



MSC Examination by course unit

Friday, 3 Jun, 2016, 10:00 am

ECS708P Machine Learning

Duration: 2 hours 30 minutes

**YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL
INSTRUCTED TO DO SO BY AN INVIGILATOR**

Answer ALL FOUR Questions.

Cross out any answers that you do not wish to be marked.

Calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough workings in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately. It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

EXAM PAPERS MUST NOT BE REMOVED FROM THE EXAM ROOM

Examiners:

Ioannis Patras

Tim Hospedales

Question 1

- a) Some emails received by users are spams, attempting to gain illicit access to employee's accounts. You are building a classifier to detect such illicit emails. You start by using the feature of whether or not an email asks for the recipients password, as this an important datum indicating if the email is in fact a hacking attempt. In particular 95% of hacking emails contain requests for a user's password, and 95% of legitimate emails do not contain requests for a user's password. 1% of emails overall are hacking attempts.
- If your classifier scans an email with a password request, what is the probability that the email is in fact a hacking attempt?
 - Comment on this value that you calculate.
 - What is the probability that your classifier makes an error?

[7+2+3=12 marks]

- b) Explain the difference between Maximum Likelihood (ML) and Maximum a Posterior (MAP) methods of parameter estimation.

[4 marks]

- c) Give the equations for ML and MAP estimates for the bias of a coin.

[3 marks]

- d) Explain the underlying probabilistic assumptions that implicitly underlie (i) the least squares error function and (ii) weight regularization in linear regression.

[6 marks]**[25 marks]**

Question 2

- a) Complete this table for the methods of linear regression and logistic regression. Give (on your answer sheet) the form of the model to be learned, the cost function, and a possible method (gradient or exact) for minimising this cost function.

	Form of Model	Cost Function	Optimisation Strategy
Linear Regr.			
Logistic Regr.			

[12 marks]

- b) Consider a three-layer neural network with d input units, h hidden units, c output units, and bias. The network has activation function f , with input-to-hidden weights labelled w_{ji} and hidden-to-output weights labelled w_{kj} . For a single input vector $\mathbf{x} = (x_1, x_2, \dots, x_d)$ with associated target vector $\mathbf{t} = (t_1, t_2, \dots, t_c)$, write down an expression for:

- the output activation y_j of a hidden unit;
- the output activation z_k of an output unit (i.e. an output of the network);
- the squared error J .

[3+3+3=9 marks]

- c) Practical challenges with training neural networks include: (i) getting stuck in local optima, (ii) underfitting or overfitting. Explain the practical strategies you would use to overcome these issues.

[4 marks]

[25 marks]

Question 3:

(a) The probability density function of the Gaussian distribution is given by

$$p(x) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

Explain what each symbol represents and give its dimensionality. Sketch the probability density function in the 1 dimensional case and use it to support your explanations.

[6 marks]

(b) The maximum likelihood principle for density estimation states that the optimal parameters θ^* are the ones that minimise the following quantity:

$$E = - \sum_{n=1}^N \ln(p(x_n; \theta))$$

- i. Explain the terms in the equation above, and state what are the parameters θ in the case of a Gaussian.
- ii. Explain how the derivatives of this function can be used to optimise E in the equation above and show that in the case of the mean vector μ this leads to

$$\mu = 1/N \sum_{n=1}^N x_n$$

For simplicity you may do the derivations for the 1 dimensional case, that is, in the case that the data are scalars and the covariance matrix a scalar as well.

[4+4 = 8 marks]

(c) Explain how density estimation with a single Gaussian can be used in a practical application. Pick one of the two examples below or devise your own problem. Explain the steps carefully. For the example that you choose explain whether you could have used a Supervised Learning algorithm and, if not, why.

- i. Background modelling and subtraction in a Computer Vision application where the goal is to determine at each frame which pixels belong to the foreground and which to the background.
- ii. Abnormality detection in any context. For example in the context of consumer buying habits in a supermarket chain.

[4+4=8 marks]

- (d) In several cases the covariance matrix is singular and cannot be inverted. Why is this a problem? Give an example, perhaps with the use of a diagram, and describe a solution to that problem.

[3 marks]

[Total 25 Marks]

Question 4: Markov Models

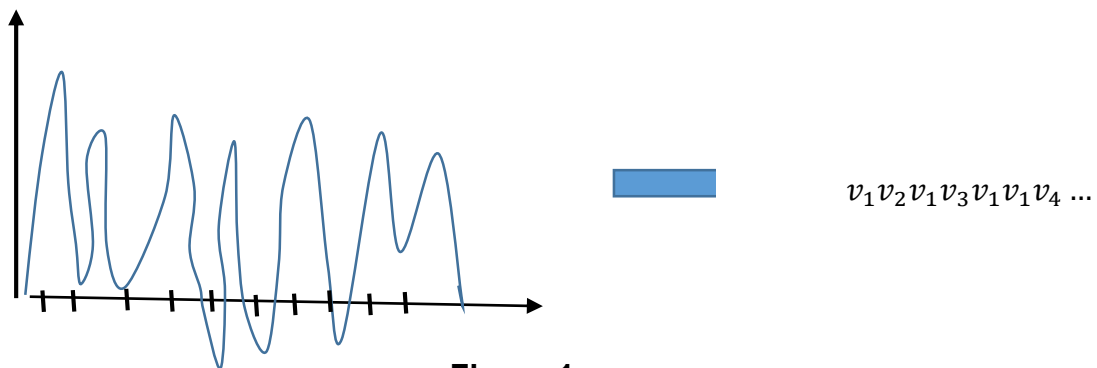
- (a) With the help of a diagram explain the main principles of the first-order Hidden Markov Model (HMM). Explain what is meant by the term “first-order”. What are the differences to a Markov model? In your answer, define the states ω_i , the symbols v_k , and the matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{jk}]$.

[6 marks]

- (b) The decoding problem can be stated as follows: Given an HMM and a sequence of observation symbols $\mathbf{V}^{1:T}$ determine the most likely sequence of hidden states $\omega^{1:T}$. What are the other two types of problems considered in the context of HMMs?

[6 marks]

- (c) You are given the task to design a speech recogniser that distinguishes between the words “yes” and “no”.
- Describe how the k-means algorithm can be used to transform a speech signal to a sequence of symbols $\{v_1, v_2, \dots, v_K\}$. An example is given in Figure 1 below:

**Figure 1**

(Hint: As a first step divide the signal in small chunks of equal and fixed length)

- Describe how you would design the training process. Make reference to what kind of data you will have, which algorithm you will use (e.g. refer to your answer to part (b)) and how many HMMs you need to train. Why can't one use a simple classifier, such as a linear logistic regression scheme?
- Describe how you would use the HMMs that you trained in order to make decisions at test time.

[4+4+5 = 13 marks]**[Total 25 marks]**