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** **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 $\mathcal{I}$?

$$\mathbf{X} = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix}$$

(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?

$$\boldsymbol{\mu}_{\omega_0} = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix}^\top$$

$$\boldsymbol{\mu}_{\omega_1} = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix}^\top$$

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

$$\mathbf{\Sigma}_{\omega_0} = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix}$$

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

$$\mathbf{\Sigma}_{\omega_1} = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix}$$

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

$$d_{\omega_0} = (\mathbf{x} - \boldsymbol{\mu}_{\omega_0})^\top \mathbf{\Sigma}_{\omega_0}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\omega_0}) = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix} \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix} \begin{bmatrix} \_\_\_ \\ \_\_\_ \\ \_\_\_ \end{bmatrix}$$

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

$$d_{\omega_1} = (\mathbf{x} - \boldsymbol{\mu}_{\omega_1})^\top \mathbf{\Sigma}_{\omega_1}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\omega_1}) = \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix} \begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \\ \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix} \begin{bmatrix} \_\_\_ \\ \_\_\_ \\ \_\_\_ \end{bmatrix}$$

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

$$p(\Omega = \omega_0) = \_\_\_, \quad p(\Omega = \omega_1) = \_\_\_$$

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

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

$$D = \_\_\_, \quad |\mathbf{\Sigma}_{\omega_0}| = \_\_\_, \quad d_{\omega_0} = \_\_\_$$

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

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

$$|\mathbf{\Sigma}_{\omega_1}| = \_\_\_, \quad d_{\omega_1} = \_\_\_$$

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

$$p(\mathbf{x}) = p(\Omega = \omega_0) p(\mathbf{x}|\Omega = \omega_0) + p(\Omega = \omega_1) p(\mathbf{x}|\Omega = \omega_1) = \_\_\_$$

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

$$P(\Omega = \omega_0|\mathbf{x}) = \frac{p(\Omega = \omega_0) p(\mathbf{x}|\Omega = \omega_0)}{p(\mathbf{x})} = \_\_\_, \quad P(\Omega = \omega_1|\mathbf{x}) = \_\_\_$$

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

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

$$p(\omega_0; n_0) = \frac{q(\_\_\_\_; n_0)}{q(\_\_\_\_; n_0) + q(\_\_\_\_; n_0)} = \_\_\_\_$$

$$p(\omega_1; n_0) = \frac{q(\_\_\_\_; n_0)}{q(\_\_\_\_; n_0) + q(\_\_\_\_; n_0)} = \_\_\_\_$$

$$H(n_0) = p(\_\_\_\_; n_0) \log_2 \left(p(\_\_\_\_; n_0)\right) + p(\_\_\_\_; n_0) \log_2 \left(p(\_\_\_\_; n_0)\right) = \_\_\_\_ \text{ bits}$$

   (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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} H(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \quad & p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_) \log_2 \Big(p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_)\Big) \\ & + p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_) \log_2 \Big(p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_)\Big) \\ = \quad & \_\_\_\_\_\_ \text{ bits} \end{aligned}$$

$$\begin{aligned} H(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) = \quad & p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_) \log_2 \Big(p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_)\Big) \\ & + p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_) \log_2 \Big(p(\_\_;\_\_\_\_,\_\_\_\_,\_\_\_\_)\Big) \\ = \quad & \_\_\_\_\_\_ \text{ bits} \end{aligned}$$

$$\begin{aligned} \mathrm{IG}(n_0; \phi_{(k=0)}, \theta_{(k=0),l=0}) = \quad & H(n_0) - \big(p(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) H(n_0 \to L, \phi_{(k=0)}, \theta_{(k=0),l=0}) \\ & + p(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0}) H(n_0 \to R, \phi_{(k=0)}, \theta_{(k=0),l=0})\big) \\ = \quad & \_\_\_\_ - \Big((\_\_\_\_)\ (\_\_\_\_) + (\_\_\_\_)\ (\_\_\_\_)\Big) = \_\_\_\_ \text{ bits} \end{aligned}$$

(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))?

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

$$\phi_{n_0} = \phi_{(k=\_\_\_\_)} = \_\_\_\_\_\_, \quad \theta_{n_0} = \theta_{(k=\_\_\_\_),l=\_\_\_\_} = \_\_\_\_\_\_$$

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

	$\phi_{(k=0)} = 0$ $\theta_{(k=0),l=1} = 1.12$	$\phi_{(k=1)} = 2$ $\theta_{(k=1),l=0} = 3.76$	$\phi_{(k=1)} = 2$ $\theta_{(k=1),l=1} = 2.25$
$q(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$q(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$q(\omega_0; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$q(\omega_1; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$q(\omega_0; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$q(\omega_1; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$p(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$p(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$p(\omega_0; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$p(\omega_1; n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$p(\omega_0; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$p(\omega_1; n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$H(n_0 \to L, \phi_{(k)}, \theta_{(k),l})$			
$H(n_0 \to R, \phi_{(k)}, \theta_{(k),l})$			
$\mathrm{IG}(n_0; \phi_{(k)}, \theta_{(k),l})$			

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

$n_0 \rightarrow L$	$n_0 \rightarrow R$
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)
(____,____,____)	(____,____,____)

(f) (2 points) Comment on the nature of nodes (strike out the incorrect one within the braces)

$n_0 \rightarrow L$ is a {Decision node} / {Leaf node with $\Omega = \{\omega_0/\omega_1\}$}

$n_0 \rightarrow R$ is a {Decision node} / {Leaf node with $\Omega = \{\omega_0/\omega_1\}$}

—— 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.