Polynomial Regression

Definition - Polynomial Regression is an extension of Simple Linear Regression that models the relationship between an input feature and a target variable by fitting a polynomial function to the data. It allows for capturing more complex, nonlinear relationships between the input and output by including polynomial terms of the input feature.

Unlike linear regression, which assumes a straight-line relationship between the dependent and independent variables, polynomial regression can model curves, making it suitable for cases where the data shows a non-linear trend.

Equation

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots + \beta_n X^n$$

- Y is the predicted output (dependent variable)
- X is the input feature (independent variable)
- B_0, B_1, …, B_n are the coefficients for each polynomial term
- n is the degree of the polynomial, determining the highest power of X

Feature Scaling - If the algorithm uses OLS method for optimization, then feature scaling generally not required because OLS can handle different feature magnitudes.

OLS computes the coefficients directly through matrix operations and does not rely on iterative optimization like Gradient Descent. Therefore, the algorithm is less sensitive to the scale of features.

How It Works

- Apply feature scaling if needed.
- Choose the degree of the polynomial.
- The algorithm generates a new feature matrix consisting of all polynomial combinations of the features with degree less than or equal to the specified degree. For example, if an input sample is two dimensional and of the form [a, b], the degree-2 polynomial features are [1, a, b, a^2, ab, b^2].
- Using the training data, the algorithm fits the polynomial regression model by finding the optimal coefficients that minimize the error.
- Once the model is trained, it can be used to predict the dependent variable y for new preprocessed (3. step) input data of the input features x_1, x_2, …, x_.

Example - You want to predict the price of a house (Y) based on size of the house (X_1) and number of rooms (X_2). The relationship between house price, size, and the number of rooms is not perfectly linear. To capture the non-linear relationship, you decide to use Polynomial Regression.

Initially, the simple Multiple Linear Regression Model would look like this:

$$Price = \beta_0 + \beta_1 \times (Size) + \beta_2 \times (Number\ of\ Rooms)$$

For Polynomial Regression with two features, you can add terms for interaction between features and higher-degree terms. For example, a degree-2 polynomial model will have the following terms:

$$Price = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_1^2 + \beta_4 \times X_2^2 + \beta_5 \times X_1 \times X_2$$

The model trains on the dataset to determine the best coefficients that minimize the prediction error.

Now, the model can predict the price of a house based on its size and number of rooms.