Total time: 70 mins Total marks: 18 Name and Roll Number: _____

Questions

1. (2 points) Write and explain the formula for feature importance of X_j^{th} feature for random forests when using M trees. The total number of samples fed to the model is N.

2. (1 point) How does a decision tree's bias and variance vary with increasing depth. Explain.

3. For each model below and the common dataset $(x, y) = \{(1, 6), (2, 3), (4, 1.5)\}$:

   1. Write the squared loss $J(\cdot)$ in terms of x_i, y_i, a, b.
   2. Find the estimator(s) (a, b) using either the normal equation or first principles.

   (a) (2 points) Model: $y_i = \dfrac{a}{x_i}$, $x_i > 0$.

   (b) (2 points) Model: $y_i = a - x_i$.

   (c) (2 points) Model: $y_i = a + \dfrac{x_i}{b}$.

4. (2 points) Show, from a geometric perspective, that the normal equation

$$\mathbf{X}^\top \mathbf{X}\,\theta = \mathbf{X}^\top y$$

   arises by requiring the residual vector

$$\mathbf{r} = \mathbf{y} - \mathbf{X}\theta$$

   to be orthogonal to the span of the columns of $\mathbf{X}$.

5. (a) (1 point) Define k-fold cross-validation and leave-one-out cross-validation (LOOCV).

   (b) (1½ points) Suppose $n = 100$. Assume that training a model on m points costs time am, and testing on m points costs bm, where $a, b > 0$ are constants.

   1. Derive the total computational cost of 5-fold CV.
   2. Derive the total computational cost of LOOCV.
   3. Compare the two costs numerically when $n = 100$, $a = 1$, and $b = 0.1$.

6. (a) (1½ points) For a binary class node with class-1 proportion $p \in [0, 1]$, the Gini impurity is

$$G(p) = 1 - (p^2 + (1 - p)^2).$$

   (i) Find the p that maximizes it. Ensure you also test via the double derivative test. (ii) Report $G_{\max}$.

   (b) (1 point) Dataset (single feature X, binary label Y):

$$(1, 0),\ (2, 0),\ (3, 1),\ (4, 1),\ (5, 1).$$

   (i) Compute the root-node Gini. (ii) Compute the *weighted* Gini for splits at $X = 2.5$ and at $X = 3.5$. Which split would a decision tree algorithm using Gini index choose? Justify.

7. (1 point) In polynomial regression, we fit

$$y \approx \theta_0 + \theta_1 x + \cdots + \theta_d x^d.$$

   Briefly state how increasing the degree d affects bias and variance of the model.

8. (1 point) You are developing a medical device that detects snoring in 10-second windows during sleep. On test data, your model achieves 90% accuracy.

   Would you recommend releasing the device based on this result alone? List at least **two important factors** that should be considered before deployment, and explain why they matter.