* EDA - Exploratory Data Analysis

Types

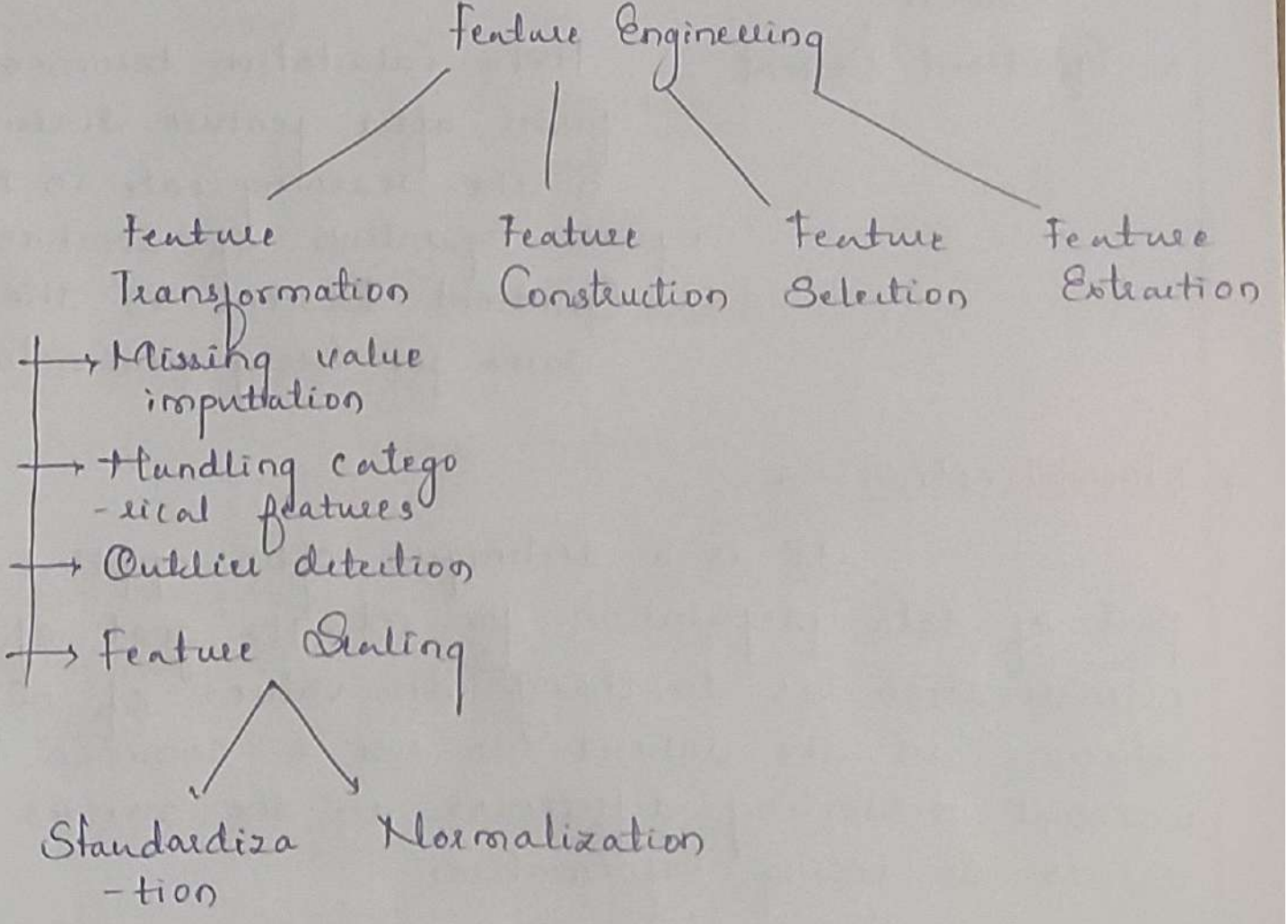
1 Univariate Analysis - It helps to describe data & find patterns within a single feature

2 Bivariate

3 Multivariate

Feature Engineering

The process of using domain knowledge to extract features from raw data. These features can be used to improve the performance of machine learning Algorithms



- Standardization

$$x_i' = \frac{x_i - \bar{x}}{\sigma}$$

Use of standardization.

Algorithm	Reason
1. K-Means	Use the Euclidean distance measure
2. K-nearest-Neighbours	Measure the distances b/w pairs of samples & these distances are influenced.
3. principal Component Analysis	Try to get the feature with maximum variance
4. Artificial Neural Network	Apply Gradient Descent
5. Gradient Descent	Theta Calculation becomes faster after feature scaling & the learning rate in the update Equation of stochastic Gradient Descent is the same for every parameter

+ Normalization :→

It is a technique often applied as part of data preparation for ml. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information.

- Minmaxscaling →
- Mean Normalization
- Max absolute
- Robust Scaling

* Minmaxscaling

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

* Mean Normalization :

$$x_i' = \frac{x_i - x_{mean}}{x_{max} - x_{min}}$$

* Robust Scaling

$$x_i' = \frac{x_i - x_{median}}{IQR}$$

* Normalization Vs Standardization

Minmax → image → CNN → 0-255

* Label Encoding
  - Nominal Data - Categories without inherent order
  - Ordinal " = " with a natural Order

* One-hot Encoding

Converts categorical data into a numerical format by creating a binary column for each unique category