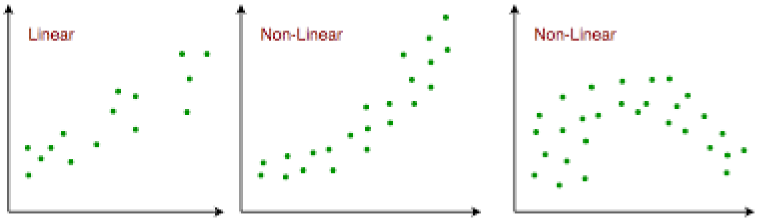
Linear Regression:

Linear regression is a widely used statistical and machine learning technique in data analytics. It is employed for modeling and analyzing the relationship between a dependent variable (target) and one or more independent variables (predictors or features). Linear regression assumes a linear relationship between the predictors and the target variable, making it a simple yet powerful tool for various data analysis tasks. Here's how linear regression works in data analytics:



1. Problem Formulation: Linear regression is often used when you want to predict or explain a continuous numeric target variable based on one or more independent variables. The problem is formulated as follows:
   - Simple Linear Regression: When there's only one predictor variable (univariate regression):

makefileCopy code

Y = β0 + β1*X + ε

- Y is the dependent variable (target).
- X is the independent variable (predictor).
- β0 is the intercept (the value of Y when X is 0).
- β1 is the slope (the change in Y for a one-unit change in X).
- ε represents the error term, accounting for the variability not explained by the linear relationship.

   - Multiple Linear Regression: When there are multiple predictor variables (multivariate regression):

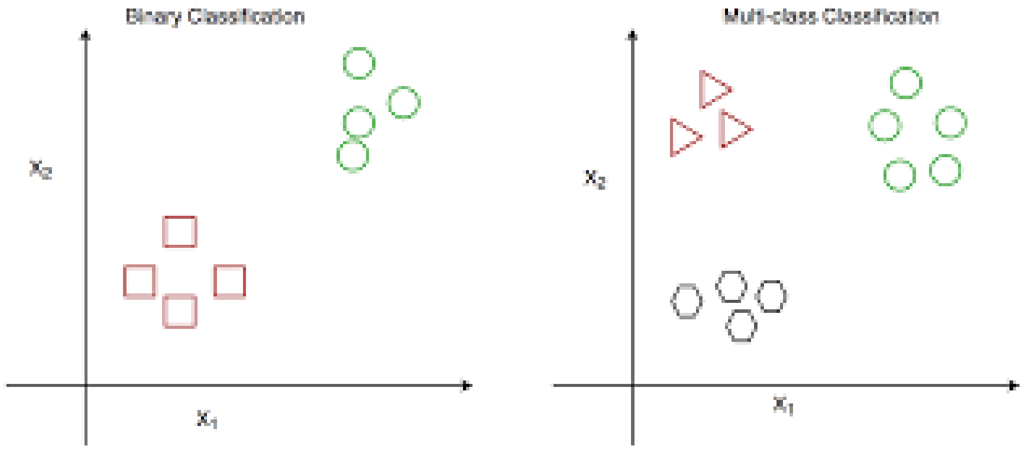
makefileCopy code

Y = β0 + β1*X1 + β2*X2 + ... + βn*Xn + ε

- X1, X2, ..., Xn are the independent variables.
- β1, β2, ..., βn are the coefficients associated with each predictor.

- ε still represents the error term.

2. Fitting the Model: The goal of linear regression is to estimate the values of the coefficients ($\beta 0$, $\beta 1$, $\beta 2$, ...) that minimize the sum of squared errors (the differences between predicted and actual values) across the training data. This process is often referred to as "training" the linear regression model.
3. Interpretation: The coefficients ($\beta 1$, $\beta 2$, ...) provide insights into the relationship between the predictors and the target variable. A positive coefficient indicates a positive correlation, while a negative coefficient indicates a negative correlation. For example, in a simple linear regression predicting exam scores based on study hours, a positive $\beta 1$ suggests that more study hours are associated with higher exam scores.
4. Model Evaluation: To assess the quality of the linear regression model, various metrics are used, including R-squared (the proportion of the variance in the dependent variable explained by the model), mean squared error (MSE), root mean squared error (RMSE), and others.
5. Assumptions: Linear regression relies on several assumptions, such as the linearity of the relationship between variables, independence of errors, constant variance of errors (homoscedasticity), and normally distributed errors. Violations of these assumptions can impact the validity of the model.
6. Applications: Linear regression is applied in various domains, including economics (predicting economic trends), finance (stock price forecasting), healthcare (predicting patient outcomes), marketing (customer behavior analysis), and more.
7. Extensions: There are extensions of linear regression, such as ridge regression and lasso regression, which introduce regularization techniques to prevent overfitting and improve model performance in cases with multicollinearity or high-dimensional data.


Binary Classification
X2
X1
Multi-class Classification
X2
X1