SML 2025, Monsoon, MidSem

Note:

- Symbols have their usual meanings. Duration 1.5 hours. Number in [.] indicate marks. [COx] indicates the question is mapped to the respective course outcome.

Q1. Compute the weights of linear regression where the data is $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{3} = \{(0,1), (1,-1), (3,-2)\}$. Choose the degree of the polynomial to be 1, that is of the form $y = mx + c$. You need to compute m and c. [CO1] [2]

Q2. Let x be a random variable distributed according to Rayleigh distribution [CO1] [3]

$$p(x|\theta) = \theta x e^{-\frac{\theta x^2}{2}}; \text{ for } x \geq 0, \text{and}, p(x|\theta) = 0; \text{ for } x < 0$$

Take the following exponential pdf as prior over θ and compute the MAP estimate for θ.

$$P(\theta) = \lambda e^{-\lambda\theta} \text{ for } \theta \geq 0 \text{ and } P(\theta) = 0 \text{ for } \theta < 0$$

Q3. Find the first principal component for given data matrix $X \in \mathcal{R}^{d \times N}$ [CO2] [3]

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Q4.a. Compute the FDA projection vector, given, classes $X_1 \in R^{d \times 2}$ and $X_2 \in R^{d \times 2}$, where [CO2] [3]

$$X_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$X_2 = -X_1$$

b. Suppose some new test sample is given as follows-

$$x_{test} = \begin{pmatrix} .25 \\ -.75 \end{pmatrix}$$

Find the class to which x_{test} belongs to. Assume equal prior for both classes. [1]

Q5. Suppose for a particular high dimensional binary classification task, you decide to first apply PCA followed by FDA. Let the PCA matrix be $U_p \in R^{n \times p}$. Let the projected data after PCA be $Y = U_p^\top X$, where $X = [X_1, X_2]$ and $X_1, X_2 \in R^{d \times n}$, $X \in R^{d \times 2n}$, $Y \in R^{p \times 2n}$, $p << d$. Let X_1 be the data matrix from first class and X_2 from another class. Let the class means of X_1, X_2 be μ_1 and μ_2 respectively. Let the scatter matrices of X_1, X_2 be S_1 and S_2 respectively. Now apply FDA on Y. Derive the FDA projection vector w in terms of $U_p, S_1, S_2, \mu_1, \mu_2$. [CO2] [3]

dxp