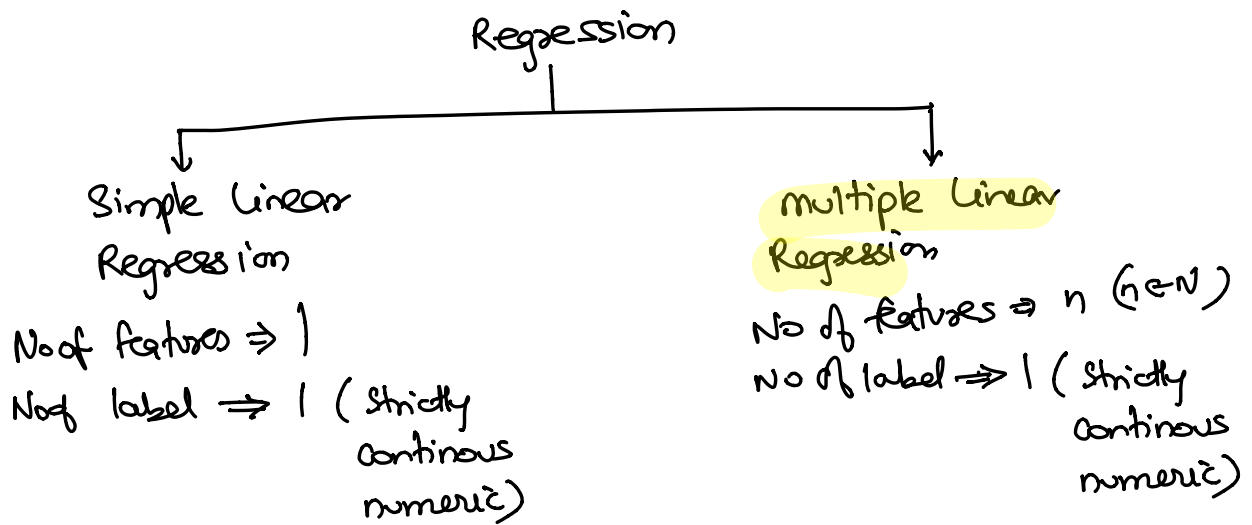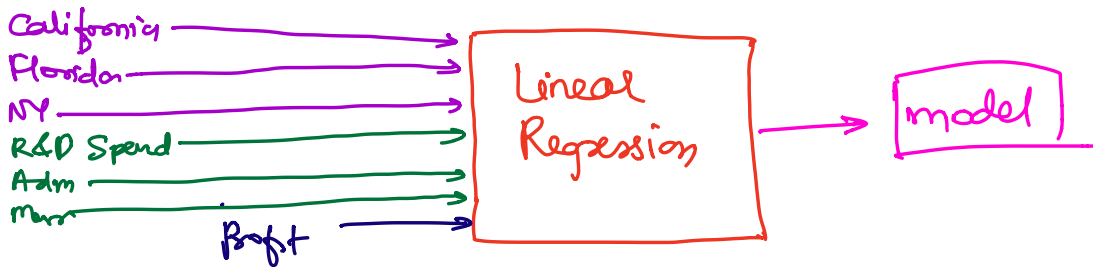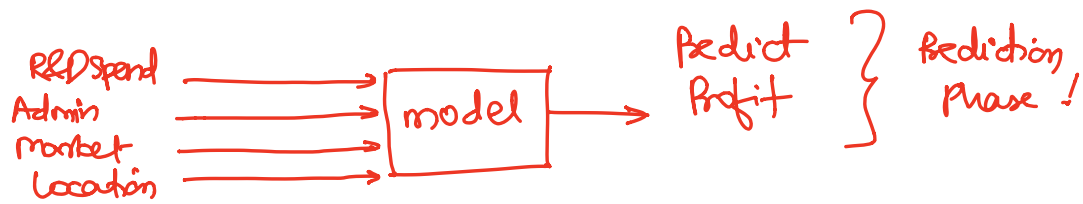- What is a ML model?
- How to create the ML model?
- How to optimize the model by controlling the sampling?
- How to deploy the model.
- How to use a deployed model

Regression

Simple Linear Regression

No of features $\Rightarrow$ 1
No of label $\Rightarrow$ 1 (Strictly continous numeric)

multiple Linear Regression

No of features $\Rightarrow$ n ($n \in N$)
No of label $\Rightarrow$ 1 (Strictly continous numeric)

Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n$$

$$\ldots (n \in N)$$

y - intercept

co-eff. of respective features

n represents no. of features

R&DSpend $\longrightarrow$
Admin $\longrightarrow$ model $\longrightarrow$ Predict Profit $\Big\}$ Prediction Phase !
Market $\longrightarrow$
Location $\longrightarrow$

California $\longrightarrow$
Florida $\longrightarrow$
NY $\longrightarrow$ Linear Regression $\longrightarrow$ model
R&D Spend $\longrightarrow$
Adm $\longrightarrow$
Mar $\longrightarrow$
Profit $\longrightarrow$

HYPOTHESIS $\longrightarrow$ ASSUMPTIONS.

HYPOTHESIS TESTING $\longrightarrow$ TESTING YOUR ASSUMPTIONS

How to do HYPOTHESIS TESTING $\rightarrow$ Using STATISTICAL $\Big\}$ $\longrightarrow$ * TESTING

Its all about finding whether the said statistical concept/technique is present or absent in the data.

① Test for Normalization
② Test for co-relation
③ Test for feature elimination

HYPOTHESIS   Q// Is salt sweet? —yes
TESTING   [Ass]→ Salt is sour (+ve/-ve)

↓ positive ☺   ↓ negative ☹

ALTERNATE   (your   NULL   (your
HYPOTHESIS   assumption is   HYPOTHESIS assumption
             correct)   is wrong)
Salt is sour   Salt is NOT sour
                                                     sla

eg: R&D Spend & Profit are related or not.
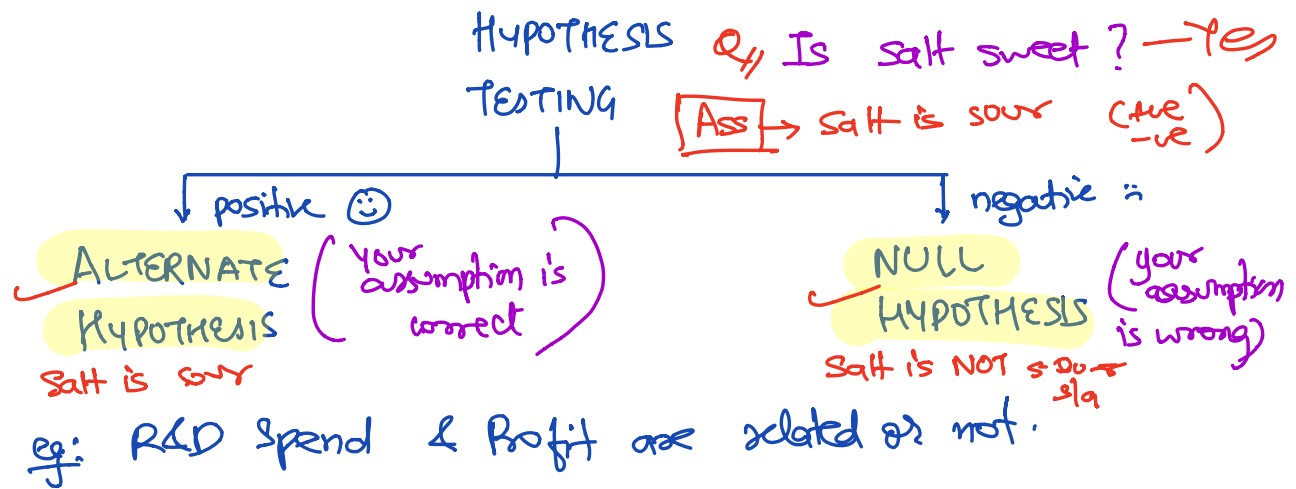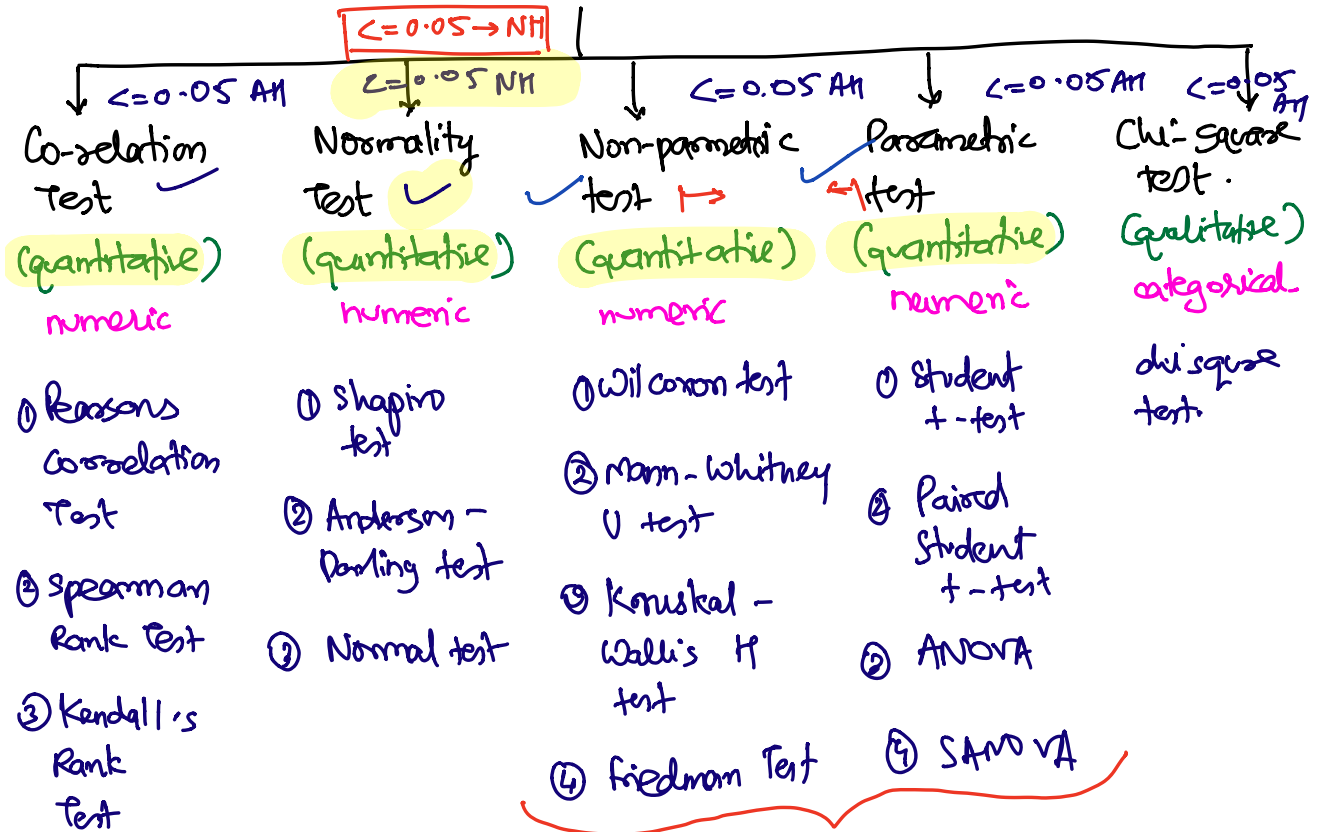
NULL HYPO: R&D Spend & Profit are NOT
             having Linear Relationship

ALTERNATE HYPO: R&D Spend & Profit is Linear
                in Nature.

50_startup. csv

R&D Spend   AdmSpend   MarkSpend   Profit   State
(Normally   (Normally   (Normally   (Normally (Normally
 Dist)   Dist)   Dist)   Dist)   Dist)

Statistical Test

<=0.05→NM

<=0.05 AM	<=0.05 NM	<=0.05 AM	<=0.05 AM	<=0.05 AM
Co-relation Test	**Normality Test**	**Non-parametric test**	**Parametric t-test**	**Chi-square test.**
(quantitative)	(quantitative)	(quantitative)	(quantitative)	(qualitative)
numeric	numeric	numeric	numeric	categorical

Co-relation Test (quantitative) numeric

① Reasons Correlation Test

② Spearman Rank Test

③ Kendall's Rank Test

Normality Test (quantitative) numeric

① Shapiro test

② Anderson-Darling test

③ Normal test

Non-parametric test (quantitative) numeric

① Wilcoxon test

② Mann-Whitney U test

③ Kruskal-Wallis H test

④ Friedman Test

Parametric t-test (quantitative) numeric

① Student t-test

② Paired Student t-test

③ ANOVA

④ SANOVA

Chi-square test (qualitative) categorical

chi square test

if col is normally dist
      do Parametric test
else
      do Non-parametric test

Statistical modelling

(feature selection & elimination) Regression → Correlation test

Normality

Classification (feature selection & elimination)

Y → Parametric

N

Non Parametric