EE 626 : Pattern Recognition and Machine Learning

Mid-Semester Examination

Duration: 2 hours   Date: Mar 2' 2025   Marks: 35

Please outline all the steps systematically to get full marks

1. Consider a one-dimensional two-category classification problem with prior $P(\omega_1) = 0.4$ and continuous class conditional densities of the form

$$p(x|\omega_i) = \theta_i e^{-\theta_i x} \qquad x \geq 0$$

The parameters θ_i for $i = 1, 2$ are positive but unknown. The prior distribution of the parameters is of the form

$$p(\theta_i) = \frac{1}{\sqrt{\pi\theta_i}} e^{-\theta_i}$$

Consider zero one loss function.

(i) The following i.i.d training observations were collected: $\mathbb{D}_1 = \{1, 2, 5\}$ and $\mathbb{D}_2 = \{3, 6, 7\}$ for ω_1 and ω_2, respectively. Using this information, find the MAP estimate values for θ_1 and θ_2.

(ii) Based on your answer of the estimate to part (i), determine the Bayes threshold (decision boundary) x^*.

(iii) Compute the probability of error.

[5 + 3 + 4.5 = 12.5 marks]

2. (a) Consider a training set with 100 apples, 150 bananas and 200 mango samples. Based on the values of an attribute , we split it to 2 nodes using the decision tree classifier. The first node contains 60 apples, 10 bananas and 40 mango samples. Compute the information gain resulting from the split

(b) Consider 2 matrices $\mathbf{A} = \mathbf{X}\mathbf{X}^\mathbf{T}$ and $\mathbf{B} = \mathbf{X}^T\mathbf{X}$, where $\mathbf{X}$ is of size $m \times n$ $(m \geq n)$. Derive a relationship between the eigenvalues and eigenvectors of $\mathbf{A}$ and $\mathbf{B}$.

[4 + 2 = 6 marks]

3. (a) Given the 2-dimensional data for two classes ω_1 and ω_2 :

$$\omega_1 : \{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \} \qquad \omega_2 : \{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix} \}$$

Compute the optimal weight vector $\mathbf{w}$ (projection line) in a reduced dimension using the Fisher criterion.

(b) Consider a two-dimensional data set D, obtained by pooling the training samples from 2 classes ω_1 and ω_2. It is desired to learn three centroids by means of a k-means algorithm.

$$\mathbf{D} = \{ \begin{bmatrix} -6 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -8 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \end{bmatrix} \}$$

The algorithm is initialized with cluster means: $\boldsymbol{\mu}_1 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$, $\boldsymbol{\mu}_2 = \begin{bmatrix} -2 \\ -7 \end{bmatrix}$ and $\boldsymbol{\mu}_3 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$. What are the values of the centroids in the next iteration?

[4 + 2.5 = 6.5 marks]

4. A one-dimensional Gaussian Mixture Model, is defined below:

$$p(x|\omega) = \tfrac{1}{4}N(2,1) + \tfrac{1}{2}N(3,1) + \tfrac{1}{4}N(5,2)$$

(i) If a data point is to be generated from this GMM, what will the probability of choosing the first or second Gaussian density

(ii) Compute the responsibilities of each of the three Gaussian densities for a sample point $x = 2.5$.

(iii) Write down the expression of the log likelihood function by considering a set of N training examples $\{x_1, x_2, x_3, ...x_N\}$

The normal distribution is given by

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

[2 + 6 + 2 = 10 marks]