REGRESSION ANALYSIS

Regression analysis is a way that can be used to **determine the relationship between the predictor variable (x) and the target variable (y).**

What are Optimization Problems?

Optimization problems are mathematical problems that involve finding the best solution from a set of possible solutions.

These problems are typically formulated as maximization or minimization problems, where the goal is to **either maximize or minimize a certain objective function.**

The **objective function** is a mathematical expression describing **the quantity to be optimized**, and a set of constraints defines the set of possible solutions.

Optimization is a crucial component of many machine learning algorithms. **In machine learning, optimization is used to find the best set of parameters for a model that minimizes the difference between the model's predictions and the true values.**

- **In supervised learning**, optimization is used to find the parameters of a model that minimize the difference between the model's predictions and the true values for a given training dataset. For example, linear regression and logistic regression use optimization to find the best values of the model's coefficients. In addition, some models like decision trees, random forests, and gradient boosting models are built by iteratively adding new models to the ensemble and optimize the parameters of the new models that minimize the error on the training data.

- **In unsupervised learning**, optimization helps to find the best configuration of clusters or mapping of the data that best represents the underlying structure in the data. In **clustering**, optimization is used to find the best configuration of clusters in the data. For example, the K-Means algorithm uses an optimization technique called Lloyd's algorithm, which iteratively reassigns data points to the nearest cluster centroid and updates the cluster centroids based on the newly assigned points. Similarly, other clustering algorithms such as hierarchical clustering, density-based clustering, and Gaussian mixture models also use optimization techniques to find the best clustering solution. In **dimensionality reduction**, optimization finds the best data mapping from a high- to a lower-dimensional space. For example, Principal Component Analysis (PCA) uses Singular Value Decomposition (SVD), an optimization technique, to find the best linear combination of the original variables that explains the most variance in the data. Additionally, other dimensionality reduction techniques like Linear Discriminant Analysis (LDA) and t-distributed Stochastic Neighbor Embedding (t-SNE) also use optimization techniques to find the best representation of the data in a lower-dimensional space.

- **In deep learning**, optimization is used to find the best set of parameters for neural networks, which is typically done using gradient-based optimization algorithms such as stochastic gradient descent (SGD) or Adam/Adagrad/RMSProp, etc.

Why do we Need OLS?

The ordinary least squares (OLS) algorithm is a method for **estimating the parameters of a linear regression model.**

The OLS algorithm aims to find the values of the linear regression model's parameters (i.e., the coefficients) **that minimize the sum of the squared residuals**.

The residuals are the differences between the observed values of the dependent variable and the predicted values of the dependent variable given the independent variables.

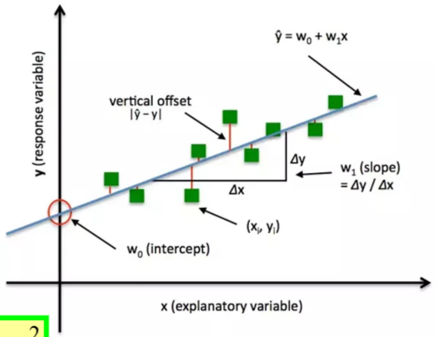
It is important to note that the OLS algorithm assumes that the errors are normally distributed with zero mean and constant variance and that there is no multicollinearity (high correlation) among the independent variables.

Possible Loss (Cost) functions for Regression

(1) Sum of errors (SE): $L = \sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)$

(2) Sum of Absolute Errors (SAE): $L = \sum_{i=1}^{N}\left|\hat{Y}_i - Y_i\right|$

(3) Sum of Squares of Errors (SSE): $L = \sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2$



(4) Mean of Squares of Errors (MSE): $L = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2$

X1	Y
1	4.8
3	11.4
5	17.5
..	…

(5) Root Mean of Squares of Errors (RMSE): $L = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}$