Artificial Intelligence and Machine Learning Fundamentals

Activity 6: Stock Price Prediction with Quadratic and Cubic Linear Polynomial Regression with Multiple Variables

In this section, we will discuss how to perform linear, polynomial, and support vector regression with scikit-learn. We will also learn how to find the best fit model for a given task. We will be assuming that you are a software engineer at a financial institution and your employer wants to know whether linear regression or support vector regression is a better fit for predicting stock prices. You will have to load all of the data of the S&P 500 from a data source. Then, you will need to build a regressor using linear regression, cubic polynomial linear regression, and a support vector regression with a polynomial kernel of degree 3 before separating the training and test data. Plot the test labels and the prediction results and compare them with the y=x line. Finally, compare how well the three models score.

1. Load the S&P 500 index data using Quandl, and then prepare the data for prediction.
2. Use a polynomial of degree 1 for the evaluation of the model and for the prediction. The closer the dots are to the y=x line, the less error the model works with. Perform a linear multiple regression with quadratic polynomials. The only change is in the Linear Regression model.
3. Perform Support Vector regression with a polynomial kernel of degree 3.

The model does not look efficient at all. For some reason, this model clearly prefers lower values for the S&P 500 that are completely unrealistic, assuming that the stock market does not lose 80% of its value within a day.