Polynomial regression - Polynomial regression is a type of regression analysis used in machine learning and statistics to model the relationship between a dependent variable and one or more independent variables. While linear regression models assume a linear relationship between the variables, polynomial regression accommodates more complex, nonlinear patterns.

In polynomial regression, the relationship is modeled as an nth-degree polynomial, where '*n*' is the degree of the polynomial. The general form of a polynomial regression equation with one independent variable is:

$Y = \beta_0 + \beta_1 X + \beta_2 x_2 + ... + \epsilon$

Y - dependant variable

x - independent variable

$\beta_0, \beta_1, \beta_2$ - coefficients

ϵ - *error term*

Polynomial regression allows for a more flexible fit to the data when the relationship is not strictly linear. By including higher-degree polynomial terms, the model can capture curvature and non-linear patterns in the data. However, it's important to note that increasing the degree of the polynomial can lead to overfitting, where the model fits the training data too closely and performs poorly on new, unseen data.

Polynomial regression is a useful technique when the relationship between variables is more complex than a simple straight line, and it provides a flexible tool for capturing non-linear patterns in the data. However, careful consideration should be given to model complexity and overfitting when selecting the degree of the polynomial.

KNN - K-Nearest Neighbors is a simple algorithm used for both classification and regression tasks in machine learning. It's a type of instance-based or lazy learning algorithm, meaning that it doesn't learn a specific model during training. Instead, it memorizes the entire training dataset and makes predictions based on the similarity between new, unseen instances and the instances in the training data.

- During the training phase, the algorithm simply stores the feature vectors and their corresponding labels from the training dataset.
- When making a prediction for a new instance, KNN identifies the k-nearest neighbors of that instance in the training data. The term "k" represents the number of neighbors to consider, and it is a user-defined parameter.
- The distance metric, often Euclidean distance, is commonly used to measure the similarity between instances.
- The algorithm calculates the distance between the new instance and every instance in the training set.
- It selects the k instances with the smallest distances to the new instance.
- For a classification task, the algorithm assigns the most frequent class label among the k-nearest neighbors to the new instance. This is often done through a majority voting mechanism.
- For a regression task, the algorithm assigns the average or weighted average of the target values of the k-nearest neighbors as the prediction for the new instance.

The choice of the number of neighbors (k) can significantly impact the performance of the KNN algorithm. A smaller value of k may lead to more flexible models, but it can be sensitive to noise, while a larger value of k may smooth out the decision boundaries but might miss local patterns.

The choice of distance metric is important and depends on the nature of the data. Euclidean distance is commonly used, but for categorical data, other metrics like Hamming distance may be more appropriate.

KNN can be computationally expensive, especially with large datasets.

Feature scaling is often crucial in KNN, as it ensures that all features contribute equally to the distance computation.It is often used for smaller datasets or as a baseline model for comparison with more complex algorithms.