UNIVERSITY OF SOUTHAMPTON COMP3008W1

SEMESTER 2 EXAMINATION 2006/2007

MACHINE LEARNING

Duration: 120 mins

Answer ALL *questions from section A (20 marks)*
and ONE *question from section B (25 marks)*
and ONE *question from section C (25 marks).*

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

Calculators without text storage MAY be used.

Section A

Question 1

(a) Explain what is meant when a problem is under-constrained in the context of machine learning. How are under-constrained problems solved?

(3 marks)

(b) Explain what is meant when a problem is over-constrained in the context of machine learning. How are over-constrained systems solved?

(3 marks)

(c) Describe the perceptron learning algorithm for the step perceptron.

(4 marks)

(d) Steve's questions *(10 marks)*

Section B

Question 2

(a) Write down the mean squared training error, $E(\boldsymbol{w}|\mathcal{D})$, for a leaning machine $f(\boldsymbol{x}|\boldsymbol{w})$ given a finite data set $\mathcal{D} = \{(\boldsymbol{x}_k, y_k)\}_{k=1}^{P}$
(1 marks)

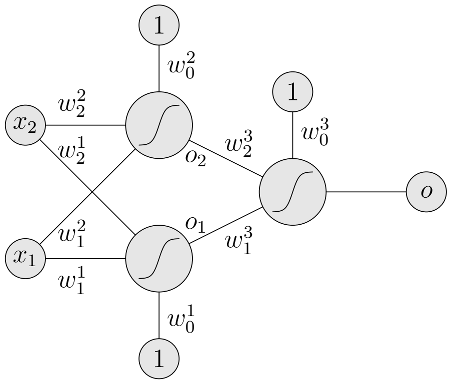
(b) The bias-variance dilemma is a theoretical analysis for understanding the causes of generalisation error. In this approach we consider the generalisation error when averaged over infinitely many randomly chosen data sets $\mathcal{D}$. Write down an equation for the generalisation error in terms of an average over training sets.
(2 marks)

(c) Denoting the set of weights that minimise the training error for a data set by $\boldsymbol{w}(\mathcal{D})$ write down an expression for the response of the average machine, $f_{av}(\boldsymbol{x})$ (i.e. the mean response averaged over machines trained on all data sets). *(2 marks)*

(d) Show the generalisation error can be written as a bias term and variance term. (Hint, add and subtract the average machine response to the generalisation error, then expand the square.) Explain what these two terms means and explain the dilemma.
(17 marks)

(e) Explain why regularisation terms can improve generalisation performance in the context of the bias-variance dilemma. Why do regularisation terms allow more complex machine to be used.
(3 marks)

Question 3 Consider the multilayer perceptron shown below



where the response function for the nodes are some squashing function $g(x)$.

(a) Write down a formula describing the output, $o = f(\boldsymbol{x}|\boldsymbol{w})$, of this network shown above. *(2 marks)*

(b) Write down an expression for the mean squared error and find the derivative with respect to w_1^3 and w_1^1. *(5 marks)*

(c) Assuming the following weights $w_0^1 = -10$, $w_1^1 = 0$, $w_2^1 = 20$, $w_0^2 = 30$, $w_1^2 = 0$, $w_2^2 = -20$, $w_0^3 = -30$, $w_1^3 = 20$, and $w_2^2 = 20$. Calculate the output o given inputs, (x_1, x_2) equal to $(0, 0)$, $(-1, 1)$, $(1, 1)$, and $(0, 2)$ for a logistic perceptrons $g(x) = 1/(1 - \mathrm{e}^{-x})$, (you may assume $g(x) = 0$ for $x \leq -10$ and $g(x) = 1$ for $x \geq 10$). *(8 marks)*

(d) Draw the separating planes for the two hidden nodes in input space (i.e. when their input fields equal zero). Which features are relevant to solving this problem? *(8 marks)*

(e) Why could this problem not be solved using a single layer perceptron? *(2 marks)*

Section C

Question 4

(25 marks)

Question 5

(25 marks)

END OF PAPER