Indian Institute of Technology Kharagpur
Department of Electrical Engineering

Subject No.: EE60020 **Subject:** Machine Learning for Signal Processing
Date of Assignment: 12 February 2024 **Semester:** Spring 2023-24
Assignment Number: 2 Solution **Duration:** 1 hour 50 mins **Full points:** 140

Name: ____________________ **Roll No:** ____________________

1. Here $\mathcal{I}$ is an image in RGB representation format and $\mathbf{\Omega}$ is the class labels associated with each pixel in it.

$$\mathcal{I} = \begin{bmatrix} (1,3,1) & (2,2,1) & (5,6,6) & (6,5,5) \\ (3,1,2) & (1,3,3) & (4,4,4) & (4,6,4) \end{bmatrix} \quad \mathbf{\Omega} = \begin{bmatrix} \omega_1 & \omega_1 & \omega_0 & \omega_0 \\ \omega_1 & \omega_1 & \omega_0 & \omega_0 \end{bmatrix}$$

When each pixel is represented as a sample $\mathbf{x}_i \in \mathbb{R}^{D\times 1}$ such that the complete dataset is represented as $\mathbf{X} \in \mathbb{R}^{D\times N}$, then

(a) (4 points) Write $\mathbf{X}$ corresponding to $\mathcal{I}$?

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 2 & 1 & 5 & 4 & 6 & 4 \\ 3 & 1 & 2 & 3 & 6 & 4 & 5 & 6 \\ 1 & 2 & 1 & 3 & 6 & 4 & 5 & 4 \end{bmatrix}$$

(b) (6 points) When solving the Bayes' decision rule employing maximum *aposteriori* probability (MAP) employing the maximum likelihood estimation (MLE) with a multivariate Normal distribution, find?

$$\boldsymbol{\mu}_{\omega_0} = \begin{bmatrix} 4.75 & 5.25 & 4.75 \end{bmatrix}^\top$$
$$\boldsymbol{\mu}_{\omega_1} = \begin{bmatrix} 1.75 & 2.25 & 1.75 \end{bmatrix}^\top$$

(c) (9 points) Compute the covariance matrix $\mathbf{\Sigma}_{\omega_0}$?

$$\mathbf{\Sigma}_{\omega_0} = \begin{bmatrix} 0.6875 & 0.0625 & 0.4375 \\ 0.0625 & 0.6875 & 0.3125 \\ 0.4375 & 0.3125 & 0.6875 \end{bmatrix}$$

(d) (9 points) Compute the covariance matrix $\mathbf{\Sigma}_{\omega_1}$?

In the traditional approach, computing the covariance matrix directly may lead to a non-invertible matrix. However, for Bayesian estimation, it is necessary to obtain the inverse of the covariance matrix for each class. To address this, PCA is applied to $\mathbf{X}_{\omega_1}$, creating a new representation of the data. Subsequently, the covariance matrix is calculated based on this transformed data representation.

$$\boldsymbol{\Sigma}_{\omega_1} = \begin{bmatrix} 1.386 & 0 & \underline{\qquad} \\ 0 & 0.676 & \underline{\qquad} \\ \underline{\qquad} & \underline{\qquad} & \underline{\qquad} \end{bmatrix}$$

(e) (12 points) Compute the square of the Mahalanobis distance for the class ω_0 for the sample $\mathbf{x} = (2, 2, 2)$?

$$d_{\omega_0} = (\mathbf{x} - \boldsymbol{\mu}_{\omega_0})^\top \boldsymbol{\Sigma}_{\omega_0}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\omega_0}) = \begin{bmatrix} -2.75 & -3.25 & -2.75 \end{bmatrix} \begin{bmatrix} 2.66 & 0.66 & -2 \\ 0.66 & 2 & -1.33 \\ -2 & -1.33 & 3.33 \end{bmatrix} \begin{bmatrix} -2.75 \\ -3.25 \\ -2.75 \end{bmatrix}$$

(f) (12 points) Compute the square of the Mahalanobis distance for the class ω_1 for the sample $\mathbf{x} = (2, 2, 2)$?

Here we need to project our data points in the eigenvector space as well to calculate the distance.

$$d_{\omega_1} = (\mathbf{x} - \boldsymbol{\mu}_{\omega_1})^\top \boldsymbol{\Sigma}_{\omega_1}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\omega_1}) = \begin{bmatrix} -0.25 & 1.98 & \underline{\qquad} \end{bmatrix} \begin{bmatrix} 0.721 & 0 & \underline{\qquad} \\ 0 & 1.478 & \underline{\qquad} \\ \underline{\qquad} & \underline{\qquad} & \underline{\qquad} \end{bmatrix} \begin{bmatrix} -0.25 \\ 1.98 \\ \underline{\qquad} \end{bmatrix}$$

(g) (1 point) Compute the prior probability per class

$$p(\Omega = \omega_0) = \frac{1}{2}, \quad p(\Omega = \omega_1) = \frac{1}{2}$$

(h) (2 points) Compute the likelihood of the sample $\mathbf{x} = (2, 2, 2)$ for class ω_0

$$p(\mathbf{x}|\Omega = \omega_0) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_{\omega_0}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} d_{\omega_0}\right) = 8.82 \times 10^{-7}$$

$$D = 3, \quad |\boldsymbol{\Sigma}_{\omega_0}| = 0.1406, \quad d_{\omega_0} = 24.33$$

(i) (2 points) Compute the likelihood of the sample $\mathbf{x} = (2, 2, 2)$ for class ω_1

$$p(\mathbf{x}|\Omega = \omega_1) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_{\omega_1}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} d_{\omega_1}\right) = 9.6 \times 10^{-3}$$

$$|\boldsymbol{\Sigma}_{\omega_1}| = 0.9375, \quad d_{\omega_1} = 5.667$$

(j) (1 point) Find the evidence for the sample $\mathbf{x} = (2, 2, 2)$

$$p(\mathbf{x}) = p(\Omega = \omega_0) p(\mathbf{x}|\Omega = \omega_0) + p(\Omega = \omega_1) p(\mathbf{x}|\Omega = \omega_1) = 9.6 \times 10^{-3}$$

(k) (2 points) Find the posterior probability for each class for the sample $\mathbf{x} = (2, 2, 2)$

$$P(\Omega = \omega_0|\mathbf{x}) = \frac{p(\Omega = \omega_0)p(\mathbf{x}|\Omega = \omega_0)}{p(\mathbf{x})} = 4.56 \times 10^{-5}, \quad P(\Omega = \omega_1|\mathbf{x}) = \frac{1}{2}$$

2. Considering the dataset and class labels provide earlier in Q. 1 we build a Binary Decision Tree (BDT)

(a) (5 points) Compute the histogram ($q(\cdot)$), prior probability ($p(\cdot)$) and Entropy ($H(\cdot)$) for the samples of $\mathbf{X}$ arriving at the root node n_0 of the BDT

$$q(\omega_0; n_0) = 4, \quad q(\omega_1; n_0) = 4$$

$$p(\omega_0; n_0) = \frac{q(\omega_0; n_0)}{q(\omega_0; n_0) + q(\omega_1; n_0)} = \frac{1}{2}$$

$$p(\omega_1; n_0) = \frac{q(\omega_1; n_0)}{q(\omega_0; n_0) + q(\omega_1; n_0)} = \frac{1}{2}$$

$$H(n_0) = -p(\omega_0; n_0)\log_2(p(\omega_0; n_0)) - p(\omega_1; n_0)\log_2(p(\omega_1; n_0)) = 1 \text{ bits}$$

(b) (15 points) Let us consider the samples at node n_j to be bifurcated using axis aligned split function defined as

$$S(\mathbf{x}; \phi_{n_j}, \theta_{n_j}) = \begin{cases} 1 & \text{if } x^{(\phi_{n_j})} > \theta_{n_j} \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{x} = [x^{(0)}, x^{(1)}, \cdots, x^{(d)}, \cdots, x^{(D-1)}]$, $\phi_{n_j} \in \{0, 1, \cdots, d, \cdots, D-1\}$, $\theta_{n_j} \in \mathbb{R}$; such that $S(\mathbf{x}; \phi_{n_j}, \theta_{n_j}) = 1$ is associated with bifurcation of the samples to the right child of n_j denoted by $n_j \to R$, and $S(\mathbf{x}; \phi_{n_j}, \theta_{n_j}) = 0$ is associated with bifurcation of the samples to the left child of n_j denoted by $n_j \to L$. We consider a set of randomly chosen axis denoted by $\{\phi_{(k)}\}$ from which an optimal one would represent ϕ_{n_j}, and also a set of values $\{\theta_{(k),l}\}$ from which an optimal one would represent θ_{n_j}. When $\phi_{(k=0)} = 0$ and $\theta_{(k=0),l=0} = 2$, compute the following

$$q(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 3, \quad q(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 5$$

$$q(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 0, \quad q(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 3$$

$$q(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 4, \quad q(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = 1$$

$$p(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})} = \frac{3}{8}$$

$$p(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})} = \frac{5}{8}$$

$$p(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})} = 0$$

$$p(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})} = 1$$

$$p(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})} = \frac{4}{5}$$

$$p(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \frac{q(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})}{q(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) + q(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})} = \frac{1}{5}$$

$$\begin{aligned} H(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = & -p(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) \log_2 \left(p(\omega_0; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})\right) \\ & -p(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) \log_2 \left(p(\omega_1; n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0})\right) \\ = & \quad 0 \text{ bits} \end{aligned}$$

$$\begin{aligned} H(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = & -p(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) \log_2 \left(p(\omega_0; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})\right) \\ & -p(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) \log_2 \left(p(\omega_1; n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})\right) \\ = & \quad 0.541 \text{ bits} \end{aligned}$$

$$\begin{aligned} \mathrm{IG}(n_0; \phi_{(k=0)}, \theta_{(k=0),l=0}) = & \; H(n_0) - \big(p(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) H(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) \\ & + p(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) H(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})\big) \\ = & \; 1 - \left(\left(\frac{3}{8}\right)(0) + \left(\frac{5}{8}\right)(0.541)\right) = 0.661 \text{ bits} \end{aligned}$$

(c) (5 points) Find the optimal value of ϕ_{n_0} and θ_{n_0} based on $\mathrm{IG}(\cdot)$(after completing (b) and (d))?

$\phi_{(k=0)} = 0$	$\phi_{(k=1)} = 2$
$\theta_{(k=0),l=0} = 2$ $\mathrm{IG}(\cdot) = 0.661$	$\theta_{(k=1),l=0} = 3.76$ $\mathrm{IG}(\cdot) = 1$
$\theta_{(k=0),l=1} = 1.12$ $\mathrm{IG}(\cdot) = 0.311$	$\theta_{(k=1),l=1} = 2.25$ $\mathrm{IG}(\cdot) = 0.548$

$$\phi_{n_0} = \phi_{(k=1)} = 2, \quad \theta_{n_0} = \theta_{(k=1),l=0} = 3.76$$

(d) (45 points) Compute the above set of parameters for the following randomly selected splits

	$\phi_{(k=0)} = 0$, $\theta_{(k=0),l=1} = 1.12$	$\phi_{(k=1)} = 2$, $\theta_{(k=1),l=0} = 3.76$	$\phi_{(k=1)} = 2$, $\theta_{(k=1),l=1} = 2.25$
$q(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	2	4	3
$q(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	6	4	5
$q(\omega_0; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	0	0	0
$q(\omega_1; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	2	4	3
$q(\omega_0; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	4	4	4
$q(\omega_1; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	2	0	1
$p(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{8}$
$p(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{5}{8}$
$p(\omega_0; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	0	0	0
$p(\omega_1; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	1	1	1
$p(\omega_0; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	$\frac{2}{3}$	1	$\frac{4}{5}$
$p(\omega_1; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	$\frac{1}{3}$	0	$\frac{1}{5}$
$H(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$	0	0	0
$H(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$	0.9182	0	0.72
$\mathrm{IG}(n_0; \phi_{(k)}, \theta_{(k),l})$	0.311	1	0.5487

(e) (8 points) Write the set of samples in $n_0 \to L$ and $n_0 \to R$ after bifurcation of samples in n_0 using the optimal split identified in (c).

$n_0 \to L$	$n_0 \to R$
(1, 3, 1)	(5, 6, 6)
(3, 1, 2)	(4, 4, 4)
(2, 2, 1)	(6, 5, 5)
(1, 3, 3)	(4, 6, 4)
(____, ____, ____)	(____, ____, ____)
(____, ____, ____)	(____, ____, ____)
(____, ____, ____)	(____, ____, ____)
(____, ____, ____)	(____, ____, ____)

(f) (2 points) Comment on the nature of nodes (strike out the incorrect one within the braces)
$n_0 \to L$ is a ~~{Decision node}~~ / {Leaf node with $\Omega = \{\omega_1\}$}
$n_0 \to R$ is a ~~{Decision node}~~ / {Leaf node with $\Omega = \{\omega_0\}$}

——— End of assignment. Use the remaining pages for rough work. ———

This page is intentionally left blank for rough work.

This page is intentionally left blank for rough work.