UNIVERSITY OF SOUTHAMPTON CM311W1

SEMESTER 1 EXAMINATION 2001/2002

NEURAL COMPUTATION

Duration: 120 mins

Answer ALL *questions from section A*
and ONE *question from section B*
and ONE *question from section C.*

Calculators MAY be used

Section A

Question 1

a) What is the difference between supervised and unsupervised learning? Give examples of each.

(2 marks)

b) Why is building a multi-layer network using linear perceptrons futile?

(2 marks)

c) There are three dice on the table. Two of them are normal dice, while the other dice has 6 on two faces and 1, 2, 3 and 5 on the other face. A dice is chosen at random and is thrown 10 times. On three occasions the top face is a 6. What is the probability that the dice chosen is the dishonest dice? Show all your workings.

(5 marks)

d) What problems arise when training a multi-layer perceptron? In what sense is a support vector machine easier to train?

(3 marks)

e) Describe the difference between learning and generalisation error. Explain why minimising one may not minimise the other.

(3 marks)

f) Suggest how you might encode the following attributes

i) sex (male or female)

ii) final educational qualifications (none, GCSE, A-levels, degree)

iii) marital status (single, married, separated, widowed).

Give a short explanation of your decision.

(5 marks)

Section B

Question 2

a) Derive a matrix equation for the coefficients of the cubic polynomial
$$y = w_1 + w_2 x + w_3 x^2 + w_4 x^3$$
which goes through the points $(x, y) = (0, 1), (1, 2), (2, 1), (3, 2)$.

(5 marks)

b) Write down the mean squared error for the regression of the above polynomial with respect to an arbitrary set of points $\{(x_i, y_i)\}_{i=1}^{P}$ in terms of the matrix
$$X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_P \\ x_1^2 & x_2^2 & \cdots & x_P^2 \\ x_1^3 & x_2^3 & \cdots & x_P^3 \end{pmatrix}$$
the vector $\boldsymbol{y} = (y_1, y_2, \cdots, y_P)^\mathsf{T}$ and the weight vector $\boldsymbol{w} = (w_1, w_2, w_3, w_4)^\mathsf{T}$.

(5 marks)

c) Derive the least square solution for the weights.

(10 marks)

d) Describe what happens as the degree of the polynomial used in regression is increased.

(5 marks)

Question 3

a) Write down the definition of the covariance matrix.

(3 marks)

b) Show that the covariance matrix is positive semi-definite (i.e. the eigenvalues are non-negative).

(3 marks)

c) Describe how to perform principal component analysis.

(4 marks)

d) Explain why principal components can be useful in data analysis.

(5 marks)

e) Describe the K-means clustering algorithm. How could this be used in training a radial basis function neural network.

(10 marks)

Section C

Question 4

a) Explain what is meant by the term over-parameterisation. State how it is removed from the hyperplane, $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$, in the linear Support Vector Machine formulation to produce a canonical hyperplane.

(3 marks)

b) State the condition for separability of the two-class data-set $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\}$ with this canonical hyperplane.

(3 marks)

c) Describe the maximum margin principle and show that the resulting optimisation problem is given by the Lagrangian,

$$\Phi(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \tfrac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left[\boldsymbol{w}^\top \boldsymbol{x}_i + b \right] - 1 \right), \quad \alpha_i \geq 0.$$

(8 marks)

d) Solve the Lagrangian problem, $\max_{\boldsymbol{\alpha}} \big(\min_{\boldsymbol{w}, b} \Phi(\boldsymbol{w}, b, \boldsymbol{\alpha}) \big)$, to show that the solution for the Lagrange multipliers can be written as a quadratic program,

$$\min_{\boldsymbol{\alpha}} \tfrac{1}{2} \boldsymbol{\alpha}^\top H \boldsymbol{\alpha} + \boldsymbol{c}^\top \boldsymbol{\alpha},$$

subject to the constraints,

$$\alpha_i \geq 0, \quad \textstyle\sum_{j=1}^n \alpha_j y_j = 0.$$

(8 marks)

e) What are the Support Vectors and how do these relate to the Lagrange multipliers?

(3 marks)

Question 5

a) What is an ill-posed problem?

(5 marks)

b) Describe the method of regularisation, making reference to examples in machine learning.

(9 marks)

c) Show that the solution to the regularisation problem is equivalent to a Maximum A Posteriori (MAP) estimate.

(6 marks)

d) Discuss how regularisation parameters can be chosen.

(5 marks)

END OF PAPER