SCHOOL OF ENGINEERING

ADVANCED CONCEPTS IN SIGNAL PROCESSING

PGEE11020

Exam Date: **01/05/15** From and To: **1430-1630** Exam Diet: **May 2015**

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

a) Explain the terms *supervised learning* and *unsupervised learning*, indicating their role in probabilistic modelling. Illustrate your answer with examples. **(5)**

b) Explain what is meant by the term *Maximum Likelihood* (ML) learning. Indicate any issues with this technique and explain how it differs from Bayesian learning techniques? **(4)**

c) Write down Bayes decision rule and explain its significance in classification problems. **(2)**

d) Write down the log likelihood for the data $\{x_i\}$, $i = 1, \dots, N$, drawn from a scalar Gaussian mixture model with K equal weights but unknown means, μ_k, and variances, σ_k^2, stating any assumptions made. Explain the difficulty in directly learning the unknown parameters for this model. **(4)**

e) Write down the Expectation Maximization (EM) algorithm for the model in part (d) and indicate how it resolves these difficulties. **(5)**

SECTION B

Question B1

a) Sketch a diagram of a standard 2-category, 3-layer multilayer perceptron (MLP) with inputs $x_1, \ldots, x_d$, n_H hidden units, a bias unit and a single output unit, z. **(3)**

b) Assuming that the activation functions for the network in part (a) have the form $f(u) = \tanh(u)$, write down the input-output relationship for the overall network in terms of the input-to-hidden weights, $\mathbf{w}$, and the hidden-to-output weights, $\tilde{\mathbf{w}}$. **(3)**

c) Explain the purpose of the *Backpropagation of errors* algorithm and identify its function in terms of optimization. **(2)**

d) Suppose the network were to be trained using a sum of squared errors measure and is presented with N input vectors $\mathbf{x}(\mathbf{n})$, $n = 1, \ldots, N$ with target values $t(n)$. Using the result in part (b) or otherwise, derive the explicit backpropagation update equations for the unknown bias weights for both the output unit and the n_H hidden units.

[*In your answer you may use the fact that:* $\frac{d}{du}\tanh(u) = \text{sech}^2(u)$.] **(8)**

e) How would the learning rules be changed if the following sum of absolute errors measure was used instead:

$$J = \sum_{n=1}^{N} |z(n) - t(n)|$$

where $z(n)$ is the network output for the nth input vector? **(2)**

f) Explain the term weight sharing and indicate how it can help in training a MLP. **(2)**

Question B2

a) Explain how the Mean Square Error (MSE) criterion, $J(\mathbf{w})$, can be used to learn the weights of a linear discriminant function even when the samples are not linearly separable. **(3)**

b) Consider the data points:

$$\omega_1 : (2,-1)(-1,-1)(-2,-1)$$

$$\omega_2 : (1,-1)(0,4)$$

Write down the augmented data matrix Y for these samples and hence show that $Y^T Y$ is diagonal. **(4)**

c) Using target outputs::

$$t(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \omega_1 \\ -1, & \text{if } \mathbf{x} \in \omega_2 \end{cases}$$

calculate the minimum MSE weights for the data in part (b) and identify which of the samples are correctly classified. **(6)**

d) Given multi-category linear discriminant functions $g_k(\mathbf{x}; \mathbf{w})$ with target outputs:

$$t(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \omega_k \\ 0, & \text{otherwise} \end{cases}$$

show that the expected MSE can be written as:

$$\mathbb{E}_{p(\mathbf{x})}(J(\mathbf{w})) = \int p(\mathbf{x}) \left(g_k(\mathbf{x}; \mathbf{w}) - p(\omega_k|\mathbf{x})\right)^2 d\mathbf{x} + \text{ constant}$$

where the *constant* term is independent of $\mathbf{w}$. **(5)**

e) Indicate the extent to which the result in part (d) can be interpreted in terms of Bayes Decision Theory. **(2)**

Question B3

a) Describe the key components of a Hidden Markov Model (HMM) based isolated word recognition system, identifying their specific roles. **(5)**

b) Let $\mathbf{A}^{(yes)}$ and $\mathbf{B}^{(yes)}$, below, define the state transition probabilities and emission probability matrix for a 4-state HMM for the word 'yes':

$$\mathbf{A}^{(yes)} = [\mathbf{a}_{ij}^{(yes)}] = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 1 \end{pmatrix} ; \mathbf{B}^{(yes)} = [\mathbf{b}_{jk}^{(yes)}] = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.0 \\ 0.2 & 0.7 & 0.1 & 0.0 \\ 0.1 & 0.2 & 0.7 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.9 \end{pmatrix}$$

where the indices i, j and k range from 1 to 4.

Determine the possible distinct state sequences that could generate the observed sequence $\{v(1), v(2), v(3), v(4)\} = v_2 v_2 v_2 v_4$, assuming that the starting state at time $t = 0$ is ω_1. How many possible state sequences are there? **(4)**

c) Given a second HMM defined below for the word 'no', determine which of the two HMMs is more likely to have generated the observed sequence in part (c). Justify your answer.

$$\mathbf{A}^{(no)} = [\mathbf{a}_{ij}^{(no)}] = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{pmatrix} ; \mathbf{B}^{(no)} = [\mathbf{b}_{jk}^{(no)}] \begin{pmatrix} 0.2 & 0.3 & 0.5 & 0.0 \\ 0.5 & 0.3 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

(7)

d) Describe the principles of HMM parameter selection. **(4)**

END OF PAPER