UNIVERSITY OF SOUTHAMPTON COMP3008W1

SEMESTER 2 EXAMINATION 2011/2012

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.

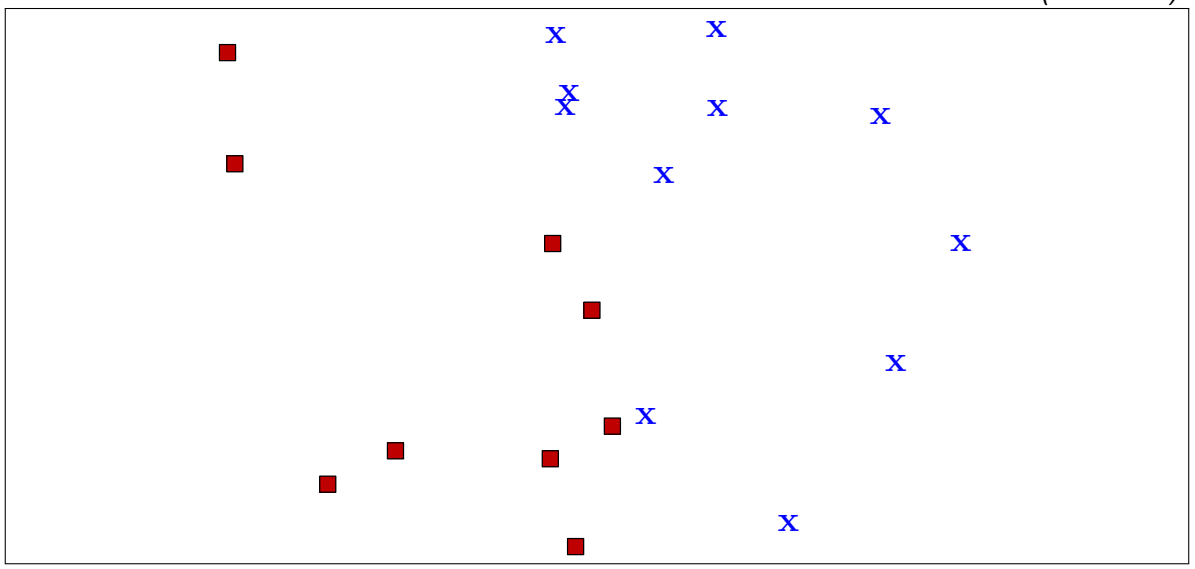
You may use a blue book for scratch—it will be discarded without being looked at.

 Number of pages 11

Section A

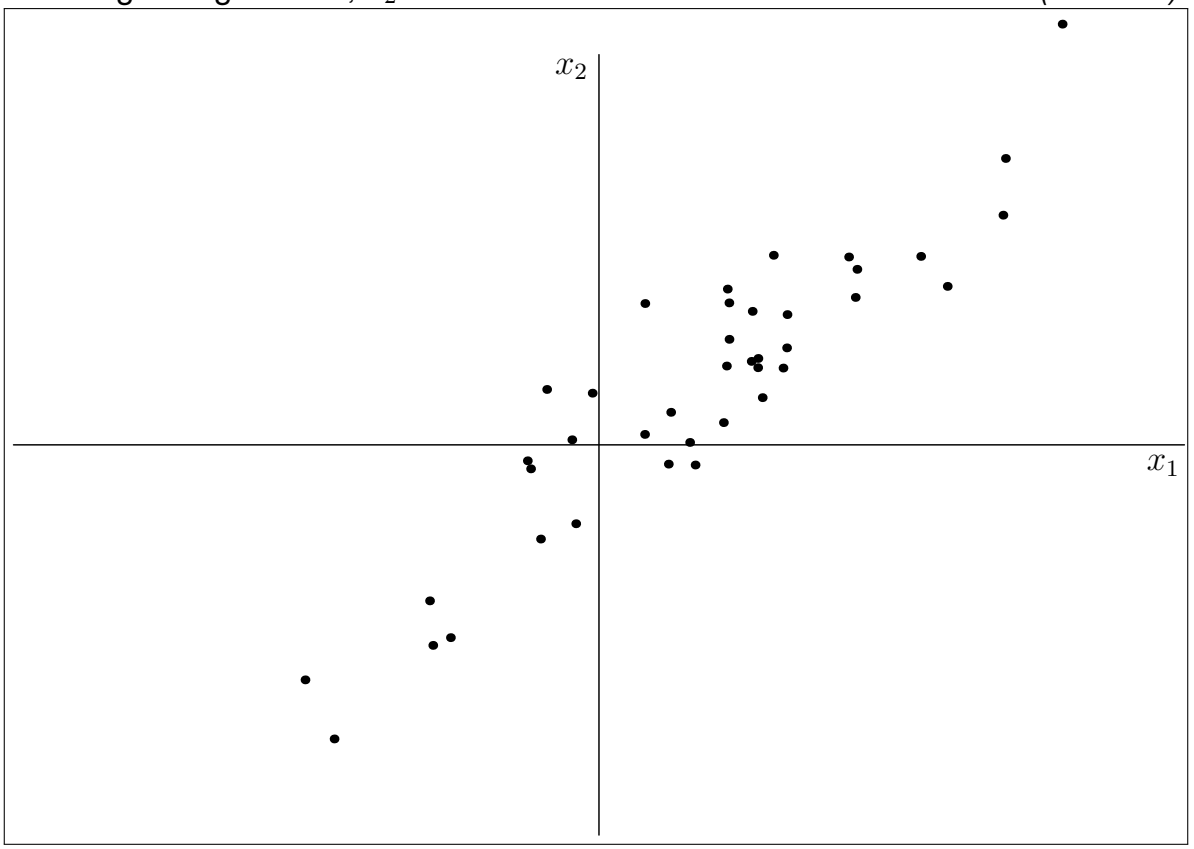
Question A 1

(a) Roughly sketch the dividing curve for a 3-nearest-neighbours classifier for the data shown below. *(5 marks)*



(b) Explain why the kernel trick allows an SVM to separate data that is not linearly separable. *(5 marks)*

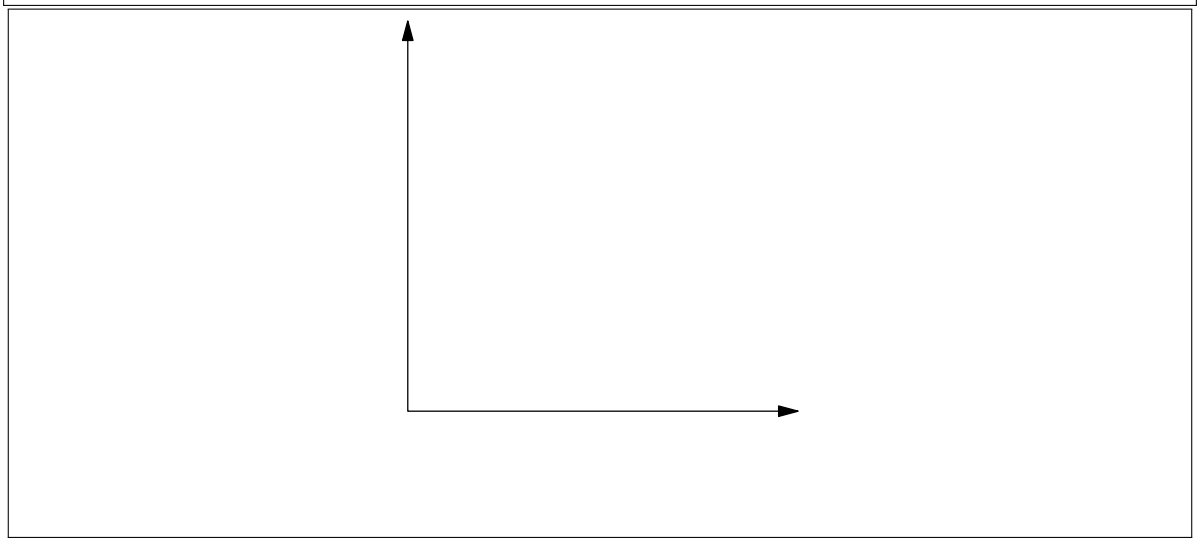
(c) Below we show a set of two-dimensional data points. Show the approximate position of the mean, $\boldsymbol{\mu}$ and sketch the contour $(\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{C}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = 1$, where $\mathbf{C}$ is the covariance matrix. Also show the direction of the leading eigenvector (principle component), $\boldsymbol{v}_1$, and the eigenvector with the second largest eigenvalue, $\boldsymbol{v}_2$. *(5 marks)*



• Do not write in this space •

TURN OVER

(d) Explain when you would use an ROC curve and sketch a typical curve, labelling the axes. *(5 marks)*



End of question 1

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

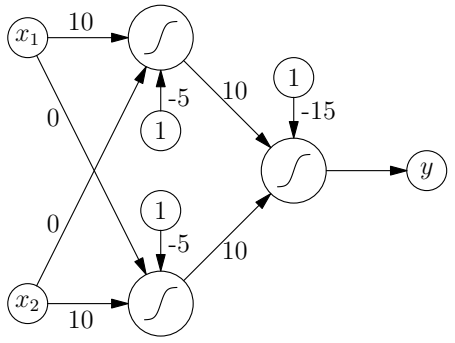
Section B

Question B 2

(a) Show that a multi-layer perceptron (MLP) with linear nodes is no more powerful than a single-layer perceptron. *(5 marks)*

(b) Describe how to train a multi-layer perceptron. *(8 marks)*

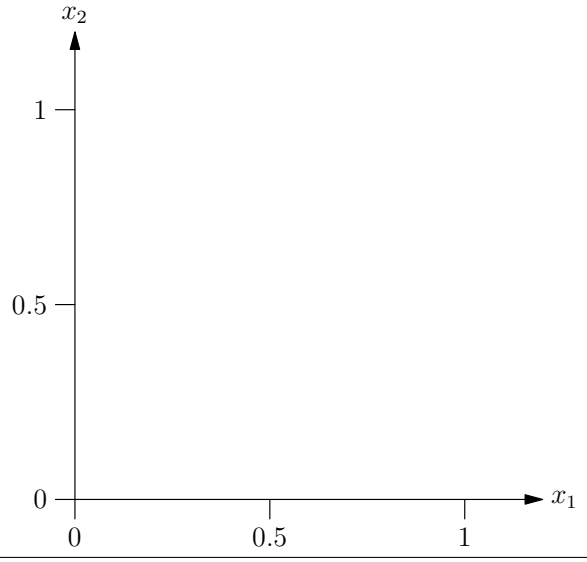
(c) For the multi-layer perceptron shown, write the response function $f(\boldsymbol{x}|\boldsymbol{w})$ explicitly as a function of the input $\boldsymbol{x} = (x_1, x_2)$. Compute the response for the inputs given in the table. Finally sketch the line where $f(\boldsymbol{x}|\boldsymbol{w}) = 1/2$. Approximate values are sufficient.



Assume that the output of the nodes are given by $g(V) = \frac{1}{1+\mathrm{e}^{-V}}$. *(12 marks)*

$f(\mathbf{x}|\mathbf{w}) =$

$\mathbf{x}$	$(0,0)$	$(0,0.5)$	$(0,1)$	$(0.5,0)$	$(0.5,0.5)$	$(0.5,1)$	$(1,0)$	$(1,0.5)$	$(1,1)$
$f(\mathbf{x}\|\mathbf{w})$									



End of question 2

Q2: (a) $\frac{}{5}$ (b) $\frac{}{8}$ (c) $\frac{}{12}$ Total $\frac{}{25}$

Question B 3

(a) Consider a finite set of hypotheses $\mathcal{H}$ for a binary classification task. If $h \in \mathcal{H}$ has an error rate ϵ, calculate the probability that it will make no error on P randomly selected (i.e. independent) patterns. *(5 marks)*

(b) Explain why the probability of any hypothesis with an error rate greater than ϵ will correctly classifying P patterns is bounded by $|\mathcal{H}|\ \mathrm{e}^{-\epsilon P}$. *(5 marks)*

(c) Obtain a bound on the number of patterns required to ensure that a consistent learner (i.e. a machine that finds a hypothesis which is consistent with all the input patterns) will have an error less than ϵ with a probability of, at least, δ. *(5 marks)*

(d) Given a hypothesis space of with $|\mathcal{H}| = 10^{10}$ hypotheses, how many patterns do you need to guarantee an error rate less that 0.1% with a probability of at least 99.99%? *(5 marks)*

(e) Explain what the VC-dimension is and why it is needed? *(5 marks)*

End of question 3

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

Question B 4

(a) A patient comes to the doctor with a symptom S. The doctor knows that all patients with a rare disease A have this symptom. However, 0.5% of patients with a common disease B also exhibit this symptom. If the probability of disease A is 10^{-6} and the probability of disease B is 10^{-3} what is the probability of the patient having disease A? (Show your working.) *(10 marks)*

(b) Explain how to set up a collaborative filter recommender system as a matrix completion problem. Describe an approximation in terms of the mean rating for users and items. Explain how to approximate the residual matrix using a low rank approximation. Show how we can introduces priors into the problem and outline how we can find a MAP solution to the problem. *(15 marks)*

End of question 4

Q4: (a) $\frac{}{10}$ (b) $\frac{}{15}$ Total $\frac{}{25}$

END OF PAPER