3_Non_Linear

January 28, 2025

1 Exercise: Fitting a Non-Linear Function with Polynomial Regression

In this exercise, you will analyze a **univariate non-linear dataset** and determine the best polynomial fit for the function using model selection. You will evaluate the models based on the **Mean Squared Error (MSE)** on the validation set.

1.1 Instructions

1. **Load the Dataset**:
   - Read the data from the file `non_linear_data.csv`.
   - The dataset contains two columns:
     - `x`: The independent variable.
     - `y`: The dependent variable.
2. **Visualize the Data**:
   - Create a scatter plot to visualize the relationship between `x` and `y`.
3. **Split the Data**:
   - Split the dataset into three sets: **training (60%)**, **validation (20%)**, and **test (20%)**.
   - Use the training set to train models, the validation set for model selection, and the test set for final evaluation.
4. **Fit Polynomial Models**:
   - For each degree (d) in ($\{1, 2, ..., 10\}$), fit a polynomial regression model:
     - Generate polynomial features up to degree (d).
     - Fit the model on the training set.
5. **Evaluate Models**:
   - For each polynomial degree:
     - Compute the **MSE** on the training and validation sets.
     - Store the results for comparison.
6. **Model Selection**:
   - Identify the degree of the polynomial that results in the lowest **MSE** on the validation set.
7. **Test the Best Model**:
   - Use the test set to evaluate the performance of the selected model.
   - Compute the **MSE** on the test set.
8. **Visualize the Results**:
   - Plot the original data and the predictions of the best-fitting polynomial model.
9. **Discuss the Findings**:
   - Analyze the results and discuss how well the selected polynomial fits the data.

- Compare the model's performance on the validation and test sets.

1.2 Deliverables

- A scatter plot of the data (x vs. y).
- A table showing the MSE for training and validation sets for polynomial degrees 1 to 10.
- The degree of the best-fitting polynomial based on validation MSE.
- The MSE on the test set for the best model.
- A plot showing the data points and the fitted polynomial curve for the best model.
- A short discussion of the findings.

[ ]: