

I. Conceptual Questions

1. Please explain the meaning of the bias-variance tradeoff and how to strike a balance between model complexity and prediction error.
2. Please briefly describe the differences between L1 regularization and L2 regularization and their roles in model optimization.
3. Please explain the basic principle of the K-Nearest Neighbors (KNN) algorithm and how to choose an appropriate K value.
4. Please explain the basic concepts of Support Vector Machines (SVM), including support vectors, margin, and kernel function.

II. Calculation Questions

1. Given the following dataset:

X: [2, 4, 6, 8, 10, 12] Y: [3, 5, 7, 9, 11, 13]

Please calculate the regression coefficient w and the intercept b using simple linear regression (either manually or by writing code).

2. Assume a binary classification problem with the following confusion matrix:

Actual \ Predicted	Positive	Negative
Positive	80	20
Negative	10	90

Please calculate the following evaluation metrics: Accuracy, Precision, Recall, and F1-score.

3. Suppose we have a regression problem with the following true and predicted values: True values: [10, 15, 20, 25, 30] Predicted values: [12, 18, 19, 24, 32] Please calculate the Mean Squared Error (MSE) and R^2 score.

III. Inference Questions

1. In a binary classification problem, you used a logistic regression model for prediction. The model performed well during training but had poor predictive performance on new data. Please analyze the possible reasons and provide corresponding solutions.

2. When dealing with a regression problem with multiple features, we may encounter the issue of multicollinearity. Please explain what multicollinearity is and how to detect and address it.