COMP 61011

Two hours

EXAM PAPER MUST NOT BE REMOVED FROM
THE EXAM ROOM AND MUST BE RETURNED

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Foundations of Machine Learning

Date: Tuesday 16th January 2018

Time: 14:00 - 16:00

Please answer ALL Questions provided

Section A and Section B should be answered in separate answerbooks.

Answer ALL Questions in Section C
Write your answers directly on the exam paper. Only answers written in the boxes on the exam paper will be marked.

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text

[PTO]

Section A

Answer *all* questions.

All answers below can be stated briefly, and still gain full marks. Please try to state your answer briefly rather than writing down everything you know about the topic.

1. Imagine we had a dataset $\{\mathbf{x}_n, y_n\}_{n=1}^N$, where each $\mathbf{x}$ is a d-dimensional vector and y is a scalar. For the model $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - t$, regard the following loss function:

   $e = \sum_{n=1}^N (f(\mathbf{x}_n) - y_n)^2$

   Write down the plain stochastic gradient descent algorithm, with a learning rate α, for this loss with respect to the parameters $\{\mathbf{w}, t\}$, being sure to state the update equations explicitly, e.g. not simply as the expression $\frac{\partial e}{\partial \mathbf{w}_i}$.

   (6 marks)

2. If we instead used the logistic regression model, what would be a more appropriate loss function? 1 mark for the name in words, 1 for the formal definition as an equation.

   (2 marks)

3. State the equation for the entropy of a discrete feature. Now, being sure to state what log base you use, calculate it for a distribution $\{0.2, 0.25, 0.05, 0.5\}$. (2 marks)

Section B

Answer *all* questions.

1. What does it mean for two propositions, X and Y, to be independent of each other? (2 marks)

2. If two propositions X and Y are such that $P(X|Y) = P(X)$, are they independent or not independent of each other? Demonstrate this algebraically. (3 marks)

3. You are in either World 1 or World 2. In World 1 there are one million birds, of which 100 are white swans. In World 2 there are two million birds, of which 200,000 are white swans and 1,800,000 are black swans. You observe one bird that is a white swan. Provide a Bayes theorem argument for which of the two worlds you are in. (3 marks)

4. Why is the result of the previous question paradoxical? (2 marks)

5. What is the Naive Bayes Assumption? (2 marks)

6. Show the Bayesian net corresponding to the following joint probability distribution. $p(x1,x2,x3,y) = p(y)p(x1|y,x2)p(x2|y,x3)p(x3|y)$ (3 marks)

Section C contains Multiple Choice Questions and is restricted