# Machine Learning Concepts: Definitions and Functions

## Linear Regression

Definition: Linear Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables.

What it does: It predicts outcomes by fitting a straight line to the data points, minimizing the error between actual and predicted values.

Application: Commonly used in forecasting, risk assessment, and understanding relationships between variables.

Limitation: Assumes a linear relationship, which may not hold in complex datasets.

## Multiple Linear Regression

Definition: Multiple Linear Regression models the relationship between one dependent variable and multiple independent variables.

What it does: Fits a hyperplane to the data, allowing for prediction based on multiple factors simultaneously.

Application: Used in economics, biology, and other fields for analyzing multidimensional data.

Advantage: Provides a more comprehensive understanding of variable interactions compared to simple linear regression.

## Polynomial Linear Regression

Definition: Polynomial Regression extends Linear Regression by fitting a polynomial curve to the data instead of a straight line.

What it does: Captures non-linear relationships between variables by including polynomial terms.

Application: Used in scenarios like curve fitting, stock price modeling, and environmental data analysis.

Limitation: Overfitting can occur if the polynomial degree is too high.

## Naive Bayes

Definition: Naive Bayes is a probabilistic algorithm based on Bayes' Theorem, assuming independence between predictors.

What it does: Classifies data by calculating the probability of a class given the features.

Application: Widely used in spam filtering, sentiment analysis, and document classification.

Advantage: Efficient for large datasets but assumes feature independence, which may not always hold.

## Random Forest

Definition: Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy.

What it does: Aggregates predictions from several trees to reduce overfitting and enhance generalization.

Application: Used in classification, regression, and feature selection across various industries.

Advantage: Handles missing data well and works effectively with large, complex datasets.

## Decision Tree

Definition: Decision Tree is a supervised learning algorithm that splits data into branches based on decision rules.

What it does: Breaks down complex decisions into simpler structures using a tree-like model.

Application: Common in medical diagnoses, credit scoring, and business decision-making.

Limitation: Prone to overfitting, which can reduce generalizability.

## Logistic Regression

Definition: Logistic Regression is used for binary classification, predicting the probability of an event occurring.

What it does: Fits a logistic function to the data, outputting probabilities instead of direct values.

Application: Widely used in healthcare (disease prediction) and marketing (customer churn analysis).

Advantage: Simple to implement and interpretable for binary classification tasks.

## Support Vector Machine (SVM)

Definition: SVM is a supervised learning algorithm for classification and regression tasks.

What it does: Finds the optimal hyperplane that separates data points into distinct classes.

Application: Effective for image recognition, text classification, and bioinformatics.

Advantage: Works well with high-dimensional data and provides flexibility with kernel functions.