UNIVERSITY OF SOUTHAMPTON COMP3008W1

SEMESTER 2 EXAMINATION 2009/2010

MACHINE LEARNING

Duration: 120 mins

Answer all parts of the question in section A (20 marks) and TWO *questions from section B (25 marks each)*

This examination is worth 70%. The coursework was worth 30%.

University approved calculators MAY be used.

 Number of pages 5

Section A

Question 1

(a) Explain what is meant by generalisation error and describe how it is estimated. *(2 marks)*

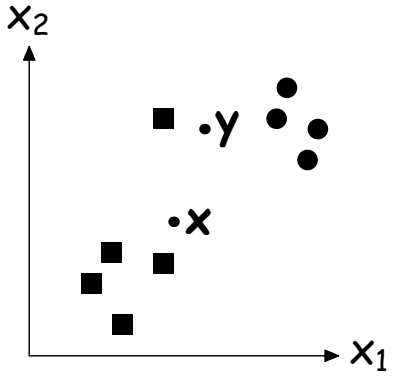
(b) Give a Bayesian interpretation for minimising the sum of the mean squared error plus a regularisation term. *(3 marks)*

(c) Show that a MLP using linear nodes is no more powerful than a linear perceptron. *(5 marks)*

(d) Describe what K-fold cross-validation is and explain what its purpose is.

(3 marks)

(e) The dataset below consists of two classes, squares and circles.



Give the classification of the points $\boldsymbol{x}$ and $\boldsymbol{y}$ produced by a K-Nearest Neighbours (KNN) algorithm with $K = 1$ and $K = 3$. Explain why increasing K acts like a regulariser? *(4 marks)*

(f) Explain why normalising data can be important in the context of KNN. *(3 marks)*

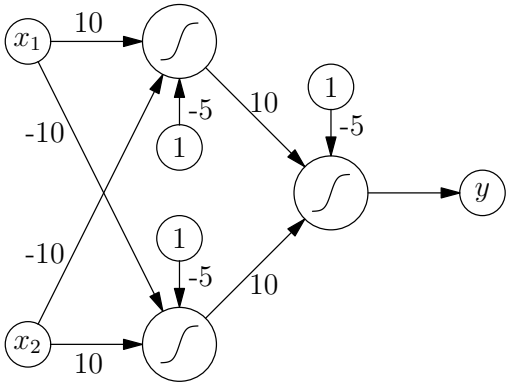
Section B

Question 2 You are given a dataset $\mathcal{D} = \{(\boldsymbol{x}_k, y_k)|k = 1, 2, 3, 4\}$ where

k	$\boldsymbol{x}_k$	y_k
1	(0,0)	0
2	(1,0)	1
3	(0,1)	1
4	(1,1)	0

(a) Explain why a perceptron is not capable of correctly classifying this dataset. *(3 marks)*

(b) The diagram below shows a MLP with two input nodes, two nodes in the hidden layer and an output node. The additional nodes, labelled 1, are pseudo inputs for implementing a threshold for each node. The weights connecting the nodes are shown on the connecting lines. The output of the nodes is equal to $g(V) = 1/(1 + \mathrm{e}^{-V})$ where V is the weighted sum of the inputs.



Show that this MLP will accurately classify the data above.
(15 marks)

(c) The problem shown is a parity problem in two dimensions. Explain why high dimensional parity problems are hard for MLPs to learn.
(7 marks)

TURN OVER

Question 3

(a) Describe the steps in Principal Component Analysis. *(7 marks)*

(b) Explain the benefits of performing PCA as a preprocessing stage for supervised learning. *(3 marks)*

(c) Explain why performing PCA on relatively small dimensional feature vectors may be advantageous despite throwing away information.

(5 marks)

(d) Explain why conventional PCA cannot be used with large dimensional images and explain how PCA can be carried out.

(7 marks)

(e) Describe the connection between how PCA works on large images and the kernel trick used in SVMs.

(3 marks)

Question 4

(a) Describe the similarities and differences between multi-layer perceptrons (MLPs), radial basis function networks (RBFs) and support vector machines (SVMs).

(10 marks)

(b) Why are regularisation terms added to the error function?

(5 marks)

(c) A weight decay term has the form $\lambda \sum_i w_i^2$. Show how adding such a term modifies the update rule for the weights and hence explain why it is known as a weight decay term. *(10 marks)*

END OF PAPER