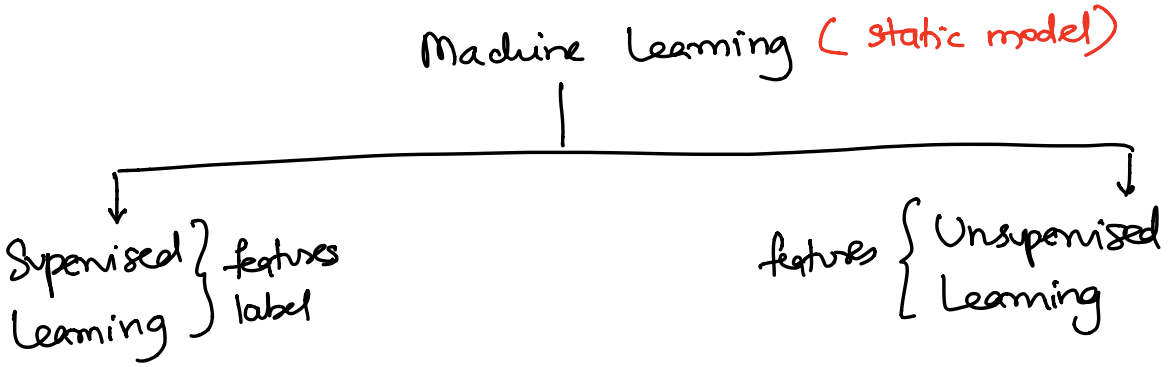
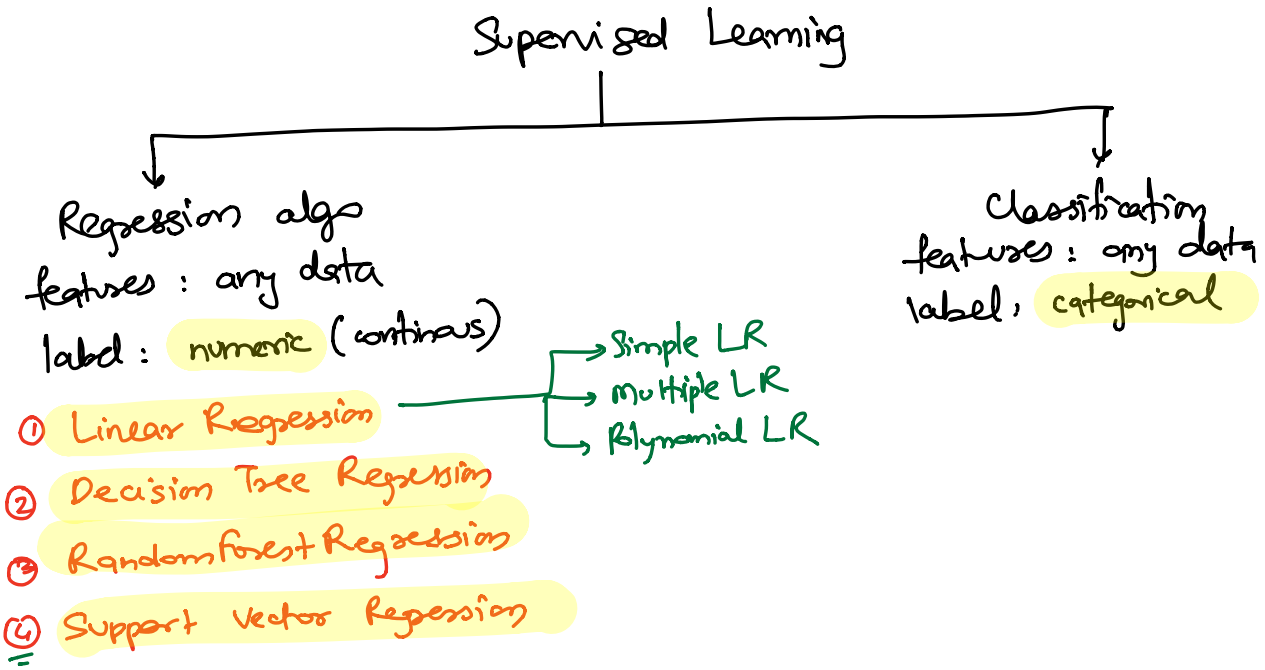
- ~~Outliers~~
- Supervised Learning
- Regression.

OUTLIERS → An outlier is Extreme values in your given column in the dataset.

IQR, boxplot, check domain and get appropriate range





⑤ KNN Regressor
⑥ ElasticNet Regression
⑦ Ridge Regression
⑧ Lasso Regression
⑨ Xgboost (Extreme Gradient Boosting)

Linear Regression

getting the best fit line for the given use-case

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

$$\text{Profit} = b_0 + b_1(R\&D) + b_2(Adm) + b_3(\text{Marketing}) + b_4(NY) + b_5(Cal) + b_6(\text{Florida})$$

b_0: intercept; b_1, b_2, b_3: coeffs; b_4: coeff; b_5: coeff; b_6: coeff

Multiple Linear Regression } multiple features and one label

$$\text{Salary} = b_0 + b_1(\text{YearsExp})$$

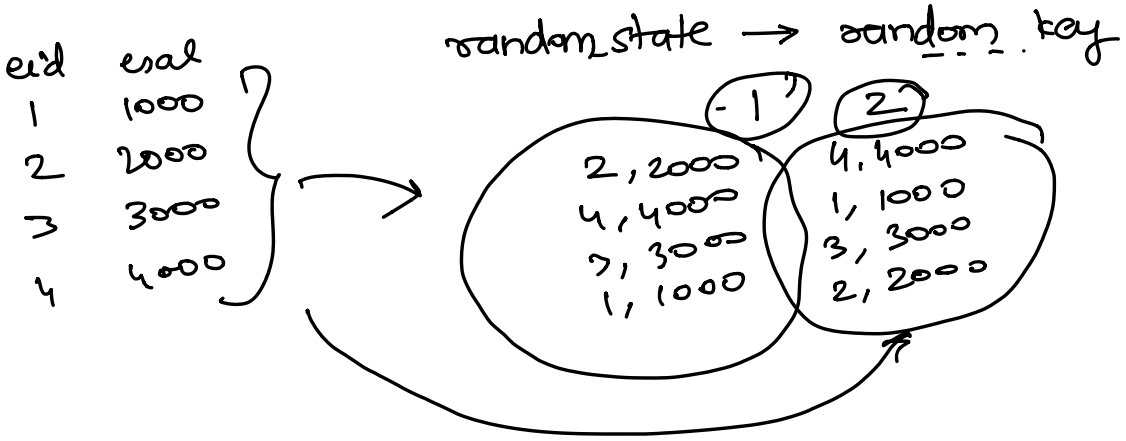
b_0: int; b_1: coef

} Simple Linear Regression — one feature and one Label

$$\text{Salary} = b_0 + b_1(YE) + b_2(YE)^2 + \dots + b_n(YE)^n$$

} Polynomial Linear Regression — one feature and one label

degree of features



Equation of linear regression

$$y = b_0 + b_1 x_1$$

$$y = b_0 x^0 + b_1 x_1 + b_1 x_1^2 + b_1 x_1^3$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{bmatrix}$$

Statistics → Sample

(perfect)

when it covers all possible scenario

Usecase

by sampling techniques

- → Thompson Sampling
- → Neilson - ? sampling