Indian Institute of Technology Kharagpur
Department of Electrical Engineering

Subject No.: EE60020 **Subject:** Machine Learning for Signal Processing
Date of Assignment: 29 January 2024 **Semester:** Spring 2023-24
Assignment Number: 1 **Duration:** 1 hour 50 mins **Full points:** 80

Name: Solution Key **Roll No:** ____________

---

1. An image in RGB representation format is provided below

$$\mathcal{I} = \begin{bmatrix} (1,1,0) & (6,3,1) & (5,3,1) \\ (2,1,0) & (3,2,0) & (4,2,1) \end{bmatrix}$$

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) (3 points) Write $\mathbf{X}^*$ corresponding to $\mathcal{I}$?

$$\mathbf{X}^* = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

(b) (3 points) What is the value of $\boldsymbol{\mu}$?

$$\boldsymbol{\mu} = \begin{bmatrix} 3.5 & 2.0 & 0.5 \end{bmatrix}^\top$$

(c) (4 points) Transform $\mathbf{X}^*$ to obtain $\mathbf{X}$ for Principal Component Analysis (PCA)?

$$\mathbf{X} = \mathbf{X}^* - \boldsymbol{\mu} = \begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \\ -1.0 & -1.0 & 0.0 & 0.0 & 1.0 & 1.0 \\ -0.5 & -0.5 & -0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

(d) (3 points) Compute the sample correlation matrix $\mathbf{R}_X$?

$$\mathbf{R}_X = \begin{bmatrix} 2.91 & 1.33 & 0.75 \\ 1.33 & 0.66 & 0.33 \\ 0.75 & 0.33 & 0.25 \end{bmatrix}$$

(e) (3 points) Compute the matrix $\mathbf{A}$ when $\mathbf{Y} = \mathbf{A}^\top \mathbf{X}$ represents the PCA operation?

$$\mathbf{A} = \begin{bmatrix} 0.88 & 0.05 & 0.46 \\ 0.40 & -0.57 & -0.70 \\ 0.22 & 0.81 & -0.53 \end{bmatrix}$$

(f) (1 point) What are the Eigenvalues corresponding to the components of $\mathbf{A}$?

$$\lambda_0 = 3.720, \lambda_1 = 0.067, \lambda_2 = 0.033$$

(g) (3 points) Represent $\mathbf{Y}$ such that every component $\mathbf{y}_i \in \mathbb{R}^{2\times 1}$

$$\mathbf{Y} = \begin{bmatrix} -2.73 & -1.84 & -0.55 & 0.55 & 1.84 & 2.73 \\ 0.03 & 0.08 & -0.43 & 0.43 & -0.08 & -0.03 \end{bmatrix}$$

2. Considering the dataset provide earlier in Q. 1 we perform Singular Value Decomposition (SVD) represented as $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^\top$

(a) (3 points) Compute $\mathbf{U}$?

$$\mathbf{U} = \begin{bmatrix} 0.88 & 0.05 & 0.46 \\ 0.40 & -0.57 & -0.70 \\ 0.22 & 0.81 & -0.53 \end{bmatrix}$$

(b) (3 points) Compute $\mathbf{\Lambda}^{\frac{1}{2}}$?

$$\mathbf{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} 4.73 & 0 & 0 \\ 0 & 0.62 & 0 \\ 0 & 0 & 0.47 \end{bmatrix}$$

(c) (6 points) Compute $\mathbf{V}$?

$$\mathbf{V} = \begin{bmatrix} -0.57 & -0.03 & 0.40 & -0.36 & 0.42 & -0.42 \\ 0.57 & 0.03 & -0.40 & -0.36 & 0.42 & -0.42 \\ 0.39 & 0.15 & 0.56 & 0.60 & 0.26 & -0.26 \\ -0.39 & -0.15 & -0.56 & 0.60 & 0.26 & -0.26 \\ -0.11 & 0.68 & -0.10 & 0 & 0.50 & 0.50 \\ 0.11 & -0.68 & 0.10 & 0 & 0.50 & 0.50 \end{bmatrix}$$

(d) (3 points) Compute $\mathbf{U}^*$ when $\mathbf{X}^* = \mathbf{U}^*\mathbf{\Lambda}^{*\frac{1}{2}}\mathbf{V}^{*\top}$?

$$\mathbf{U}^* = \begin{bmatrix} -0.86 & 0.18 & -0.46 \\ -0.47 & -0.57 & 0.66 \\ -0.14 & 0.79 & 0.58 \end{bmatrix}$$

(e) (3 points) Compute $\mathbf{\Lambda}^{*\frac{1}{2}}$?

$$\mathbf{\Lambda}^{*\frac{1}{2}} = \begin{bmatrix} 11.00 & 0 & 0 \\ 0 & 0.85 & 0 \\ 0 & 0 & 0.47 \end{bmatrix}$$

(f) (2 points) What is the reason for dis-/similarity of $\mathbf{U}$ and $\mathbf{U}^*$?

3. Consider the symbols in $\mathcal{I}$ in Q. 1.

(a) (1 point) The minimum number of bits required to represent the values in unsigned integer format is 3 bits.

(b) (7 points) List the unique symbols and their probability such that the symbol with highest probability is on top of the table

Symbol	Probability
0	1/6
1	1/3
2	1/6
3	1/6
4	1/18
5	1/18
6	1/18

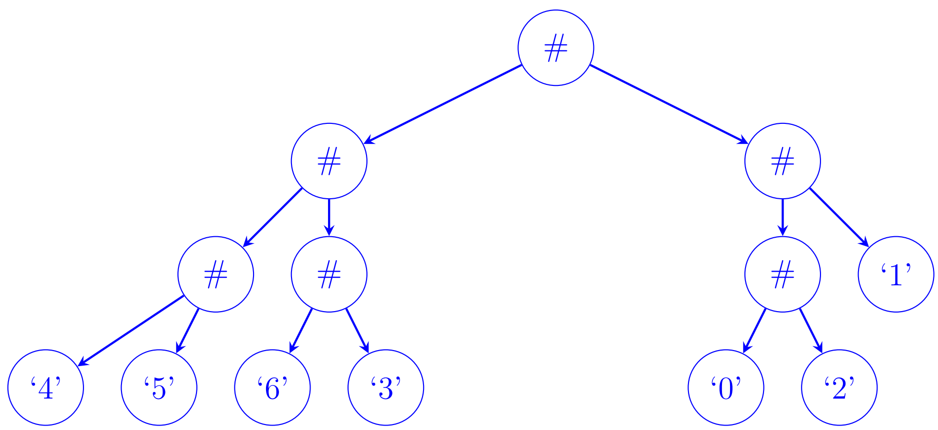
(c) (5 points) Create the Huffman code book (after completing (e))?

Symbol	Bit code
0	100
1	11
2	101
3	011
4	000
5	001
6	010

(d) (18 points) Write the Huffman coded dataset arranging $\mathcal{I}$ in column major format? Indicate the code within braces and corresponding symbol in the blank below it.

$$\underbrace{11}_{1}\ \underbrace{010}_{6}\ \underbrace{001}_{5}\ \underbrace{101}_{2}\ \underbrace{011}_{3}\ \underbrace{000}_{4}$$
$$\underbrace{11}_{1}\ \underbrace{011}_{3}\ \underbrace{011}_{3}\ \underbrace{11}_{1}\ \underbrace{101}_{2}\ \underbrace{101}_{2}$$
$$\underbrace{100}_{0}\ \underbrace{11}_{1}\ \underbrace{11}_{1}\ \underbrace{100}_{0}\ \underbrace{100}_{0}\ \underbrace{11}_{1}$$

(e) (7 points) Create the min-heap to build the Huffman tree for generating the Huffman code book. Indicate the symbol and its probability in the leaf node. Indicate the sum of probability of children in a parent node. The probability of the left child is lower than or equal to that of the right child. A left child transition is to be denoted by boolean 0 and a right child transition is to be denoted by boolean 1. Strike off the nodes and edges whichever do not contain any valid probability. Add nodes and edges as required.



(f) (2 points) Find the Entropy of $\mathcal{I}$ and the average code length per symbol obtained from the coded dataset in (d)?

$$H(\mathcal{I}) = 2.51 \;\; bits$$

$$\text{Average Code Length} = 2.66 \;\; bits/symbol$$