UNIVERSITY OF SOUTHAMPTON COMP3008W1

SEMESTER 2 EXAMINATION 2012/2013

MACHINE LEARNING

Duration: 120 mins

You must enter your Student ID and your ISS login ID (as a cross-check) on this page. You must not write your name anywhere on the paper.

Student ID:	
ISS ID:	

Question	Marks
1	
2	
3	
4	
Total	

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.

Each answer must be completely contained within the box under the corresponding question. No credit will be given for answers presented elsewhere.

You are advised to write using a soft pencil so that you may readily correct mistakes with an eraser.

You may use a blue book for scratch—it will be discarded without being looked at.

Section A

Question A 1

(a) Explain what the **bias** and **variance** terms in the expected generalisation error is and explain the bias variance dilemma. *(6 marks)*

1

2

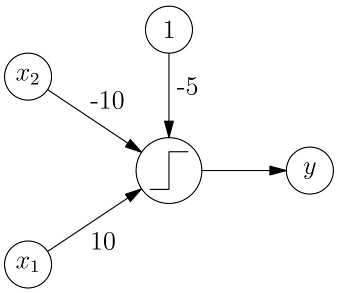
3

$\overline{6}$

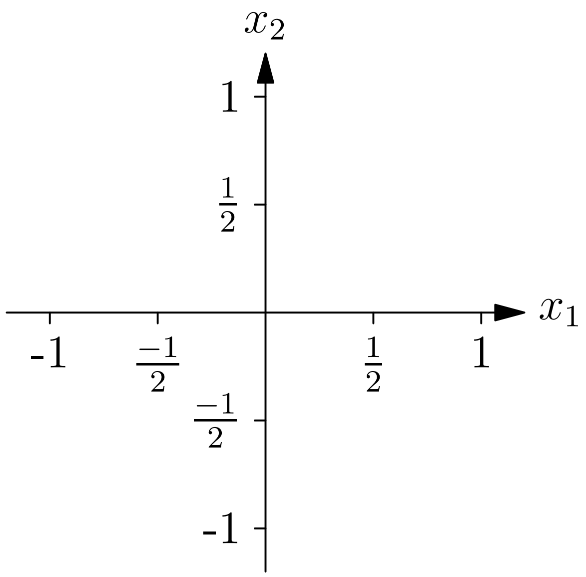
(b) Show that the kernel function $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\phi}^{\mathsf{T}}(\boldsymbol{x})\boldsymbol{\phi}(\boldsymbol{y})$, where $\boldsymbol{\phi}(\boldsymbol{x})$ is a vector equal to $(x_1^2, x_2^2, x_3^2, \sqrt{2}\,x_1x_2, \sqrt{2}\,x_1x_3, \sqrt{2}\,x_2x_3)$, can be written as $(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y})^2$ is $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors of length 3. *(4 marks)*

$\overline{4}$

(c) The figure below shows a step perceptron with an output $\Theta(V)$ which is equal to 1 if $V > 0$ and 0 otherwise. Write down the formula that describes the response of the perceptron and draw the separating surface in the input space, indicating the response y.



(5 marks)



$\overline{5}$

(d) Give (high level) pseudo code for k-means clustering. *(5 marks)*

$\overline{5}$

End of question 1

Q1: (a) $\frac{}{6}$ (b) $\frac{}{4}$ (c) $\frac{}{5}$ (d) $\frac{}{5}$ Total $\frac{}{20}$

Section B

Question B 2

(a) Show that the squared error of a linear perceptron with data $(\boldsymbol{x}_k, y_k)$ for $k = 1, 2, \ldots, P$ can be written as $\|\mathbf{X}^{\mathsf{T}}\boldsymbol{w} - \boldsymbol{y}\|^2$ where $\mathbf{X}$ is the usual matrix of input patterns and $\boldsymbol{y}$ is a vector of target values. *(5 marks)*

$\overline{5}$

(b) Write down the cost you would minimise in vector form if you include a weight decay regularisation term with a regularisation parameter ν. *(3 marks)*

$\overline{3}$

(c) Obtain an equation for the weights that minimise this cost (show your working). *(8 marks)*

$\overline{8}$

(d) Explain why adding the regularisation term guarantees that the problem is never ill-posed and makes the solution better conditioned. *(9 marks)*

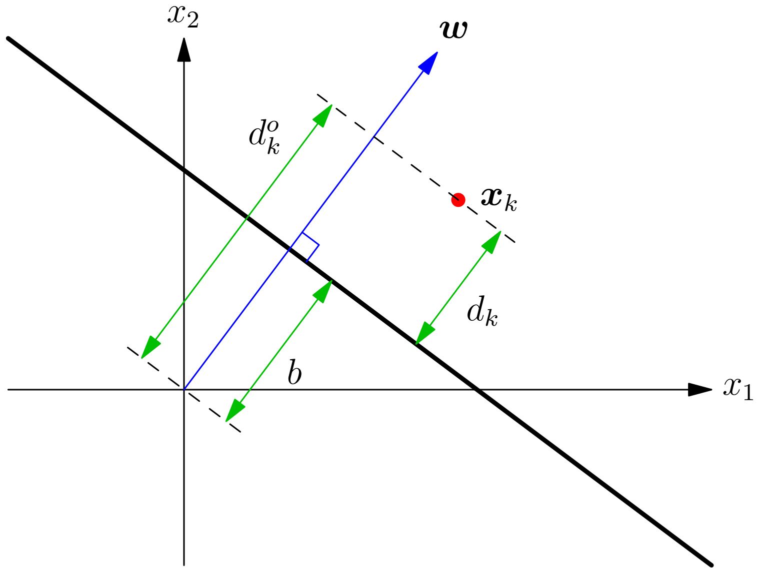
$\overline{9}$

End of question 2

Q2: (a) $\frac{}{5}$ (b) $\frac{}{3}$ (c) $\frac{}{8}$ (d) $\frac{}{9}$ Total $\frac{}{25}$

Question B 3

(a) Write down a formula for the minimum distance d_0^k between $\boldsymbol{x}_k$ and a hyperplane through the origin perpendicular to $\boldsymbol{w}$, and the minimum distance d_k from $\boldsymbol{x}_k$ to the hyperplane perpendicular to $\boldsymbol{w}$ displaced by b. *(5 marks)*



$d_k^o =$

$d_k =$

$\overline{5}$

(b) Depending on the category $y_k \in \{-1, 1\}$, write down the condition for a data point to be at least a distance m above (or below if $y_k = -1$) the hyperplane shown in part (a). *(3 marks)*

$\overline{3}$

(c) Define $\boldsymbol{w'} = \boldsymbol{w}/(m\|\boldsymbol{w}\|)$ and $b' = b/m$, rewrite the condition above and explain why minimising $\|\boldsymbol{w}'\|^2$ is equivalent to maximising the margin m. *(3 marks)*

$\overline{3}$

(d) Write down a Lagrangian for finding the maximal margin hyperplane for an SVM given data $(\boldsymbol{x}_k, y_k)$ for $k = 1, 2, \ldots, P$. *(3 marks)*

$\overline{3}$

(e) Write down the optimisation condition for the Lagrangian (i.e. what are you maximising or minimising with respect to) and what are the conditions on the Lagrange multipliers. *(3 marks)*

1

2

$\overline{3}$

(f) Find the weight vector $\boldsymbol{w}'$ and threshold b' which minimises the Lagrangian and by substituting the result back into the Lagrangian find the dual form for optimisation problem. *(8 marks)*

$\overline{}$ 8

End of question 3

Q3: (a) $\frac{}{5}$ (b) $\frac{}{3}$ (c) $\frac{}{3}$ (d) $\frac{}{3}$ (e) $\frac{}{3}$ (f) $\frac{}{8}$ Total $\frac{}{25}$

• Do not write in this space •

Question B 4

(a) Describe Bayes' rule giving a description of all the parts. *(5 marks)*

$\overline{5}$

(b) Show that if the likelihood of observing n events is given by the Poisson distribution $p(n|\mu) = \mu^n \mathrm{e}^{-\mu}/n!$, then the Gamma distribution $p(\mu) = \mu^{a_0-1}\mathrm{e}^{-b_0\mu}$ is a conjugate distribution and compute the updated parameters of the posterior. *(5 marks)*

$\overline{5}$

(c) Describe the MAP solution and explain its advantages and disadvantages over computing the posterior. *(5 marks)*

5

(d) What is the naive Bayes assumption and explain how you would use it to implement a spam filter? *(10 marks)*

$\overline{}$
10

End of question 4

Q4: (a) $\frac{}{5}$	(b) $\frac{}{5}$	(c) $\frac{}{5}$	(d) $\frac{}{10}$	Total $\frac{}{25}$

END OF PAPER