

Week4 - Cross-validation, ridge regression and nonlinear regression

DS3010: Introduction to Machine Learning Lab

Timing: 02:00 PM to 04:45 PM

Max Marks:

Instructions

1. Submit one .ipynb file containing all answers. The name should be [student_name]-[rollno]-[lab].ipynb
 2. Write the questions in separate text blocks before the answers.
 3. Outputs for all sub-questions should be given and the code should be executable.
 4. Write justifications for your choices where needed.
 5. Ensure that all plots include clear labels and legends for better interpretation.
 6. Use of generative AI tools (such as ChatGPT, Gemini, etc.) is strictly prohibited. Any submission found to contain AI-generated or plagiarized content will receive a score of zero, and disciplinary action.
-

1. Non-linear Regression

(2)

- (a). Generate 300 data points using a relation $y = \sin(x) + \epsilon$, where x is evenly spaced between 0 and 2π and $\epsilon \sim \mathcal{N}(0, 0.5)$.
- (b). Fit the Ridge regression (Alpha = 1) model to the above data.
- (c). Fit the polynomial regression model with degree 3 on the same dataset.
- (d). Fit the polynomial regression model with degree 30 on the same dataset.
- (e). Plot the dataset along with the fitted lines for all the models.
- (f). Write your observation.

2. Multivariate regression with Cross-validation

(3)

- (a). Generate 300 data points using a relation $y = \sin(x) + z^2 + \epsilon$, where x is evenly spaced between 0 and 2π , z evenly spaced between $-\pi$ and π and $\epsilon \sim \mathcal{N}(0, 0.5)$. Split the data into training and testing.
- (b). Fit Ridge regression with the best degree of polynomial regression model. Compute the MSE and R2 score on the test set.
- (c). Define a range of alpha values [0.01, 0.1, 1, 10, 100] and the best degree of polynomial [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] for ridge regression. For each alpha and degree, perform 5-fold cross-validation.
- (d). Determine the best alpha and degree based on average validation score.
- (e). Using the best alpha and degree, train the model and compute the MSE and R2 score on the test set.
- (f). Discuss and analyse the significance of alpha, degrees, and cross-validation in relation to model performance.