The University of Nottingham

SCHOOL OF COMPUTER SCIENCE

A LEVEL 3 MODULE, AUTUMN SEMESTER 2010-2011

MACHINE LEARNING

Time allowed: **Two Hours**

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced

Answer All Questions

Only silent, self contained calculators with a Single-Line Display or Dual-Line Display are permitted in this examination

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject specific translation dictionaries are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

DO NOT turn your examination paper over until instructed to do so

Question 1

(a) Using handwritten digit recognition as an example, list the major steps of building a machine learning system and for each step discuss what it is involved in detail. Use a specific or a generic learning model as appropriate to illustrate your answer.

(10 marks)

(b) Discuss the concepts of **generalization** and **over-fitting** in the context of machine learning and give one possible solution to avoid over-fitting.

(10 marks)

Question 2

(a) Design a single layer perceptron that will classify the following 3-d inputs into two classes.

Input	Output
(1, 0, 0,)	Class 1
(0, 1, 0)	Class 1
(1, 1, 0)	Class 1
(0, 0, 2)	Class 2
(0, 0, 1)	Class 2

Your answer should include:

(i) A diagram of the perceptron.
(ii) The values of the connection weights and the threshold or bias connection.
(iii) The equation of the perceptron's decision boundary.
(iv) A verification of the correctness of your design.

(10 marks)

(b) A network consists of two ADLINE units, N1 and N2 is shown in Fig. Q2(b).

(i) Derive a delta training rule for all the weights, you must show the derivation process.

(5 marks)

(ii) Assuming the weights have following values $w_0 = w_1 = w_2 = w_3 = w_4 = 0.1$, using a training rate of 0.1, show the new values of w_0 and w_4 after presenting a training sample $x_1 = 0.5$, $x_2 = 0$ with a desired output d = 0.5

(5 marks)

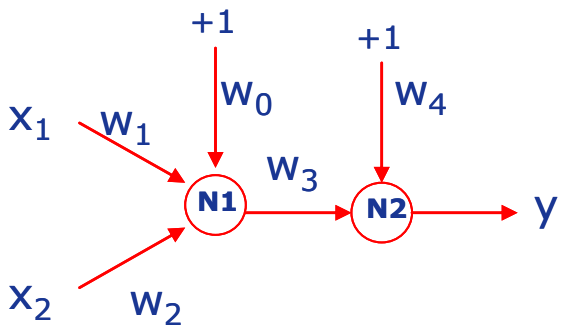

Fig. Q2(b)

Question 3

In an object recognition task, it is known that the objects come from one of two classes, C_1 or C_2. Each instance of an object X has four features, $X = (x_1, x_2, x_3, x_4)$. An experiment has collected 14 instances and their feature values and classification are shown in the following table

$X = (x_1, x_2, x_3, x_4)$.				Classification
x_1	x_2	x_3	x_4	C
1	1	1	1	C_1
1	1	1	2	C_1
2	1	1	1	C_2
3	2	1	1	C_2
3	3	2	1	C_2
3	3	2	2	C_1
2	3	2	2	C_2
1	2	2	1	C_1
1	3	2	1	C_2
3	2	2	1	C_2
1	2	2	2	C_2
2	2	1	2	C_2
2	1	2	1	C_2
3	2	1	2	C_1
1	3	1	2	=?

Using the data with known classification, we can build a Naïve Bayesian classifier to classify the unknown instance X = (1, 2, 1, 2) in the last row of the table. According to the Naïve Bayesian classifier, the classification of the unknown instance, C, can be expressed as

$$C = \arg\max_{C \in \{C_1, C_2\}} P(C)P(x_1 = 1 \mid C)P(x_2 = 2 \mid C)P(x_3 = 1 \mid C)P(x_4 = 2 \mid C)$$

Which class will the instance X = (1, 2, 1, 2) be classified into? You must show the detail of the working process. An answer without the working process will not receive any marks.

(20 marks)

Question 4

(a) The perceptron and linear SVM classifier both use a pyherplane to separate two linearly separable classes. Compare and contrast their respective training algorithms.

(8 marks)

(b) The Support Vectors are the only training samples that are useful for determining the decision hyperplane of the SVM, hence the rest training samples can be ignored in SVM training. True or false? Discuss.

(6 marks)

(c) Discuss how SVMs tackle linearly non-separable classes.

(6 marks)

Question 5

(a) Experiment 1 has produced N 2d data points $\{X(i) = (x_1(i), x_2(i))\}$, i = 1, 2, … N. Experiment 2 has produced M 2d data points $\{Y(i) = (y_1(i), y_2(i))\}$, i = 1, 2, … M}. The covariance matrices of these two sets of data, S_x and S_y, are

$$S_X = \begin{bmatrix} 0.7 & 0.85 \\ 0.85 & 0.5 \end{bmatrix} \qquad S_Y = \begin{bmatrix} 0.8 & 0.1 \\ 0.1 & 0.6 \end{bmatrix}$$

(i) What can be inferred about the relations between x_1 and x_2; and that between y_1 and y_2.

(5 marks)

(ii) If Principal Component Analysis (PCA) is applied to reduce the dimensionality of X and Y with the aim of reducing the dimensionality of the measurements, which dataset can achieve more effective dimensionality reduction and why.

(5 marks)

(b) An experiment has produced the following 3d feature vectors $X = (x_1, x_2, x_3)$ belonging to two classes. Design a decision tree classifier to class an unknown feature vector X = (1, 2, 1).

$X = (x_1, x_2, x_3)$			
x_1	x_2	x_3	Classes
1	1	1	1
1	1	1	2
1	1	1	1
2	1	1	2
2	1	2	1
2	2	2	2
2	2	2	1
2	2	1	2
1	2	2	2
1	1	2	1
1	2	1	= ?

(10 marks)