

Two hours

QUESTION PAPER MUST NOT BE REMOVED FROM THE EXAM ROOM

**UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE**

Machine Learning and Optimisation

Date: Friday 27th January 2012

Time: 14:00 - 16:00

**Answer ALL 10 short questions in Section A
Answer ONE question from Section B
Answer ONE question from Section C**

Use SEPARATE Answerbooks for EACH section

**For full marks your answers should be concise as well as accurate.
Marks will be awarded for reasoning and method as well as being correct**

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 is
restricted and
cannot be
published*

Section B

Answer *one* question only from this section.

1. a) Describe the Bagging algorithm in pseudo-code, remembering to give definitions of any technical words you use, including any properties that you think interesting, and explaining what models might be suited to the algorithm, and why.
(5 marks)

- b) Describe the Boosting algorithm, and state the main difference(s) from Bagging.
(5 marks)

- c) I recently bought a device that claimed to predict the weather anywhere in the world. Last year in Barbados it rained for 20 of the total 365 days in the year. We will call the rainy days the ‘positive’ class. Over the rainy days, my weather-prediction device correctly predicted rain 15 times. However, several times through the year, it predicted rain when in fact it turned out nice and sunny. I counted these events – it happened 30 times!

Calculate the accuracy, sensitivity, and specificity of my weather predictor.
(6 marks)

- d) Related to the previous question, if a false negative (i.e. going out without an umbrella, and it rains) costs me a laundry bill of £20, and a false positive (i.e. buying a new umbrella to use) costs me £5, what is the cost of my classifier?
(4 marks)

2. a) Give pseudo-code for the ID3 algorithm. Be sure to state the base case for the recursion and be precise when discussing the split criterion. (10 marks)
- b) State the formula for the entropy of a feature and calculate it for a binary feature with $p(X = 1) = 0.75$. (3 marks)
- c) What are 'filter' and 'wrapper' approaches to feature selection? Give a definition and computational complexity properties for each. (5 marks)
- d) What is the relationship between splitting criteria in decision trees, and filter methods for feature selection? (2 marks)

Section C

Answer *one* question only from this section.

1. The *Support Vector Machine* (SVM) is an effective and popular Machine Learning algorithm for classification.

- a) Describe the general principle of an SVM, including how it differs from a Perceptron classifier. In your description, it is essential to give definitions of the important concepts in an SVM, e.g., margin and support vectors. (6 marks)
- b) Give one *advantage* and one *disadvantage* of an SVM classifier. (2 marks)
- c) Given a training data set $X = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, the cost (or loss) function of an SVM consists of two objectives given as follows:

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\}$$

where \mathbf{w} and b are weights and bias of the SVM, and α is the vector of trade-off coefficients (or Lagrange multipliers, formally)

- i) What is the purpose of the $\frac{1}{2} \mathbf{w}^T \mathbf{w}$ term? (2 marks)
 - ii) What is the purpose of the $\sum_{i=1}^N \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\}$ term? (2 marks)
 - iii) After SVM learning, each α_i takes either zero or non-zero value. What does it indicate in each situation, respectively? (2 marks)
- d) Describe the general principle of kernel-based SVM, including why it is often more powerful than the original SVM in general and works efficiently in a very high dimensional feature space. (6 marks)

2. Clustering analysis is an unsupervised learning process that groups a set of physical or abstract objects into clusters of similar or coherent objects.
- a) *K-means* is a popular clustering algorithm. Describe this algorithm in detail and give *one advantage* and *two disadvantages* of the *K-means* algorithm. (4 marks)
- b) A modified *K-means* algorithm, named **K-medoids**, was proposed to overcome the weaknesses of the *K-means* algorithm. It is described as follows.

Input

K : number of clusters pre-defined

D : dataset containing N objects

Output : a partition of D with K clusters, C_1, \dots, C_K .

Step 1: Arbitrarily choose K objects in D as *seeds*, \mathbf{o}_k , $k = 1, \dots, K$

Step 2: Repeat

Step 2.1: Assign each remaining object to the cluster with the nearest seeds;

Step 2.2: For the current partition, calculate its cost E_{old} based on a cost function such as the one appearing in the lecture notes on *K-means*;

Step 2.3: Randomly select a non-seed object \mathbf{o}_{rand} ;

Step 2.4: Calculate the cost E_{new} after swapping seed object \mathbf{o}_k with \mathbf{o}_{rand} ;

Step 2.5: If $E_{new} < E_{old}$, then swap \mathbf{o}_k with \mathbf{o}_{rand} to form a new set of K seeds;

Until No membership change in all K clusters.

Given all the differences between *K-means* and *K-medoids*, explain what weaknesses of *K-means* are likely to be overcome and why the *K-medoids* algorithm is less likely to suffer from such weaknesses. (8 marks)

- c) Suppose that Tesco is going to develop its overseas market by establishing a chain of supermarkets/shops in Shanghai, the largest city in China. Households may be clustered so that either one large Tesco Extra supermarket or multiple small Tesco Express shops will be built per cluster. You are asked to use the *K-medoids* algorithm described above along with a clustering algorithm learnt in this module to help them make a plan to choose locations and shop types with appropriate information available. Describe your method that decides locations and types of shops (i.e., Extra supermarket or Express shop) so that each cluster must cover at least a number of household, e.g. 10,000. It is essential to justify your method and address main issues to be considered for satisfactory services, e.g., shopping by travelling a short distance, and a low operational cost for Tesco. (8 marks)