Feature Engineering

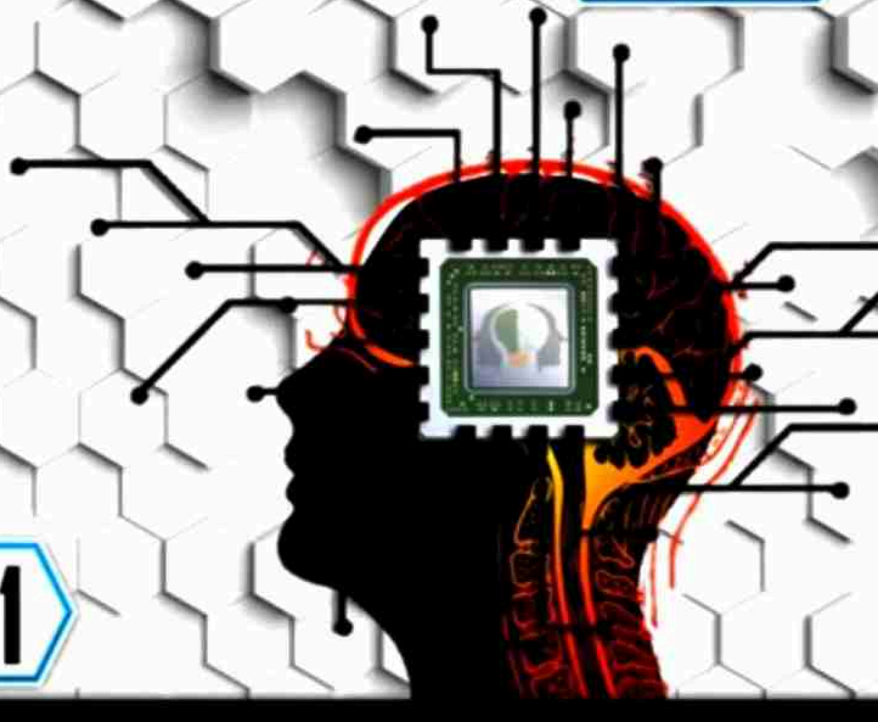
Feature Scaling

What, Use, Types, Adv, Diss

Tutorial

E/हिं

Part-1



Agenda

- What Is Feature?
- What is Scaling?
- What is Feature Scaling?
- Why & Use Feature Scaling?
- Type Of Feature Scaling?
- Advantages
- Disadvantages
- Algorithm Support Feature Scaling

What is Feature Scaling?

- **Feature Scaling is a method to scale numeric features in the same scale or range *(like:-1 to 1, 0 to 1)*.**
- **This last step involved in Data Preprocessing and before ML model training.**
- **It is also called as data normalization.**
- **We apply Feature Scaling on independent variables.**
- **We fit feature scaling with train data and transform on train and test data.**

Why Feature Scaling?

- **The scale of raw features is different according to its units.**
- **Machine Learning algorithms can't understand features units, understand only numbers.**
- ***Ex: If hight 140cm and 8.2feet***
- ***ML Algorithms understand numbers then 140 > 8.2***

Male Hight In Feet	Female Hight In CM	Life Span In Year
8.0	150	30
8.5	165	40
7.9	170	36
8.2	140	41

Which ML Algorithms Required Feature Scaling ?

- **Those Algorithms Calculate Distance**
  - K-Nearest Neighbors (KNN)
  - K-Means
  - Support Vector Machine (SVM)
  - Principal Component Analysis(PCA)
  - Linear Discriminant Analysis

$$d^E(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

- **Gradient Descent Based Algorithms**
  - Linear Regression,
  - Logistic Regression
  - Neural Network

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

- **Tree Based Algorithms not required FS**
  - Decision Tree, Random Forest, XGBoost

Types of Feature Scaling