Lab Assignment: Bias-Variance Decomposition with Linear and Polynomial Regression

Objective

The objective of this lab is to deepen your understanding of the bias-variance trade-off in regression models. You will estimate and plot the bias, variance, irreducible-error variance, and total error for models of varying complexity on two different types of datasets: linear and quadratic, both with added noise.

Detailed Tasks

Task 1: Data Generation:

- Generate a synthetic data set with two real-valued variables X and Y such that Y depends linearly from X;
- Generate a synthetic data set with two real-valued variables X and Y such that Y depends quadratically from X;
- Add Gaussian noise (e.g. with standard deviation 1) to output variable Y in both data sets.
- Plot both data sets and the true regression functions in the X, Y space.

In context of model study and bias-variance trade-off answer the following questions:

- What is the purpose of adding Gaussian noise?
- Why is it important to visualize the data and the true functions?

Task 2: Model Training with Polynomial Regression

- Define polynomial regression models with varying degrees (model complexities) (e.g. 1, 3, 5, 7, 9, 12).
- Train polynomial regression models with varying degrees and predict on both data sets.

- Plot the data and regression model curves in the X, Y space.

Answer the following questions:

- What is a polynomial regression model and how is it different from a linear regression model?
- What do you expect to happen with the bias and variance of polynomial regression models when the polynomial degree increases?

Task 3: Bias-Variance Decomposition

- Estimate bias, variance, irreducible error, and total error for each polynomial regression model on the linear and quadratic data sets.

Answer the following questions:

- Is it necessary to train multiple models to estimate the bias, variance, and irreducible error?
- Is it possible to estimate the model's bias without an access to the true regression function?
- Is it possible to estimate the model's variance without an access to the true regression function?

Task 4: Visualization and Analysis

- Plot bias, variance, irreducible error, and total error as functions of model complexity for both linear and quadratic data.
- Estimate the total error as the sum of the (squared) bias, variance, and irreducible-error variance and as the absolute difference of the total error computed directly from y and $\hat{y}$.

Answer the following questions:

- Can you describe the trade-offs between the bias and variance for polynomial regression models on the linear and quadratic data sets?
- How and why do the bias-variance trade-offs differ between the linear and quadratic datasets?
- Using the bias-variance trade-off analysis (based on the previous two questions), which polynomial degree would you choose for the linear and quadratic datasets to best balance the bias and variance?
- Discuss the concept of overfitting/underfitting in the context of polynomial regression. How can bias-variance analysis help in identifying overfitting/underfitting?

Submission Requirements

Please follow the guidelines below for submitting your assignment's solutions:

- Analytical Report: Submit a PDF file that serves as your Analytical Report. This should contain all your written answers, interpretations, and graphical visualizations for each exercise. Ensure that the graphics are clearly labeled and appropriately integrated into your explanations.
- Jupyter Notebook PDF: Additionally, submit a PDF version of your Jupyter Notebook that contains all the code used for data generation, analysis, and visualization. Make sure that the code is well-commented for readability.

Extra Assignments

- Vary the level of Gaussian noise added to the datasets and repeat the bias-variance analysis. How does increasing or decreasing the noise level affect the bias, variance, irreducible error, and total error?
- Repeat the bias-variance trade-off exercises for Ridge and Lasso regression. Note that the regularization parameter has to be taken into account.