EE 657 : End Semester Examination

Duration: 3 hours **Marks: 40** **Date: May 3' 2017**

Note: This paper comprises seven questions. Ensure that you present in a logical sequence all important steps that lead to the solution of the problem. Remember to transfer the solutions to the answer sheet being provided.

1. The following ten *i.i.d* one-dimensional training observations were collected for ω_1 and ω_2 respectively.

$$\omega_1 : \{0, 1, 2, 7, 8\} \qquad \omega_2 : \{3, 4, 5, 6, 9\}.$$

Through an iterative scheme, we consider building an ensemble of weak classifiers using the Adaboost algorithm. All the ten training samples are initially assigned a weight of 0.1. Each weak learner in the algorithm provides decision boundaries that are selected by a threshold.

(a) Choose a weak classifier/ threshold that minimizes the training error rate e_1, in the first iteration. What is the value of e_1.

(b) Compute the normalized weights for the misclassified training samples, that are to be used for the selection of the second weak classifier. (Assume that only the weights of wrongly classified samples get scaled by a factor $\frac{1-e_1}{e_1}$).

[1.5 + 2.5 = 4 marks]

2. You are given a set of 4 vectors $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}$

(a) Represent the vector $\mathbf{X} = \begin{bmatrix}1\\1\\3\\2\end{bmatrix}$ as a linear combination of these vectors

(b) If instead, we represent the vector $\mathbf{X}$ as a linear combination of the first 3 vectors, what will be the approximated vector.

(c) Compute the approximation error between $\mathbf{X}$ and the answer obtained in (b) using the Euclidean distance.

Note: `Credit will be given only for the easiest way of solving the problem. Analyze the 4 given vectors !! Do NOT start with Cramer's rule or the like and waste time.`

[2 + 2 + 2 = 6 marks]

3. Consider a one-dimensional two-category classification problem with prior $P(\omega_1) = 0.6$ and the densities of the form

$$p(x|\omega_1) = \frac{1}{2}e^{-x/2} \qquad x \geq 0$$

$$p(x|\omega_2) = \frac{1}{6}e^{-x/6} \qquad x \geq 0$$

Assuming zero-one loss function, write down expressions of the average probability that a sample :

(a) from ω_1 gets misclassified to ω_2

(b) from ω_2 gets misclassified to ω_1

correctly classified.

(c) from ω_1 gets ~~misclassified~~ to ω_1

(d) from ω_2 gets ~~misclassified~~ to ω_2

correctly classified

[2+ 1 + 1 + 1 = 5 marks]

4. During the training of a CNN, a filtered 4×4 patch is obtained as

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 4 & -8 \\ 1 & 3 & -4 & -7 \\ -1 & 7 & -1 & 3 \\ -2 & 1 & 4 & 9 \end{bmatrix} \quad \begin{bmatrix} 7 & 7 \\ 7 & 9 \end{bmatrix}$$

1.444

$13/9 \quad 0$

$8/9 \quad 15/9$

What will the output when

(a) the ReLU activation is applied on this patch.

(b) max pooling is applied on the patch obtained in part(a) , reducing it to 2×2 size

(c) average pooling is applied on the patch obtained in part(a) , reducing it to 2×2 size

(d) the logistic function is applied to the element marked in **bold** in $\mathbf{X}$.

[1.5 + 1.5 + 1.5 + 1 = 5.5 marks]

5. Consider the training of a GMM of 2 Gaussians with unknown means, variances and weights. The following table represents the responsibility probability of each of the one-dimensional training sample in $\mathbf{X} = \{2, 5, 7, 9\}$ to the first Gaussian component at a given iteration.

Sample	2	5	7	9
Probability	0.15	0.41	0.69	0.55

0.15

0.69

0.41

0.15

1.8

Compute the updated estimate of the following that are to be used for the subsequent iteration.

(i) the mean of the first Gaussian component

(ii) the variance of the first Gaussian component

(iii) the variance of the second Gaussian component

(iv) the weight of the first Gaussian component

(v) the weight of the second Gaussian component

[1.5 + 2 + 2.5 + 1 + 1 = 8 marks]

6. You are given 3 training examples (comprising 2 dimensions $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$) for classes ω_1 and ω_2 with labels $y_1 = 1$ and $y_2 = -1$. A single layer perceptron is used for the learning.

$$\omega_1 : \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad \omega_2 : \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

(a) At a given iteration, if $x_1 + x_2 - 2 = 0$ is used as a decision boundary, specify the misclassified samples (if any). Clearly state their corresponding value of the discriminant function.

(b) Recall that, in a perceptron learning, we update the weights by accumulation of sample feature vectors being misclassified. Use this knowledge to obtain the decision boundary for the subsequent iteration

(c) Based on the nature of the training samples, guess the number of iterations required for convergence of the perceptron algorithm.

[2 + 2.5 + 1 = 5.5 marks]

7. (a) Assume that we have a single training pair $(\mathbf{x}_1, y_1)$, where $\mathbf{x}_1 \neq 0$. If we used a perceptron algorithm to adjust the weights, defined by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \eta \frac{e_k \mathbf{x}_1}{\mathbf{x}_1^T \mathbf{x}_1}$$

where $e_k = (y_1 - \mathbf{w}_k^T \mathbf{x}_1)$. Find an expression for e_{k+1} as a function of e_k and η.

[1 mark]

(b) Consider an HMM representation of a coin-tossing experiment. You are given a three state model (corresponding to three different coins) with probabilities

	State 1	State 2	State 3
P(H)	0.5	0.6	0.4
P(T)	0.5	0.4	0.6

and with state transition probabilities set as: .

	State 1	State 2	State 3
State 1	0.2	0.4	0.4
State 2	0.4	0.2	0.4
State 3	0.4	0.4	0.2

The initial probabilities for state 1, state 2 and state 3 are same. Given the observation sequence $\mathbf{O} = HHT$, of length three,

i. What is the probability that this sequence came from state 1 entirely?

ii. Using the Viterbi algorithm, what is the optimal sequence of states that can be assigned to $\mathbf{O}$? (`Remember to back-trace` !!)

[1.5 + 2.5 = 4 marks]

(c) Check whether $\begin{bmatrix} 1 & 4 \\ 4 & 3 \end{bmatrix}$ is a valid covariance matrix with proper justification.

[1 mark]