AI&ML-Theory Exam on Dimensionality Reduction

Total Marks: 100

1. The Tree shown in Figure 1 can be used to generate the hash for input data $x \in \mathbb{R}^{10}$. Each split node of the tree hosts a decision function to form a binary value according to

$$hashValueBit(x; j, \theta) = \begin{cases} 1, & x[j] \geq \theta \\ 0, & x[j] < \theta \end{cases}$$

Also, the input goes to **LEFT** for $[hashValueBit = 1]$ and **RIGHT** for $[hashValueBit = 0]$

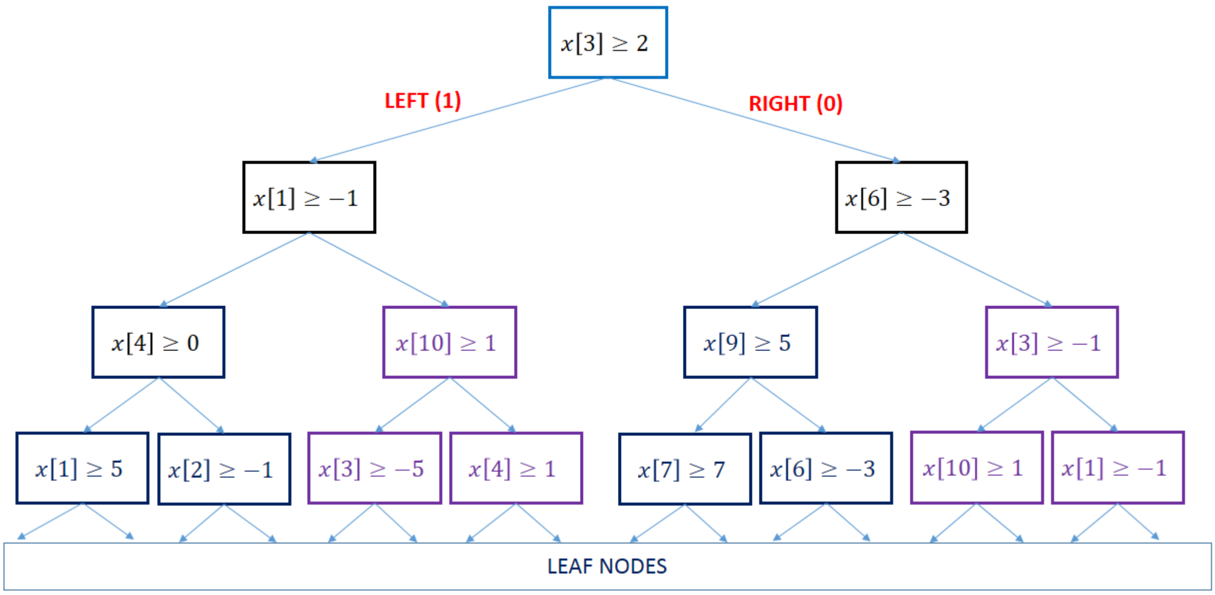

Figure1: The Tree for Indexing Input Data $x \in \mathbb{R}^{10}$

Evaluate Hash for $x = [-2,7,5,3,0,-1,4,6,9,-2]^T$ [10]

(a) Hash = 10110

(b) Hash = 1001

(c) Hash = 0101

(d) Hash = 01011

2. Consider the following Datasets $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, each containing *7* points in $\mathbb{R}^2$.

X1								X2						
1.5	2.5	3.5	5.5	4.5	6.5	8.5		-2.5	-1.5	1.5	3.5	4.5	2.5	5.5
1.5	2.5	3.5	4.5	5.5	6.5	7.5		3.5	4.5	5.5	6.5	7.5	8.5	9.5

Evaluate the Means of the Two Datasets [10]

(a) $\boldsymbol{\mu}_1 = \begin{bmatrix} 1.9286 \\ 6.5 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 4.6429 \\ 4.5 \end{bmatrix}$

(b) $\boldsymbol{\mu}_1 = \begin{bmatrix} 5.3485 \\ 3.75 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 2.3645 \\ 5.6563 \end{bmatrix}$

(c) $\boldsymbol{\mu}_1 = \begin{bmatrix} 4.6429 \\ 4.5 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 1.9286 \\ 6.5 \end{bmatrix}$

(d) $\boldsymbol{\mu}_1 = \begin{bmatrix} 4.7785 \\ 5.6563 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 1.8955 \\ 7.6525 \end{bmatrix}$

3. Evaluate Covariance Matrices of the two datasets [20]

(a) $\boldsymbol{C}_1 = \begin{bmatrix} 4.6867 & 4.6969 \\ 4.6969 & 4.0 \end{bmatrix}, \boldsymbol{C}_2 = \begin{bmatrix} 8.6987 & 6.3698 \\ 6.3698 & 4.3268 \end{bmatrix}$

(b) $\boldsymbol{C}_1 = \begin{bmatrix} 4.9796 & 4.2857 \\ 4.2857 & 4.0 \end{bmatrix}, \boldsymbol{C}_2 = \begin{bmatrix} 7.6735 & 5.0 \\ 5.0 & 4.0 \end{bmatrix}$

(c) $\boldsymbol{C}_1 = \begin{bmatrix} 8.6987 & 6.3698 \\ 6.3698 & 4.3268 \end{bmatrix}, \boldsymbol{C}_2 = \begin{bmatrix} 4.6867 & 4.6969 \\ 4.6969 & 4.0 \end{bmatrix}$

(d) $\boldsymbol{C}_1 = \begin{bmatrix} 7.6735 & 5.0 \\ 5.0 & 4.0 \end{bmatrix}, \boldsymbol{C}_2 = \begin{bmatrix} 4.9796 & 4.2857 \\ 4.2857 & 4.0 \end{bmatrix}$

4. Evaluate the (magnitude normalized) LDA direction vector $\hat{\boldsymbol{v}}$ (or $-\hat{\boldsymbol{v}}$). Select the nearest correct solution from the following choices. [10]

(a) $\hat{\boldsymbol{v}} = \begin{bmatrix} 0.6236 \\ -0.7818 \end{bmatrix}$

(b) $\hat{\boldsymbol{v}} = \begin{bmatrix} 0.6578 \\ -0.7894 \end{bmatrix}$

(c) $\hat{\boldsymbol{v}} = \begin{bmatrix} -0.6236 \\ -0.7818 \end{bmatrix}$

(d) $\hat{\boldsymbol{v}} = \begin{bmatrix} 0.6896 \\ 0.7894 \end{bmatrix}$

5. Consider the following dataset consisting of three readings obtained from a number of subjects.

#	1	2	3	4	5	6	7	8	9	10
R_1	11	3	6	16	0	16	0	14	18	13
R_2	8	13	16	8	1	8	7	19	14	19
R_3	4	3	13	9	4	1	15	14	13	18

Evaluate the mean of the dataset. [5]

(a) $\boldsymbol{\mu} = \begin{bmatrix} 9.7 \\ 11.3 \\ 9.4 \end{bmatrix}$

(b) $\boldsymbol{\mu} = \begin{bmatrix} 8.7 \\ 10.8 \\ 10.2 \end{bmatrix}$

(c) $\boldsymbol{\mu} = \begin{bmatrix} 10.3 \\ 9.5 \\ 7.8 \end{bmatrix}$

(d) $\boldsymbol{\mu} = \begin{bmatrix} 10.4 \\ 9.8 \\ 7.6 \end{bmatrix}$

6. Evaluate the Covariance Matrix of the zero-centered dataset. [15]

(a) $\boldsymbol{C} = \begin{bmatrix} 35.68 & 11.26 & 3.76 \\ 11.26 & 32.98 & 15.24 \\ 3.76 & 15.24 & 29.75 \end{bmatrix}$ (b) $\boldsymbol{C} = \begin{bmatrix} 42.61 & 14.79 & 4.32 \\ 14.79 & 30.81 & 19.58 \\ 4.32 & 19.58 & 32.24 \end{bmatrix}$

(c) $\boldsymbol{C} = \begin{bmatrix} 46.87 & 10.62 & 2.57 \\ 10.62 & 27.71 & 21.51 \\ 2.57 & 21.51 & 26.87 \end{bmatrix}$ (d) $\boldsymbol{C} = \begin{bmatrix} 40.65 & 11.89 & 5.70 \\ 11.89 & 35.28 & 17.43 \\ 5.70 & 17.43 & 30.54 \end{bmatrix}$

7. The data of these readings are used to evaluate their dependence on two latent factors $z_1$and z_2. Evaluate the factor loading matrix $\boldsymbol{V}$. Note that during eigen decomposition, negatives of eigen vectors may also occur in solution. Thus, select the nearest correct solution from the following choices. [20]

(a) $V = \begin{bmatrix} -5.8714 & -3.1067 \\ -6.5484 & 2.7953 \\ -5.0914 & 2.7659 \end{bmatrix}$ (b) $V = \begin{bmatrix} -3.1146 & -3.0913 \\ -3.4121 & 0.0956 \\ -2.4687 & 4.7869 \end{bmatrix}$

(c) $V = \begin{bmatrix} -4.7295 & -4.4248 \\ -4.8294 & 1.3852 \\ -3.9537 & 3.6008 \end{bmatrix}$ (d) $V = \begin{bmatrix} -2.0439 & -6.1724 \\ -5.6709 & 2.6745 \\ -6.3586 & 4.9456 \end{bmatrix}$

8. Evaluate the variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ of the sources. Select the nearest correct solution from the following choices. [10]

(a) $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.3487, 4.8967, 2.7116)$
(b) $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.7591, 6.2398, 1.1065)$
(c) $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.5584, 3.9014, 4.7856)$
(d) $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.6628, 5.5676, 3.6413)$