Two hours - online hybrid

EXAM PAPER MUST NOT BE REMOVED FROM THE EXAM ROOM

UNIVERSITY OF MANCHESTER
DEPARTMENT OF COMPUTER SCIENCE

Foundations of Machine Learning

Date: Thursday 23rd January 2020

Time: 14:00 - 16:00

This is a hybrid examination with sections to be answered online and questions to be answered on paper
The examination is worth a total of 50 marks

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

---

This is a CLOSED book examination

Electronic calculators may be used in accordance with the University regulations

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. (a) Table 1 shows a simple data set with two features. The initial conditions are $w_1 = 3$, $w_2 = 40$ and $t = 100$ and the perceptron algorithm learning rate is set to 0.1. The discriminant function is $\mathbf{w}^T\mathbf{x} - t$.

	$\mathbf{x_1}$	$\mathbf{x_2}$	**Y**
example 1	187	80	0
example 2	160	70	0
example 3	150	80	1
example 4	192	92	1

Table 1

Perform one update using the perceptron update rule applied to example 1, to complete Table 2.

	$\mathbf{w_1}$	$\mathbf{w_2}$	**t**
Initial State	3	40	100
After example 1			

Table 2

(4 marks)

(b) State the perceptron convergence theorem (2 marks)

(c) The logistic function is:

$$f(x) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x} - t)}}$$

Given the data set in Table 1 (and ignoring x_2, we wish to apply logistic regression to predict Y from x_1 only. If $w = 0.2$, what would be an appropriate value for t, after convergence? (1 mark)

2. (a) What is the main difference between the decision boundary induced by a perceptron and the decision boundary induced by a linear SVM?

(2 marks)

(b) The SVM error function with soft margins is:

$$E = \sum_{i=1}^{N} \max\{0, 1 - y_i f(\mathbf{x}_i) - \xi_i\} + \frac{1}{2}\sum_{j=1}^{d} w_j^2 + C\sum_{i=1}^{N} \xi_i$$

Define the terms of this expression, i.e. state what is meant by w_j, $f(\mathbf{X}_i)$, y_i, C and ξ_i. State also what is achieved by minimising the first part versus the second part versus the third part (i.e. before and after the '+' symbols) (6 marks)

3. (a) Calculate the sensitivity, specificity, and F-measure of the classifier that produced the confusion matrix in Table 3 (treat True as the positive outcome)

(3 marks)

		Predicted	
		False	True
Actual	False	50	12
	True	2	76

Table 3

(b) Figure 1 shows a set of points on an ROC graph which represents the performance of 5 machine learning classifiers that predict whether an event is True or false. Write down the best and worst classifiers and explain your reasoning. (3 marks)

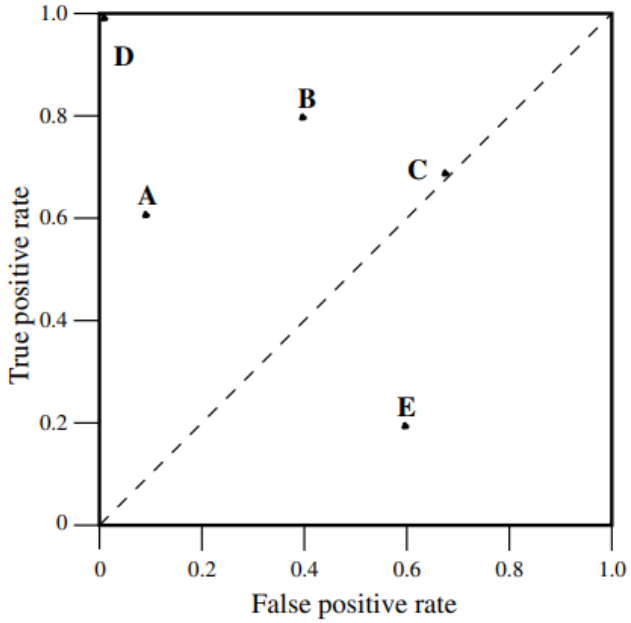

Figure 1

4. Exactly 1 in 5 people in Smallville have the disease *Bloaty Head*. When a person goes to the hospital, they will receive exactly one of two possible tests for the disease, TEST1 or TEST2. The doctor conducts TEST1 $\frac{3}{4}$ of the time. When TEST1 is conducted, the outcome is as follows:

   - If the person has Bloaty Head, TEST2 is positive with probability $\frac{5}{6}$
   - If the person does not have Bloaty Head, TEST 1 is *positive* with probability $\frac{1}{3}$

   When TEST2 is conducted, the outcomes is as follows:

   - If the person has Bloaty Head, TEST2 is positive with probability 1
   - If the person does not have Bloaty Head, TEST 1 is *positive* with probability $\frac{1}{8}$

   A person is picked at random from Smallville and sent to the hospital to test for Bloaty Head. The result comes out positive. What is the probability that the person has the disease? (4 marks)

Section B

Section B contains restricted multiple choice questions (MCQs) and is NOT published.