COMP 61011

Two hours - online hybrid

The exam is hybrid and will be taken on line and answered on paper.

EXAM PAPER MUST NOT BE REMOVED FROM THE EXAM ROOM

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Foundations of Machine Learning

Date: Tuesday 22nd January 2019

Time: 14:00 - 16:00

This is a hybrid examination with sections to be answered online and questions to be answered on paper

Please answer All Questions in Section A in a separate answerbook and All Questions in Section B online

---

This is a CLOSED book examination

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

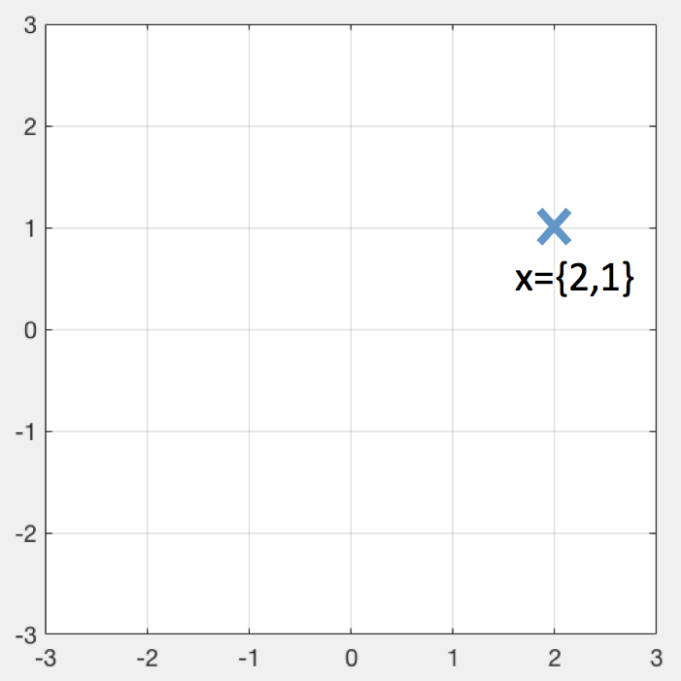
[PTO]

Section A

Answer *all* questions.

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

1. Using the axes below, draw the *decision boundary* for a model $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - t$ with parameters $\mathbf{w} = [2, -1]$, $t = 1$. State what the predicted label y would be, for $\mathbf{x} = [2, 1]$. (3 marks)



2. State the full equation for the logistic regression model. State an appropriate loss function for this model – giving both the name in words, and the equation. (3 marks)

3. Given $b = 2$, $c = 3$, $d = 4$, and a function $L = \frac{1}{2}(bc - d)^2 + \frac{1}{2}b^2$, calculate one step of a gradient descent update minimising L with respect to b, using step size $\alpha = 0.1$. After this one step, what is the value of b, and what is the value of L? (4 marks)

4. What is the Naive Bayes assumption? State the mathematical expression, its meaning in words, and draw the corresponding Bayesian network. (3 marks)

5. There is a horrible disease in the Manchester student population, "Examophobia", but 90% of all students are immune to it. Most of them (75%) wish their lecturers wouldn't set such hard questions. The reliability of a particular skin test for Examophobia is as follows: if the student has the disease, the test correctly identifies it 98% of the time. If the student does not have it, the test comes back negative 99% of the time. Imagine we take the entire student population of the University of Manchester, $50,000$ students (52% female, 48% male) and a random person is chosen. It's you. Your test comes back "positive" for Examophobia.

   What is the probability that you have Examophobia? (4 marks)

6. State the equation for the entropy of a discrete feature. Now, being sure to state what log base you use, calculate it for a categorical random variable $X \in \{1,2,3,4\}$ with distribution $P(X) = \{0.5, 0.4, 0.05, 0.05\}$. (2 marks)

7. Assuming the feature from the previous question is a potential feature in a dataset I provide to you, give the full ID3 decision tree algorithm in pseudo-code, being sure to define all notation. (6 marks)

Section B contains Multiple Choice Questions and is restricted