Homework 1: «Polynomial Regression»

Course: CS454&554
Professor: Ahmet İbrahim Ethem Alpaydın
Student Name: Roman Mordovtsev

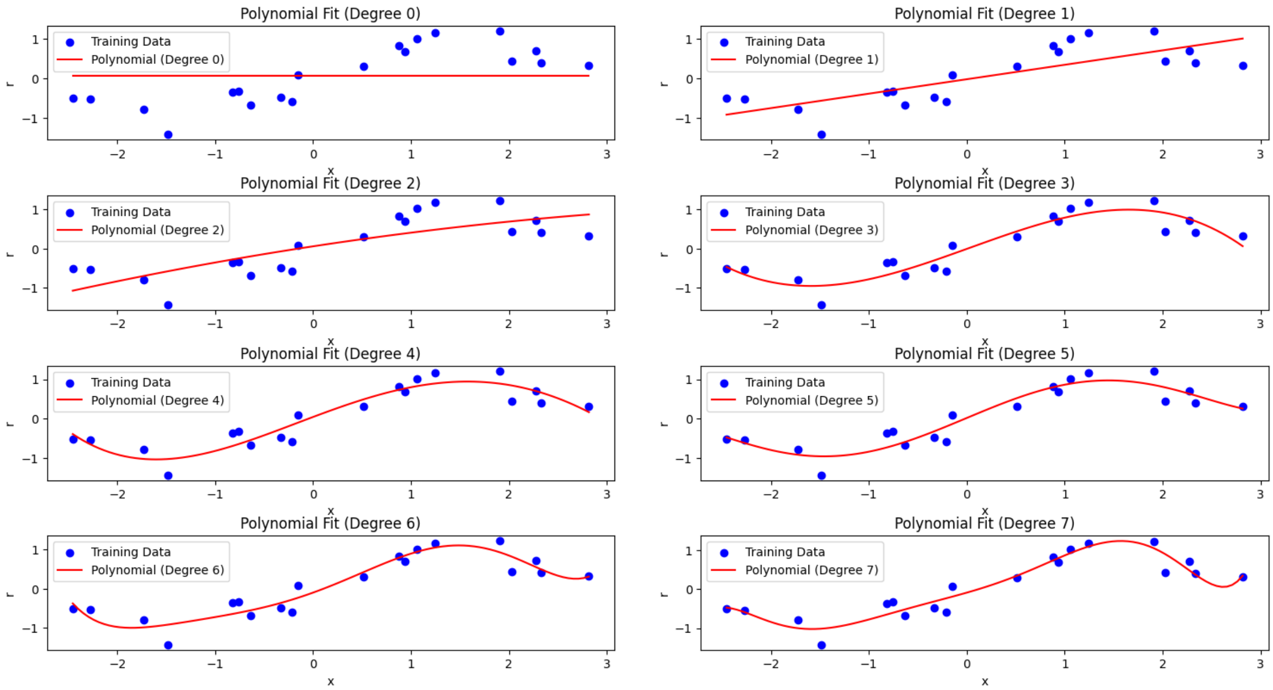
Polynomial Regression Analysis Report:

Polynomial regression is a technique used to fit a nonlinear relationship between input and output data. In this analysis, I explore how polynomials of different degrees (0 to 7) fit the given dataset and evaluate their performance on both training and test data, using Python. Source Code is attached in submission.

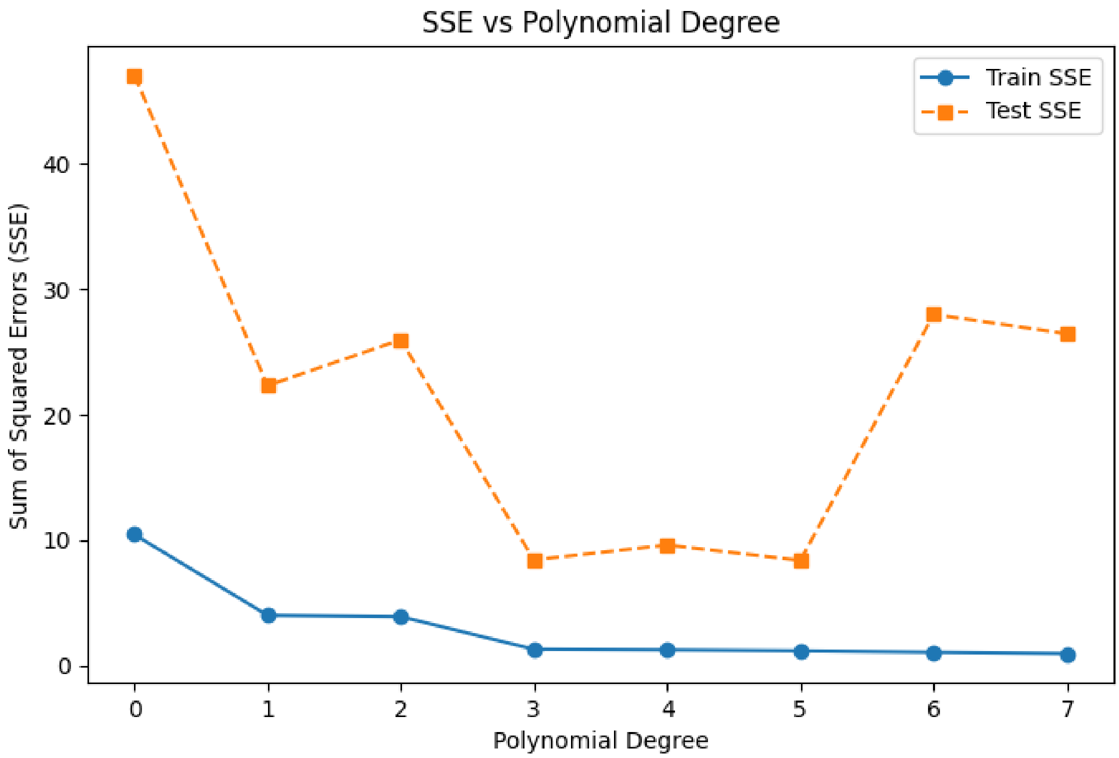
The main steps I followed:

- Load and preprocess the data: I read the CSV files and ensured numerical values were properly formatted.
- Fit polynomial models: Using NumPy's `polyfit` function, I fitted polynomials of degrees 0 through 7 to the training data.
- Make predictions: Each polynomial model was used to predict values for both training and test datasets.
- Compute Sum of Squared Errors (SSE): I calculated SSE for both datasets to measure how well each model fit the data.
- Visualize results: I plotted the fitted polynomials against training data and graphed SSE trends.

Results of visualization could be found below.



Picture 1 – Polynomials



Picture 2 - SSE

These plots illustrate how higher-degree polynomials can fit the training data more closely but may fail to generalize well to new data.

The results of computing SSE for both training and test data showed:

- Low-degree polynomials (0-2): These models underfit the data, meaning they fail to capture important patterns.
- Mid-range degrees (3-5): These models struck a balance between fitting the data well and generalizing to new data.
- High-degree polynomials (6-7): These models overfitted the training data, leading to poor generalization on test data.

Analysis demonstrates that while polynomial regression is a powerful tool, choosing the right degree is crucial. Too low a degree results in underfitting, while too high a degree leads to overfitting. Based on findings, polynomial degrees 3-5 provide the best trade-off between accuracy and generalization.