

Week 4 - Cross-validation

DS3010 - Introduction to Machine Learning

Instructions

1. Provide commented, indented code. Variables should have meaningful names.
 2. Submit one .ipynb file containing all answers. The name should be [student name][roll_number] assignment[number].ipynb
 3. Read the questions carefully before answering. If a question asks to follow a particular approach or to use a specific data structure, then it must be followed.
 4. Write questions in separate text blocks in Jupyter Notebook before the code blocks containing answers.
 5. All plots should have appropriate axis labels, titles, and legends.
-

Tasks for the Lab

1. Using Scikit-Learn Cross-Validation (3 points)

Tasks:

- A. Fit Ridge regression with the best degree of polynomial regression model. Calculate MSE and plot polynomial regression line fitted to the 'sinusoidal' dataset. (2)
- B. Use Scikit-Learn's **cross val score with 5 folds** to evaluate (Average MSE) the Ridge regression model. Use $\alpha = [0.01]$ (1)

2. Implementing k-Fold Cross-Validation from Scratch (7 points)

Tasks:

- A. Implement 5-fold cross-validation from scratch. (5)
- B. For each fold, train the Ridge regression model and compute the average MSE. Use $\alpha = [0.01]$ (1)
- C. Discuss any differences observed between the above two methods. (1)

3. Hyperparameter Tuning with Cross-Validation (5 points)

Tasks:

- A. Do a train-test split of the same data. Define a range of alpha values [0.01, 0.1, 1, 10, 100] and best degree of polynomial [1, 2, 3, 4, 5] for Ridge regression. For each alpha and degrees perform 5-fold cross-validation using your custom implementation on the training part. (3)
- B. Identify the alpha value and degree of polynomial that yields the best average validation score. (1)
- C. Retrain the Ridge regression model using the optimal alpha and evaluate its performance on the testing set. (1)

4. Analysis and Discussion (5 point)

Tasks:

- A. Analyze how the degree of polynomial features affects the model's performance and complexity. (2)
- B. Discuss the importance of cross-validation in evaluating and selecting models. (2)
- C. Reflect on the advantages and disadvantages of implementing cross-validation manually versus using library functions. (1)