INDIAN STATISTICAL INSTITUTE

End-semester Examination : Semester I (2025-2026)

M.Stat. Second Year

Pattern Recognition

Date : 24/11/2025

Maximum marks: 100 (Total marks: 105) Time: 3 hours.

1. Consider a 2-class classification problem with $Y \in \{-1, 1\}$, continuous covariate X and the exponential loss function given by $\exp\{-YF(X)\}$.
   i) Show that the minimizer of $E(\exp(-YF(X)\})$ with respect to $F(X)$ is a monotonic function of the odds-ratio given by $O(X) = \frac{Pr(Y=1|X)}{Pr(Y=-1|X)}$.
   ii) Discuss the overfitting of the AdaBoost algorithm based on the result given in i). [8+4]

2. i) What is a BART model for 2-class classification using probit link function? Write the model along with the priors and discuss the Gibbs sampling steps for posterior estimation.
   ii) In a randomly chosen tree, suppose the probability that a node at depth $d = (0, 1, 2, \dots)$ is non-terminal is given by $0.95(1+d)^{-2}$. Find the probability that a randomly chosen tree has (a)1 (b)2 (c)3 terminal nodes. [6+6]

3. Let $y = f(x) + \epsilon$ be the data generating process, with $E(\epsilon) = 0$ and $f(x)$ be an unknown function. Compare bagging and random forest for estimating $f(x^*)$ at a particular value of x^*. The comparison needs to be made in terms of mean square error and the precise conditions for the superiority of one algorithm over the other need to be stated. [12]

4. Let $x, z \in R^p$, $K(x, z)$ be a known kernel and let H denotes the Hilbert space induced by $K(x, z)$. Suppose $x_1, x_2, \dots, x_N \in R^p$ are a set of observations. Write, in details, the k-means clustering algorithm for clustering the observations in H based on the distance induced by the kernel. [12]

5. i) Describe the backpropagation algorithm for optimization in a feed-forward neural network with one hidden layer.
   ii) Prove or disprove: A mutli-layer feed-forward neural network with a polynomial activation function of degree d, where d is fixed, can approximate any continuous function with compact support. [8+4]

6. Consider the soft-margin Support Vector Machine (SVM) for binary classification. Given training data $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$, and a known constant $C > 0$, the primal optimization problem is

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, n, \\ & \xi_i \geq 0, \quad i = 1, \dots, n. \end{aligned}$$

i) Formulate the Lagrangian

$$L(w, b, \xi, \alpha, \mu)$$

for this problem using Lagrange multipliers $\alpha_i \geq 0$ (for the margin constraints) and $\mu_i \geq 0$ (for the non-negativity constraints on ξ_i).
ii) Derive the Karush–Kuhn–Tucker (KKT) conditions for the above optimization problem.
iii) Using the KKT conditions, show that at the optimal solution:

$$\begin{aligned}
&\alpha_i > 0 \;\Rightarrow\; y_i(w^\top x_i + b) = 1 - \xi_i, \\
&0 < \alpha_i < C \;\Rightarrow\; y_i(w^\top x_i + b) = 1, \\
&\alpha_i = 0 \;\Rightarrow\; y_i(w^\top x_i + b) > 1, \\
&\alpha_i = C \;\Rightarrow\; \xi_i > 0.
\end{aligned}$$

iv) Explain **geometrically** how the KKT conditions determine which training points become *support vectors*, and how the *margin* and decision boundary are established. Illustrate the role of the parameters C and α_i in this interpretation. [4+4+4+4]

7. Suppose you have a model which provides the class-conditional densities for wind direction (θ) measured on a continuous scale, where there are two classes given as Rain(C_1) and No Rain (C_2). Let, the conditional densities of $\theta \in [0, 2\pi)$ for the two classes are given as $f(\theta|C_1) = K_1 \exp(\cos(\theta))$ and $f(\theta|C_2) = K \exp(\cos^2(\theta))$. Let $\theta = 0$ represents the North direction.
i) Derive the minimax rule for this 2-class classification problem.
ii) Suppose you have a compass which measures wind directions on a categorical scale where the categories are given by North, East, South, West. Suppose the compass shows the direction to be East. Based on the rule derived in i), classify this observation to C_1 or C_2. [7.5+7.5]

8. i) Describe the kernel density estimator based classification algorithm for a multi-class classification problem with zero-one loss.
ii) Modify the algorithm for general loss. Consider the loss function of classifying an observation belonging to class i to class j as $L(i, j)$. Let, $L(i, i) = 0$, $L(i, j) = i$ when $i \neq j$. Prove the Bayes optimality of the classification algorithm. [7+7]