SCHOOL OF ENGINEERING

ADVANCED CONCEPTS IN SIGNAL PROCESSING

PGEE11020

Exam Date:	**21/05/2013**	From and To:	**1430-1630**	Exam Diet:	**May**

Please read full instructions before commencing writing

Exam paper information

- Paper consists of 2 Sections
- Candidates to answer THREE questions
- Section A: (One question) Answer whole Section
- Section B: Answer TWO out of THREE questions

Special instructions,

- Students should assume reasonable values for any data not given in a question nor available on a datasheet, and should make any such assumptions clear on their script.
- Students in any doubt as to the interpretation of the wording of a question, should make their own decision, and should state it clearly on their script.
- Please write your name in the space indicated at the right hand side on the front cover of the answer book. Also enter you examination number in the appropriate space on the front cover.
- Write **ONLY** your examination number on any extra sheets or worksheets used and firmly attach these to the answer book(s).
- This examination will be marked anonymously.

Special items

- None

Convenor of Board of Examiners: **Professor B Mulgrew**
External Examiner: **Professor D Bull**

SECTION A

Question A1

a) Write down the expression for a fixed step size gradient descent step for optimizing a quadratic cost function $J(\boldsymbol{a})$ of the vector $\boldsymbol{a}$. Show how it can be used to solve the optimization problem. **(4)**

b) Summarize the steps involved in Principal Component Analysis (PCA). Define any terms that you use. **(4)**

c) What precisely is meant by the term linearly separable when applied to (i) a 2-class and (ii) a multi-class problem. **(4)**

d) Describe the key steps in the k-means clustering algorithm. Indicate how it can be used to group a scalar sequence into two classes. **(4)**

e) What is the Levenberg-Marquardt optimization strategy? How can it be used to speed up learning in conjunction with the back propagation algorithm? **(4)**

SECTION B

Question B1

a) Explain precisely the difference between a Markov model and a hidden Markov model (HMM). **(2)**

b) Write down an emission matrix for a 4-state Markov model that is not hidden. **(2)**

c) Derive the following *backward* HMM algorithm for the evaluation of probabilities.

$$\beta_i(t) = \sum_j \beta_j(t+1) a_{ij} b_{jk}$$

The observed sequence is given by $\mathbf{V}^T = \{v(1),\ v(2), \dots, v(T)\}$.
The sequence of states is given by $\boldsymbol{\omega}^{\boldsymbol{T}} = \{\omega(1),\ \omega(2),\ \dots, \omega(T)\}$.
The Markov model is characterised by the transition probabilities such as $P\big(\omega_j(t+1)|\omega_i(t)\big) = a_{ij}$,where $\omega_i(t)$ is the state value at time t, and emission probabilities such as $P\big(v_k\,(t)|\ \omega_j(t)\big) = b_{jk}$, where $v_k\,(t)$ is the observation at time t.

The starting point for the derivation should be the sequence probability

$$P(\mathbf{V}^T) = \sum_r \prod_{t=1}^{T} P(v(t)|\omega(t))P(\omega(t)|\omega(t-1))$$

where r enumerates all possible state sequences. **(8)**

d) A HMM model is defined by the transition matrix

$$\mathbf{A} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0 & 0 & 1 \end{bmatrix}$$

and the emission matrix

$$\mathbf{B} = \begin{bmatrix} 0.1 & 0 & 0.9 \\ 0 & 1 & 0 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

Use the *backward* algorithm to evaluate the probability of the observed sequence $\{v_2,\ v_3, v_2\}$ conditioned on the knowledge that the system starts in state 2 at time zero. Here the visible states are indexed from 1, i.e. the 3 possible visible states are, v_1, v_2 and v_3 **(8)**

Question B2

a) What precisely is a linear machine for a Kclass problem? **(2)**

b) The following data vectors are from 3 classes

$$\omega_1: [1 \quad 1], [1 \quad -1]$$

$$\omega_2: [-1 \quad 1], [-1 \quad -1]$$

$$\omega_3: [0.25 \quad 2], [0.25 \quad 3]$$

Use the Pseudo-inverse method to design a linear machine for this problem. You are told that a bias term is not required. Verify whether or not the linear machine has separated the 3 classes. **(12)**

c) For a 2-class problem in 2 dimensions we wish to separate the classes using a linear classifier of the form $y = \mathbf{w}^{\boldsymbol{T}}\,\mathbf{x}$ where the classifier weight vector is given by $\mathbf{w}^{\boldsymbol{T}} = [1 \quad 1]$. The class densities are Gaussian mixtures given by

$$p(\mathbf{x}|\omega_1) = \frac{1}{2}\sum_{i=1}^{2} N(\mathbf{x}; \mathbf{x}_i, \boldsymbol{\Sigma})$$

$$p(\mathbf{x}|\omega_2) = \frac{1}{2}\sum_{i=3}^{4} N(\mathbf{x}; \mathbf{x}_i, \boldsymbol{\Sigma})$$

Where the multi-dimensional Gaussian density is

$$N(\mathbf{x}; \mathbf{x}_i, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}\,|\boldsymbol{\Sigma}|} e^{-(\mathbf{x}-\mathbf{x}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{x}_i)/2}$$

with covariance matrix, $\boldsymbol{\Sigma} = \frac{1}{10}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and means $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$,

$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$,$\mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ and $\mathbf{x}_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. Derive an expression for the probability of misclassification in terms of the cumulative distribution function (CDF), $F(x)$, of a standard Gaussian distribution. This CDF is defined as:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\,\mathrm{d}z$$

(6)

Indicate clearly any assumptions that you make in the derivation.

Question B3

a) Explain what is meant by unsupervised learning and explain how this relates to mixture models. **(3)**

b) The general formula for the probability density function (pdf) of an observation from a mixture model with classes c_k that are defined in terms of the parameters θ_k is

$$p(x|\,\theta) = \sum_{k=1}^{K} p(c_k) g(x|c_k\,, \theta_k)$$

where $p(c_k)$ are the class priors and $g(x|c_k\,, \theta_k)$are the individual class conditional density functions. Write down an expression for the log likelihood function associated with the observed sequence, $x_1, x_2, \dots, x_N$ stating any assumptions that you make. **(2)**

c) Show that the Maximum Likelihood estimate of the parameters given this observed sequence must satisfy the following relationship. **(5)**

$$\sum_{n=1}^{N} p(c_k|x_n\,, \theta_k) \frac{\partial}{\partial \theta_k} \ln p(x_n|c_k\,, \theta_k) = 0$$

d) Explain the role that this equation plays in Expectation Maximization. **(2)**

e) Assume that the observed sequence in part (b) is a scalar one drawn from a Mixture of Gaussians.

i. Using Bayes theorem (or otherwise) derive the Expectation step for calculation of the Membership Density $p(c_k|x_n\,, \theta_k)$. **(4)**

ii. Using the equation of part (c) derive the Maximization Step for calculation of the mean μ_k associated with the k^{th} class. **(4)**

END OF PAPER