 0 and 4. For each iteration, report the assignments of examples to prototypes and the new values of the prototypes. (2 marks)

c) Suppose you would like to fit a Gaussian model of data x: $P(x) = \exp(-(x-\mu)^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$. Write an expression for the conjugate prior of the mean μ. How many parameters does the conjugate prior have? (3 marks)

d) Suppose you use naive Bayes to classify a 4D test case as being from class 0 or class 1. For class 0, the likelihoods of the 4 inputs are 2, 5, 6 and 4, whereas for class 1, the likelihoods are 2, 3, 1 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. (5 marks) Circle T or F for each of the following statements about k-nearest neighbours classification with N training points.

a) The training time of k-NN is O(1) (fixed and independent of N) T / F

b) As k grows from 1 to N, the test classification accuracy consistently increases. T / F

c) For small values of k, the model is underfitting. T / F

d) The decision boundary is smoother with small values of k. T / F

e) If the data is not linearly separable, k-NN cannot achieve 100% training accuracy. T / F

3. (10 marks) Show that for a linearly separable dataset, the maximum likelihood solution for a logistic regression model is obtained by finding a vector $\mathbf{w}$ whose decision boundry $\mathbf{w} \cdot \mathbf{x} = 0$ separates the two classes and then taking the magnitude of $\mathbf{w}$ to infinity.

4. (12 marks) Bayesian inference. Suppose ECE521 has two assignments and two corresponding tests. The tests get integer scores from 0 to 10. If a student does the first assignment, A1, prior to the first test, T1, he has equal probabilities of getting a mark of 7, 8, 9 or 10, whereas if he doesn't do A1, he has equal probabilities of getting 5, 6, 7 or 8. If a student does the second assignment, A2, his chances of getting 7, 8, 9, 10 on test T2 are 0.2, 0.5, 0.2, 0.1, whereas if he doesn't do A2, his chances of getting 5, 6, 7, 8 are 0.2, 0.5, 0.2, 0.1. *You may find it helpful to first write down all of the marginal and conditional probabilities below, before answering the questions.*

a) Before taking the course, a student randomly chooses, with equal probability, either to do both assignments or to do no assignments. Given that he got a score of 8 on T1, what is the probability that he did the assignments? (3 marks)

b) Suppose the student from (a) got a score of 8 on T1 and a score of 8 on T2. Given this, what is the probability that he did the assignments? (3 marks)

c) Suppose an overworked student skips A1, but she decides that if she gets less than 7 on T1, she will do A2, and otherwise she will skip A2. What is the probability that she will do A2? (3 marks)

d) Before taking the course, a student randomly chooses, with equal probability, whether or not to do A1. He decides that if he gets less than 7 on T1, he will do A2, and otherwise he will skip A2. We don't know his performance on T1, but we know that he got 8 on T2. What is the probability that he did A2? (3 marks)

5. (10 marks) Graphical models. Suppose a distribution over variables u, v, w, x, y and z can be written in the form

$$P(u, v, w, x, y, z) = P(u)P(v)P(w|u)P(x|u)P(y|u, v)P(z|u, w, x, y).$$

a) Draw the Bayesian network describing $P(u, v, w, x, y, z)$. (3 marks)

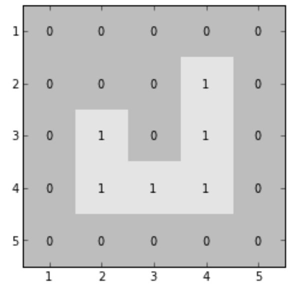
b) Draw an undirected factor graph describing $P(u, v, w, x, y, z)$ using six function nodes corresponding to the Bayesian network factorization. Give an expression for each local function using the terms $P(u)$, $P(v)$, $P(w|u)$, $P(x|u)$, $P(y|u, v)$ and $P(z|u, w, x, y)$. (3 marks)

c) Based on the structure of each of the graphical models found in parts (a) and (b), indicate whether each of the conditional independencies in the following table is true or false. (4 marks)

Graphical model	$u \perp\!\!\!\perp v$	$w \perp\!\!\!\perp y$	$w \perp\!\!\!\perp y \mid u$	$z \perp\!\!\!\perp v \mid u, y$
Bayesian network (a)				
Factor graph (b)				

6. (10 marks) ICM and Gibbs sampling. Suppose that you'd like to denoise a 5×5 image Y.

Y



Consider a grid factor graph with hidden binary variables $x_{i,j} \in \{0, 1\}$, where i indexes rows and j indexes columns. The factor graph has pairwise potentials connecting each interior hidden variable to its four hidden neighbours, so that neighboring variables are more likely to have the same value: $\phi_{xx}(0,0) = 10$, $\phi_{xx}(0,1) = 1$, $\phi_{xx}(1,0) = 1$, $\phi_{xx}(1,1) = 10$. It also has potentials connecting each hidden variable to its observed pixel. For image Y, these potentials are $\phi_{xy}(0,0) = 4$, $\phi_{xy}(0,1) = 1$, $\phi_{xy}(1,0) = 1$, $\phi_{xy}(1,1) = 4$.

The boundary hidden variables are set to their observed values, and only the interior hidden variables are updated.

a) Draw the fragment of the factor graph that consists of $x_{3,3}$ and the variables in its Markov blanket. There should be six nodes in this fragment. (1 mark)

b) Show that the potential ϕ_{xy} correspond to noise levels where the probability of flipping each pixel is 0.2 in image Y. (2 marks)

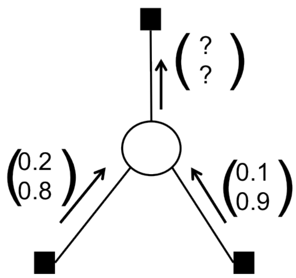
c) Suppose we initialize the hidden variables to the observed values in image Y. In Gibbs sampling, what is the probability of setting $x_{2,3} = 1$? What is the result of an ICM update for $x_{2,3}$: 0 or 1? Show your work. (3 marks)

d) During ICM, does the posterior probability of the entire set of hidden variables monotonically increase? How about for Gibbs sampling? In each case, explain why or why not. (2 marks)

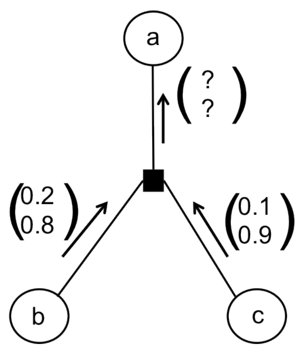
e) Repeat part (c) using the following observation potentials: $\phi_{xy}(0,0) = 40000, \phi_{xy}(0,1) = 10000, \phi_{xy}(1,0) = 10000, \phi_{xy}(1,1) = 40000$. (2 marks)

7. (8 marks) Message passing.

a) For the following fragment from a factor graph, use the sum-product algorithm rules to write down the unknown message next to the diagram. Do not normalize the message. *Partial marks will not be given, so be careful to obtain the correct numerical values.* (1 mark)

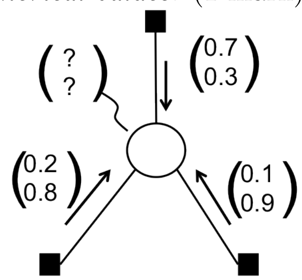


b) For the following fragment from a factor graph, where the function has the values shown in the table, determine the unknown message using the sum-product algorithm rules. Do not normalize the message. *Partial marks will not be given, so be careful to obtain the correct numerical values.* (3 marks)

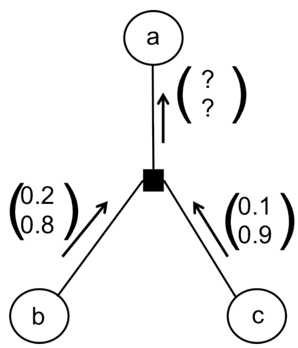


a,b,c	f
0,0,0	0
0,0,1	1
0,1,0	2
0,1,1	3
1,0,0	4
1,0,1	5
1,1,0	6
1,1,1	7

d) For the following fragment from a factor graph, use the sum-product algorithm rules to determine the *marginal probability* of the variable, making sure to normalize. *Partial marks will not be given, so be careful to obtain the correct numerical values.* (1 mark)



e) **Max-product algorithm.** For the following fragment from a factor graph, determine the unknown message using the **max-product** algorithm rules. Do not normalize the message. *Partial marks will not be given, so be careful to obtain the correct numerical values.* (5 marks)



a,b,c	f
0,0,0	0
0,0,1	1
0,1,0	2
0,1,1	3
1,0,0	4
1,0,1	5
1,1,0	6
1,1,1	7