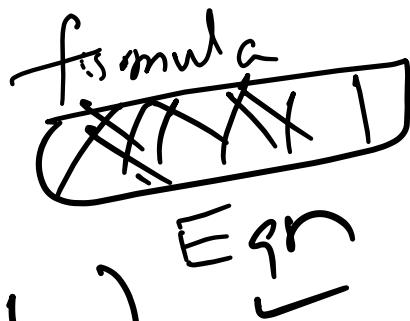
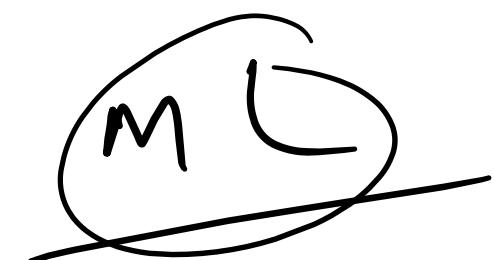
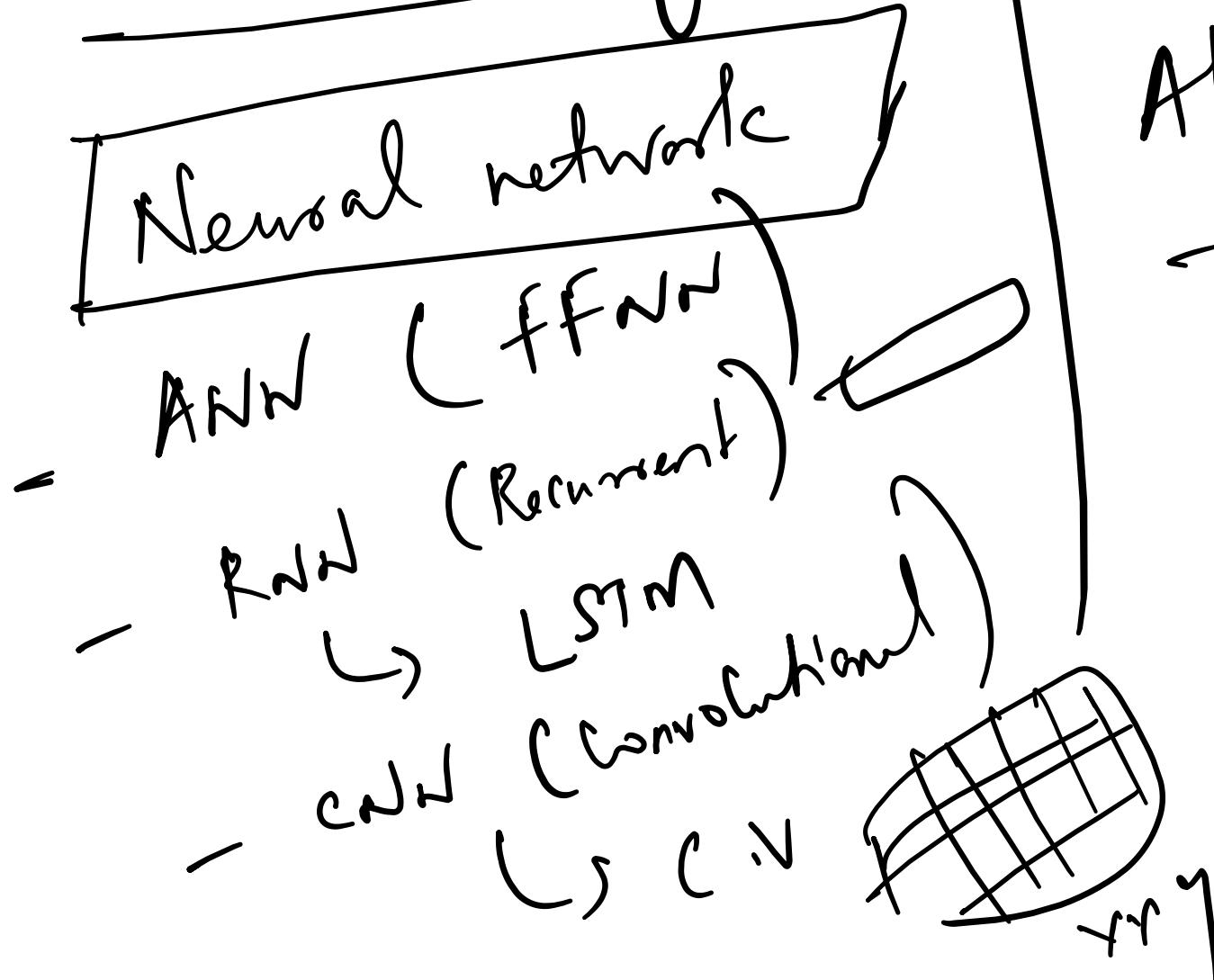
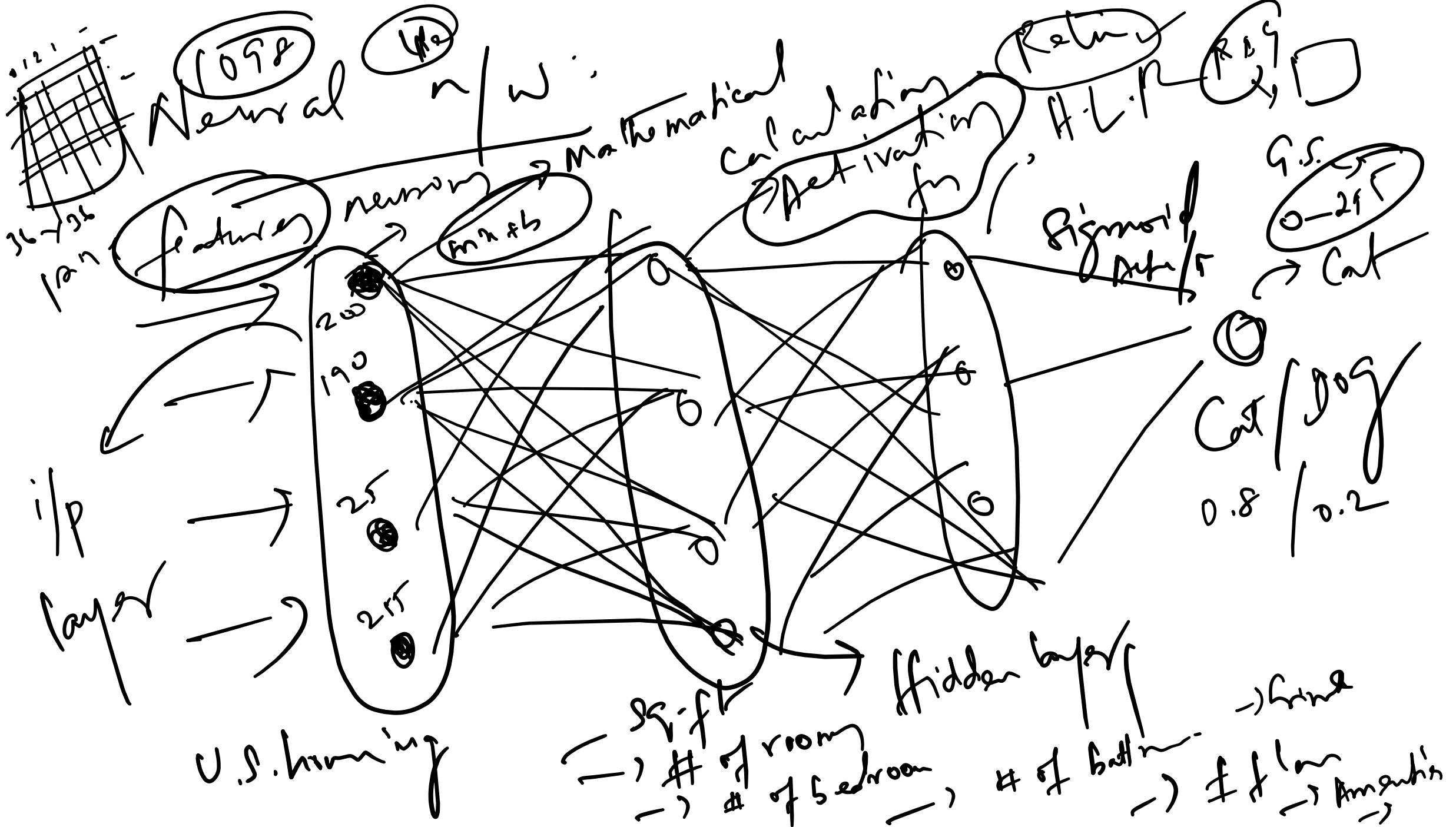


Deep Learning



- Algorithms** (m)
- L.R. / Cross Poly
 - Log Poly
 - KNN
 - D.T / R.F
 - boosting



ML

Supervised

- Labelled ✓
- Target ✓

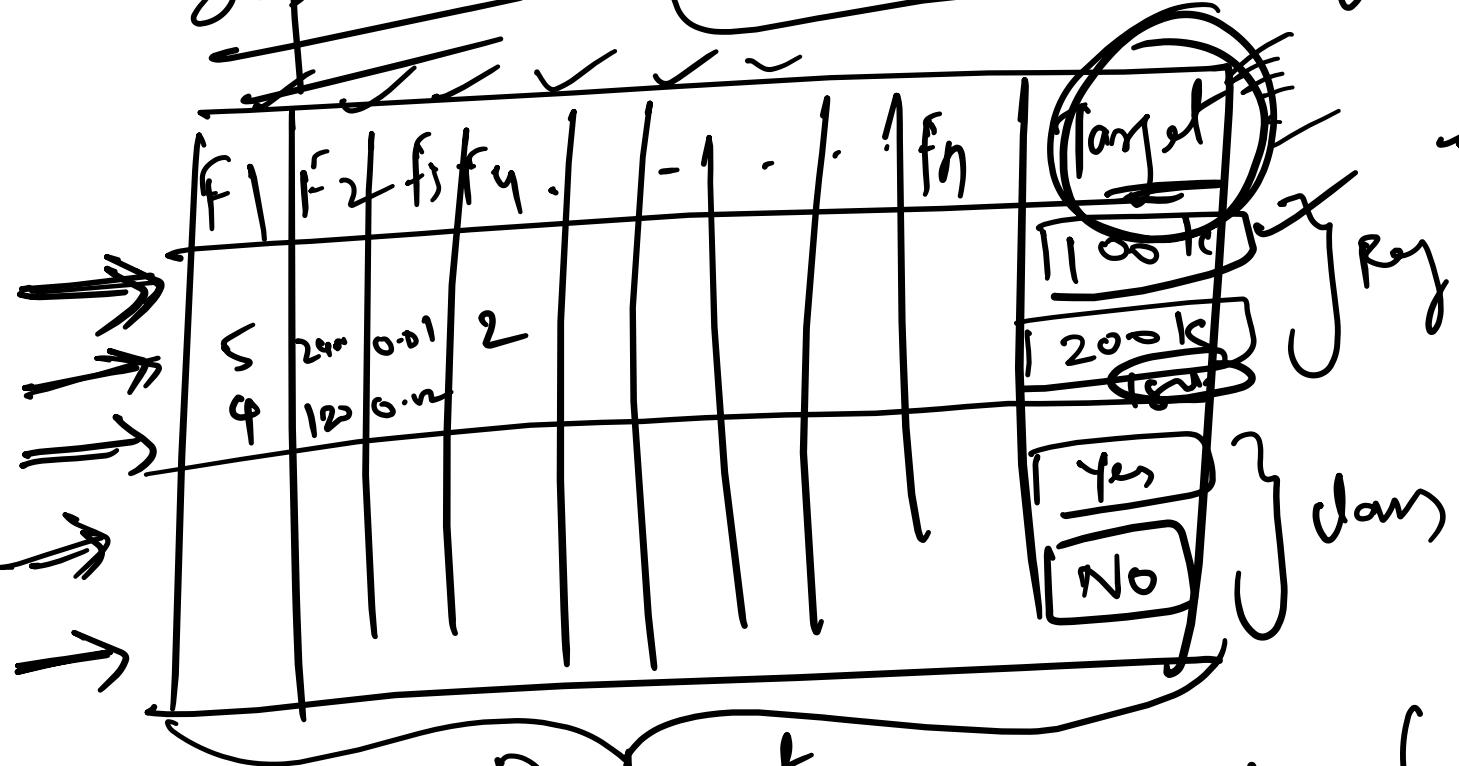
Unsupervised

- Unlabelled ✓
- Target ✗

Supervised

We'll have a Target variable

Regression (Continuum)



1

2

Classification

(Categorical)

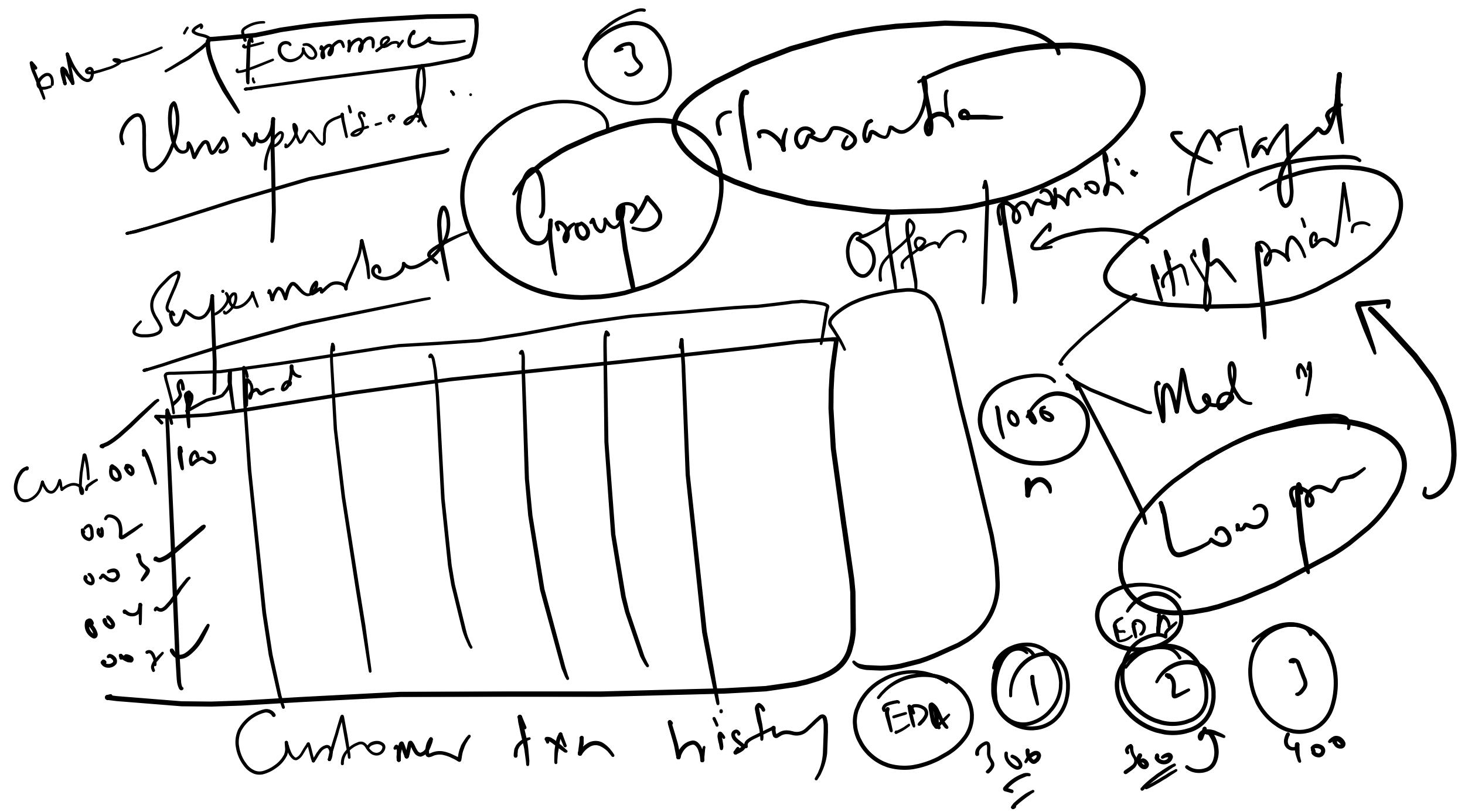
Class A
B
C₁
C₂

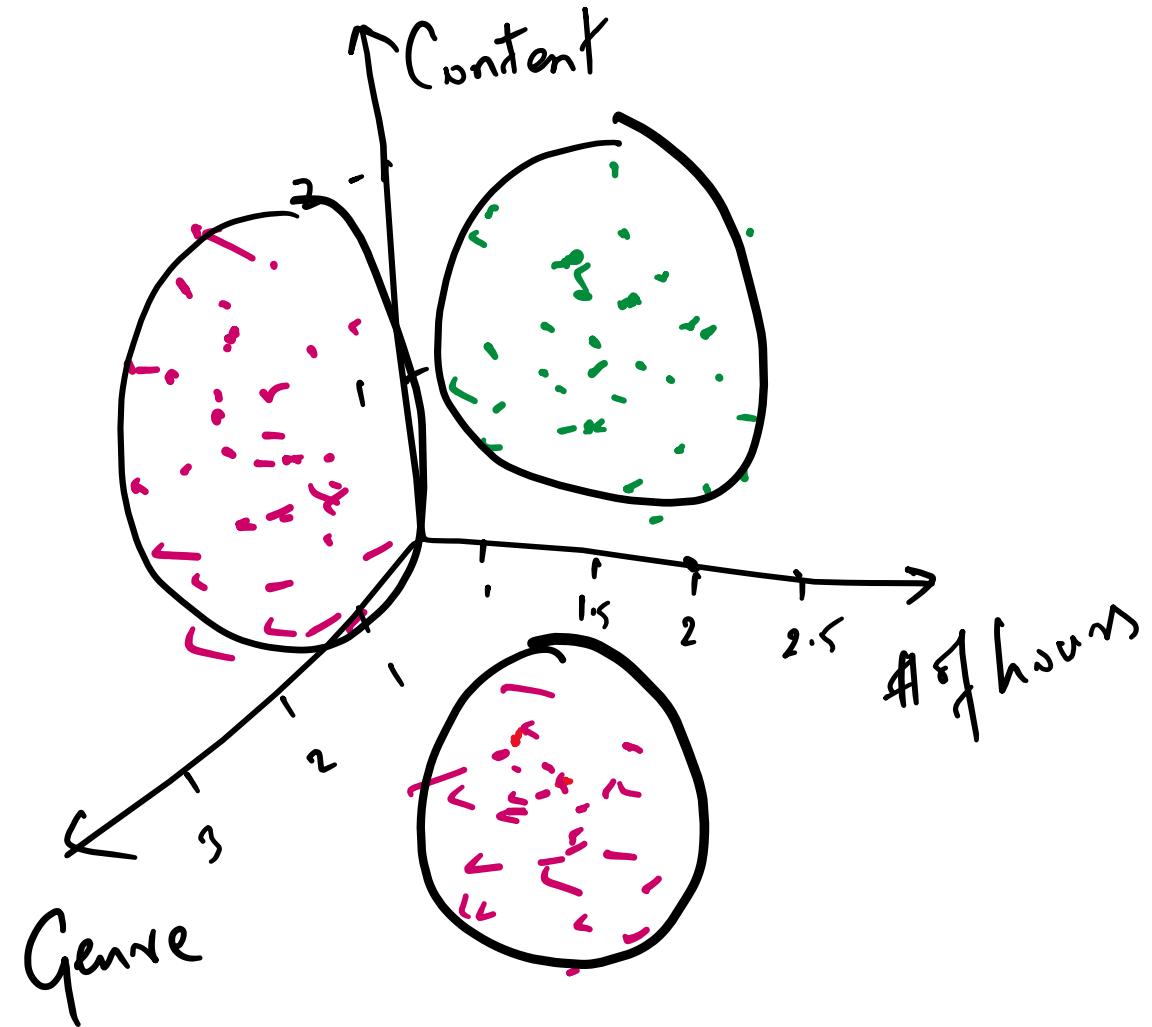
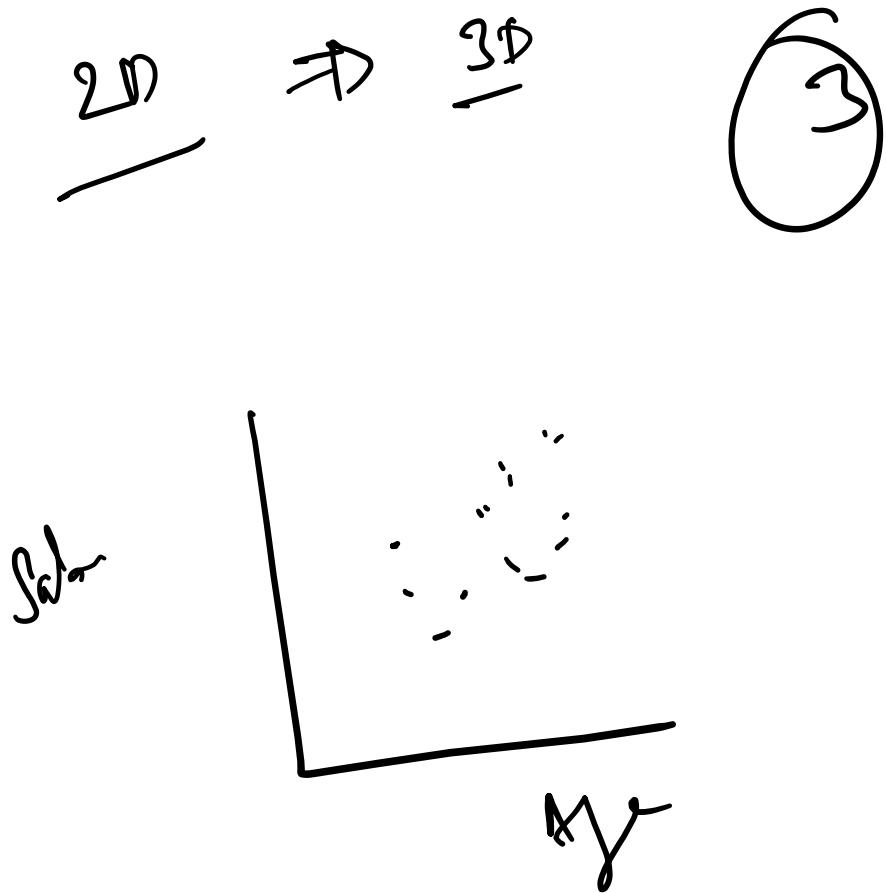
(Number) $b - \alpha$

0 / 1 (Binary)

→ Sales of a product

→ Yes / No (Disease or not)





Observation

Customer

100

A hand-drawn diagram featuring a large red circle containing the number '100'. To the left of the circle, the word 'Observation' is written in red. To the right of the circle, the word 'Customer' is written in red.

Overfitting:

Memorizing the features
(Data)

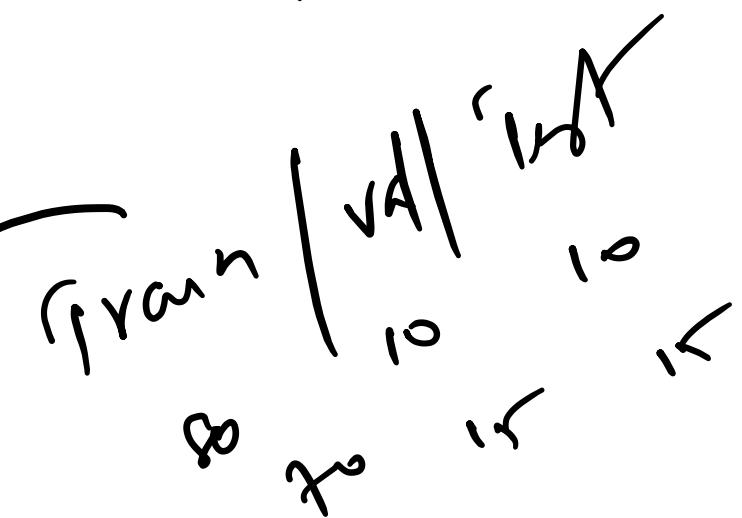
Train will perform well
Test will fail

Underfitting:

Not learning at all

Both will fail.

Modeling



Private data \rightarrow from business 85

Commercial data \rightarrow Data collected by certain org. 85.5
86 ✓

- Reuters, Dun & Bradstreet, Nielson, Change
Healthcare

(Macro-economic variable)

Open source data: - Research & teaching purposes

→ S3:

- Object level storage ✓
- Store as much data as possible eg: .csv
- AWS mgmt console → (discoverability)
- API to call S3

Sagemaker → csv

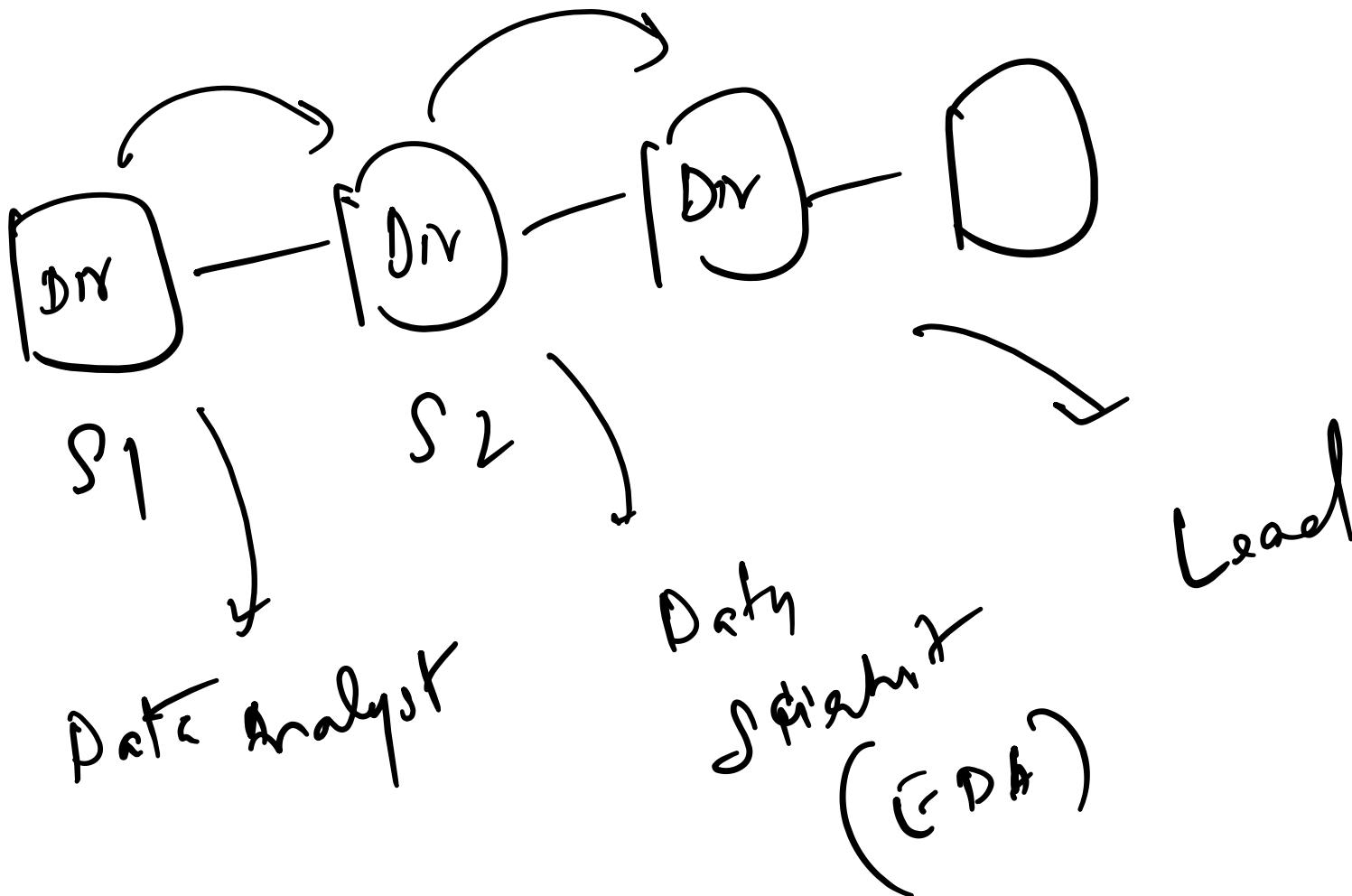
FSX:

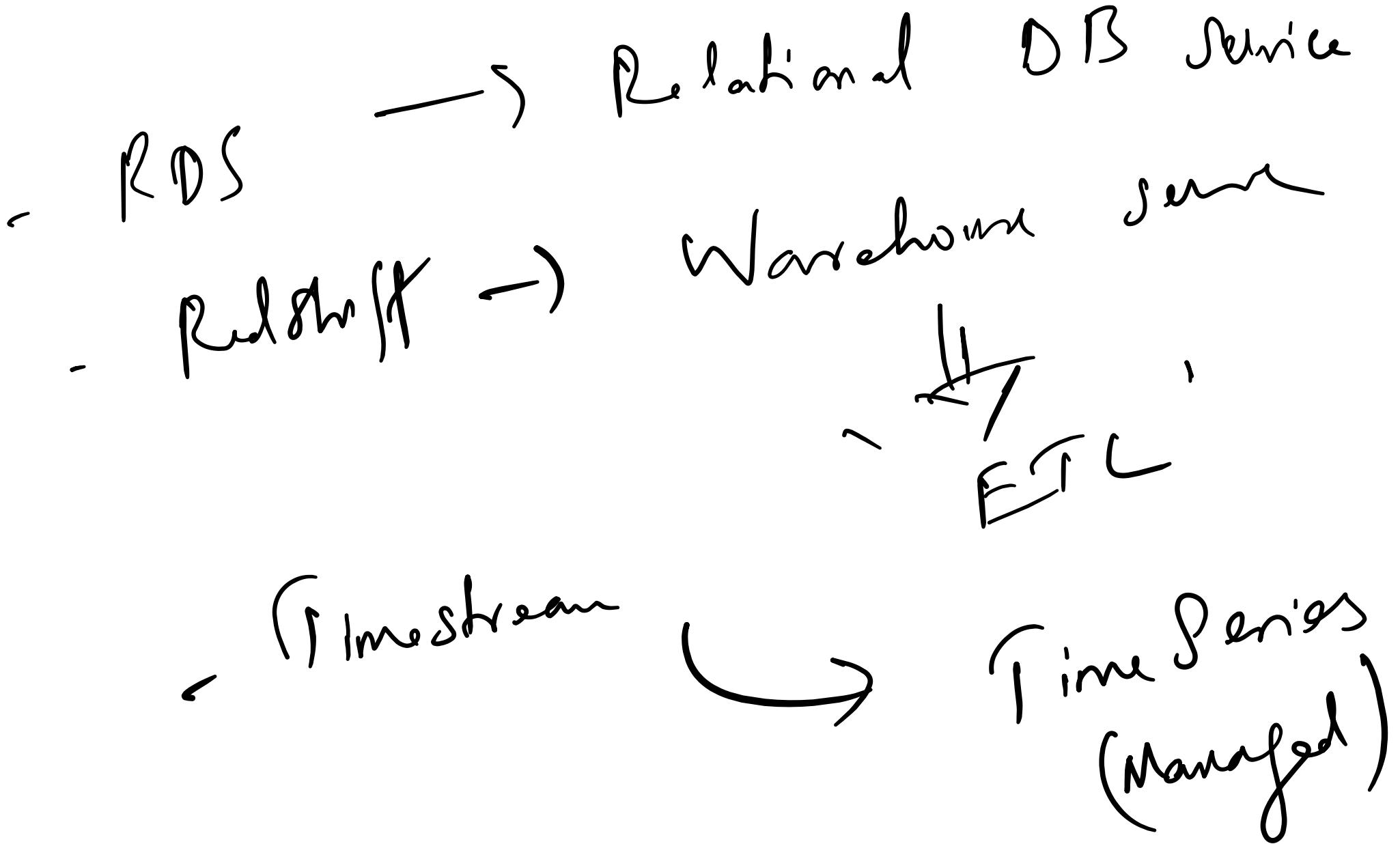
→ Recommended to use FSX

→ File system service

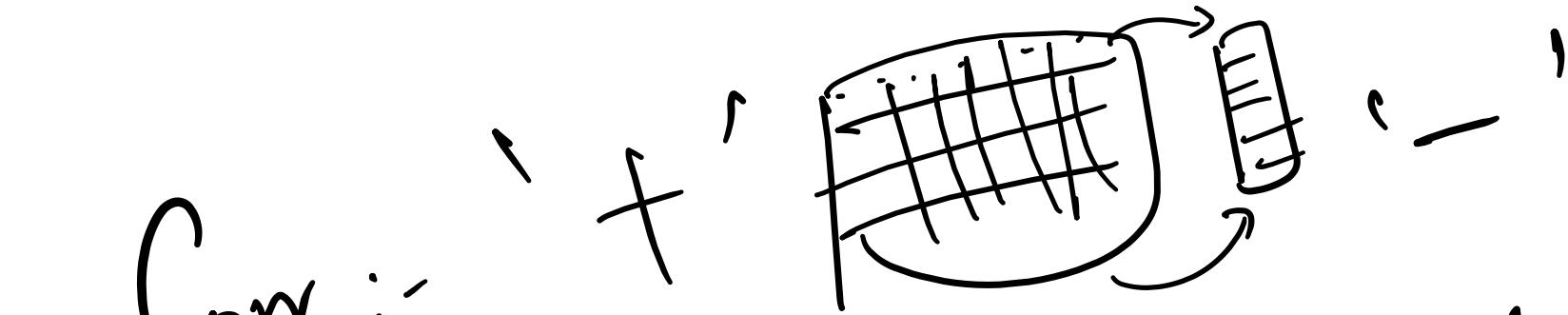
↳ High speed for training
jobs.

EFS: -
- Similar to FSx



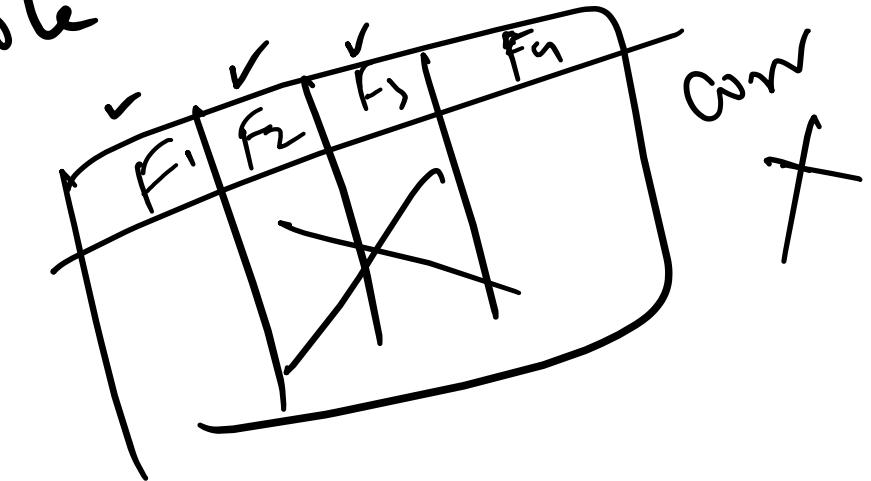


Con :- '+' '-'

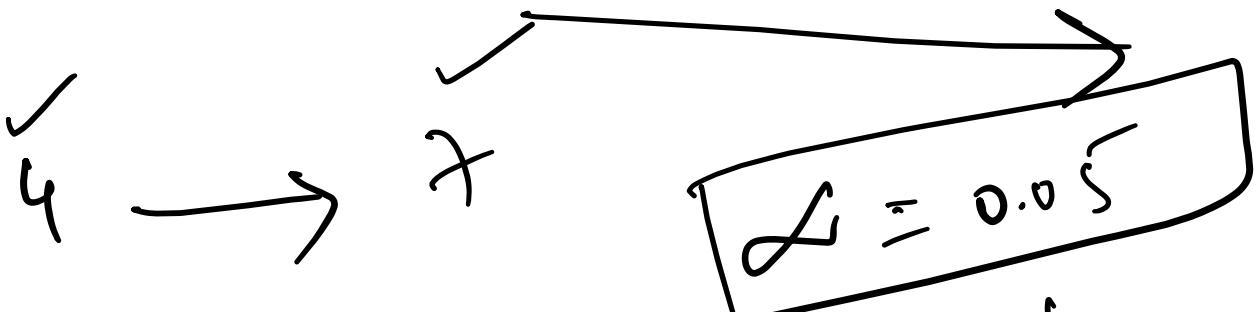


- good must have b/w features & target
(predictor var) (Response var)

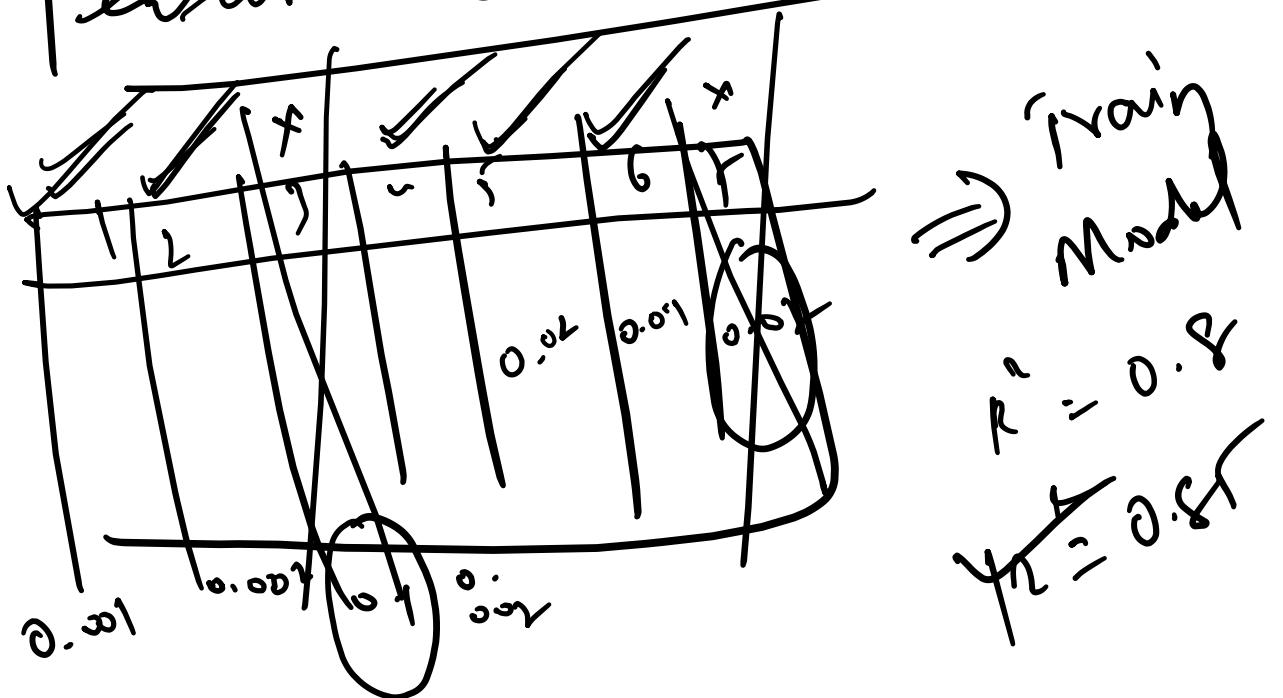
- Should not be
correlation among the variable



Feature Extraction -



Feature Selection :



Feature Extraction:

→ Builds valuable information from the raw data by reformatting, combining and transforming primary features into new ones

2021

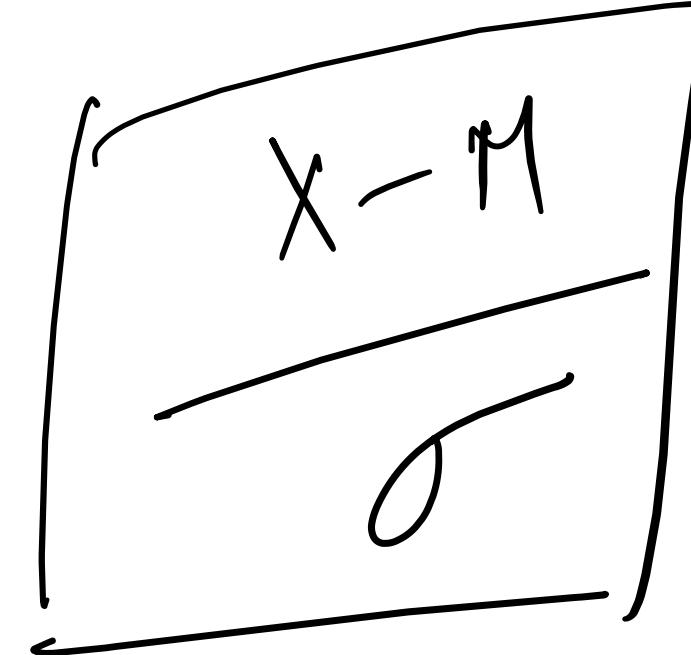
	F_1	F_2	F_3
# of doors	Purchases	Milk	
2	1400		
2	2000		
4	1600		
4	2500		
4	2600		

~~Order~~ \times ~~Variable~~

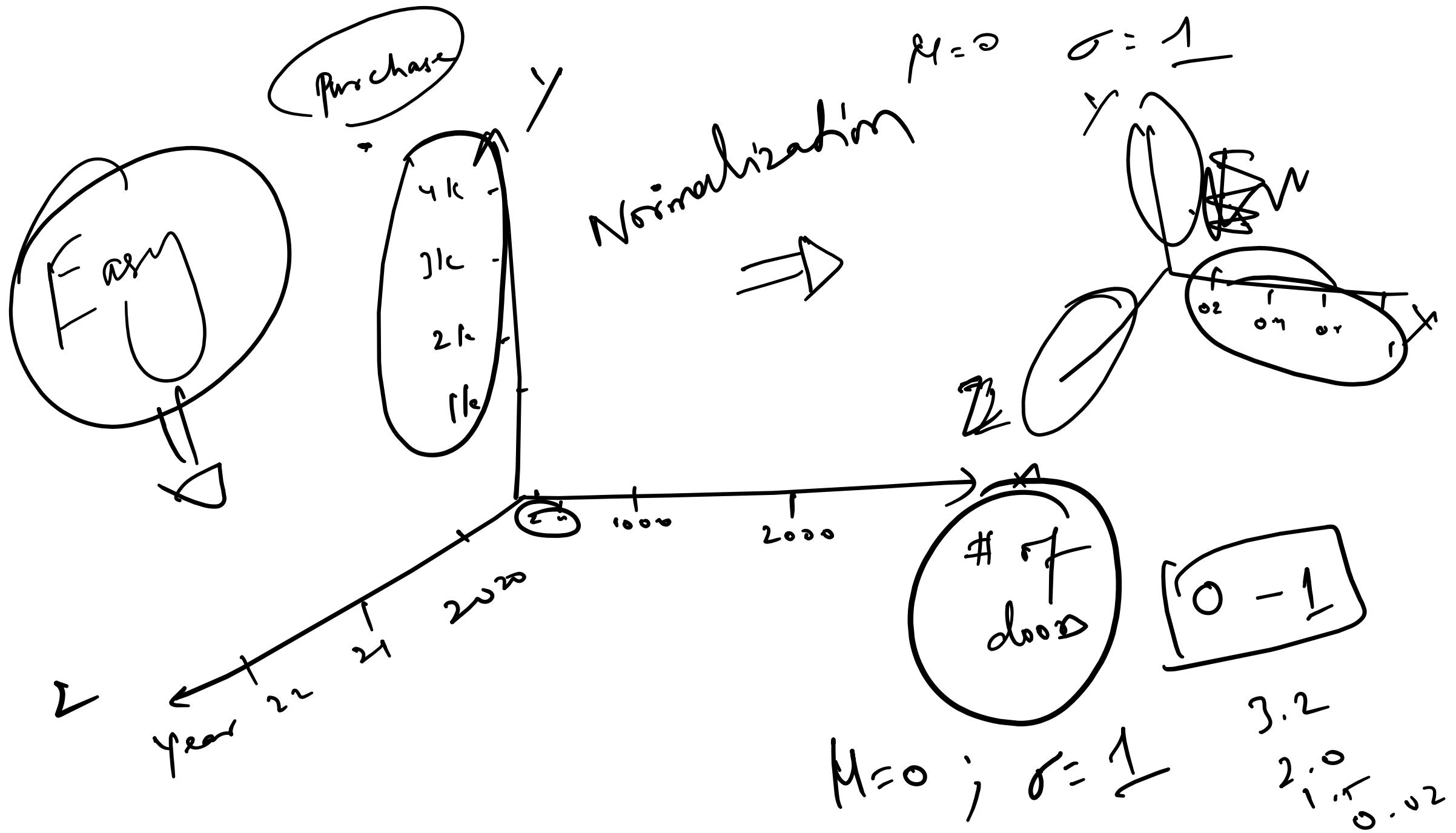
- Min max

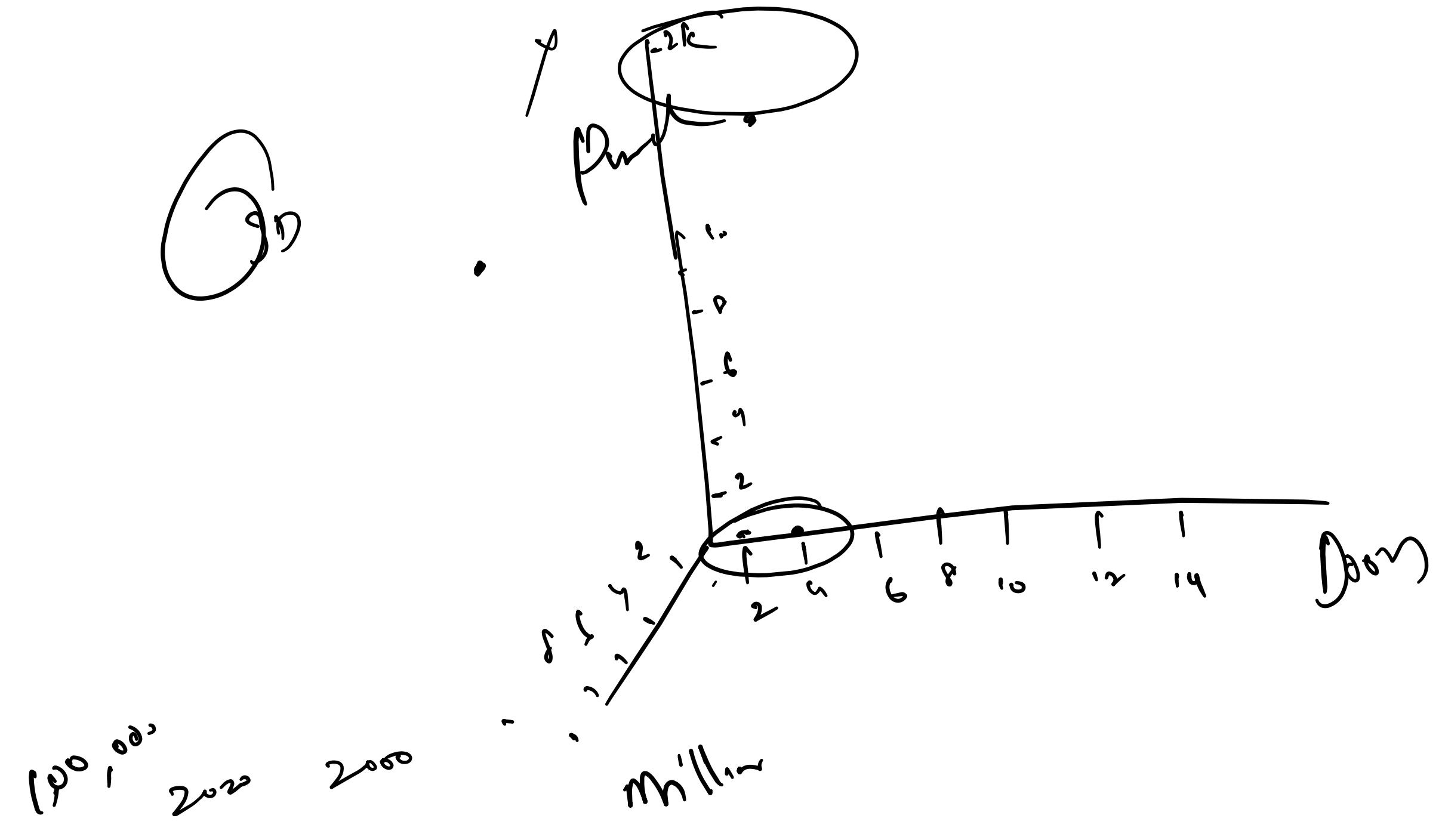
Normalization

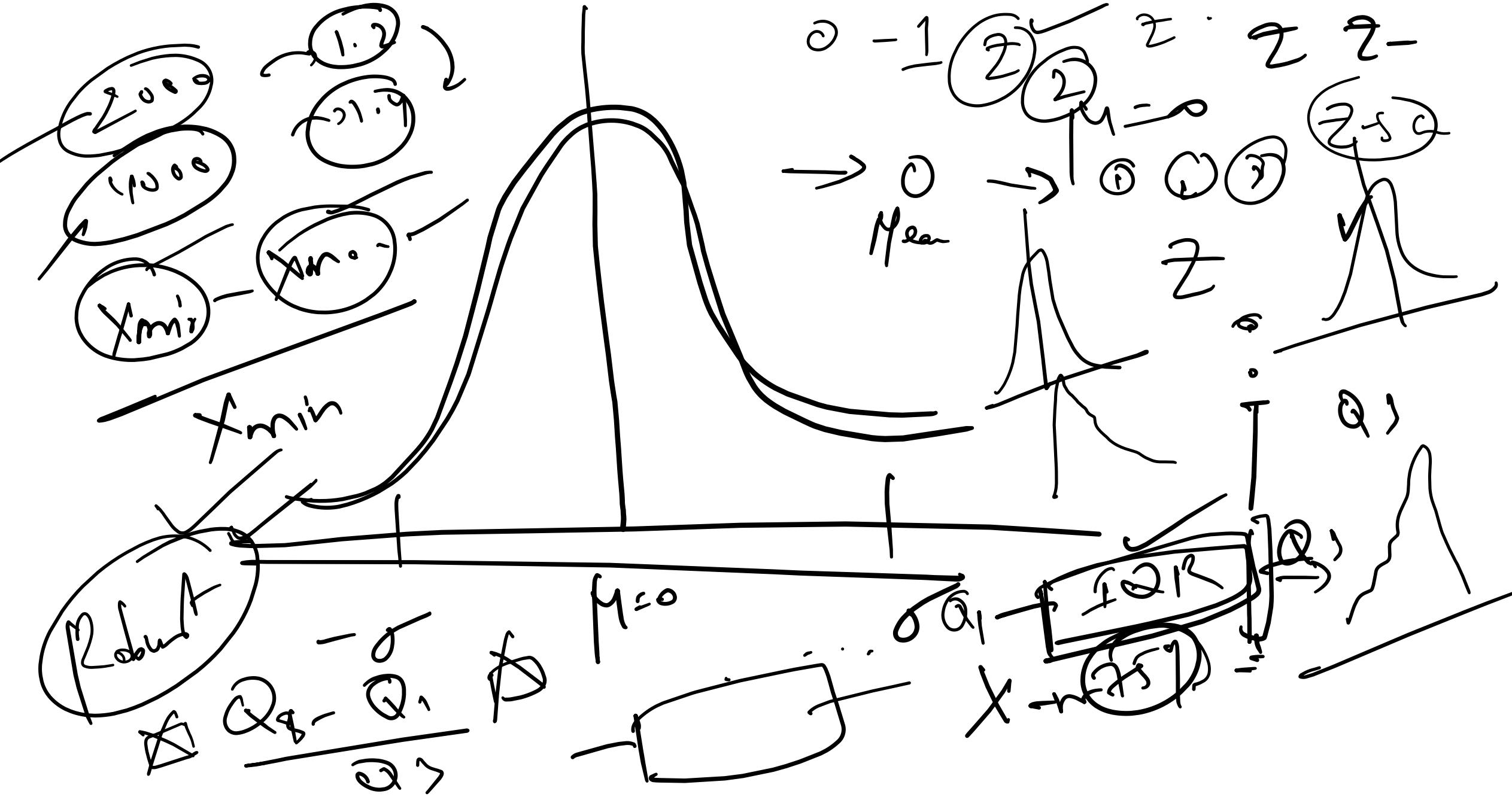
- Z score



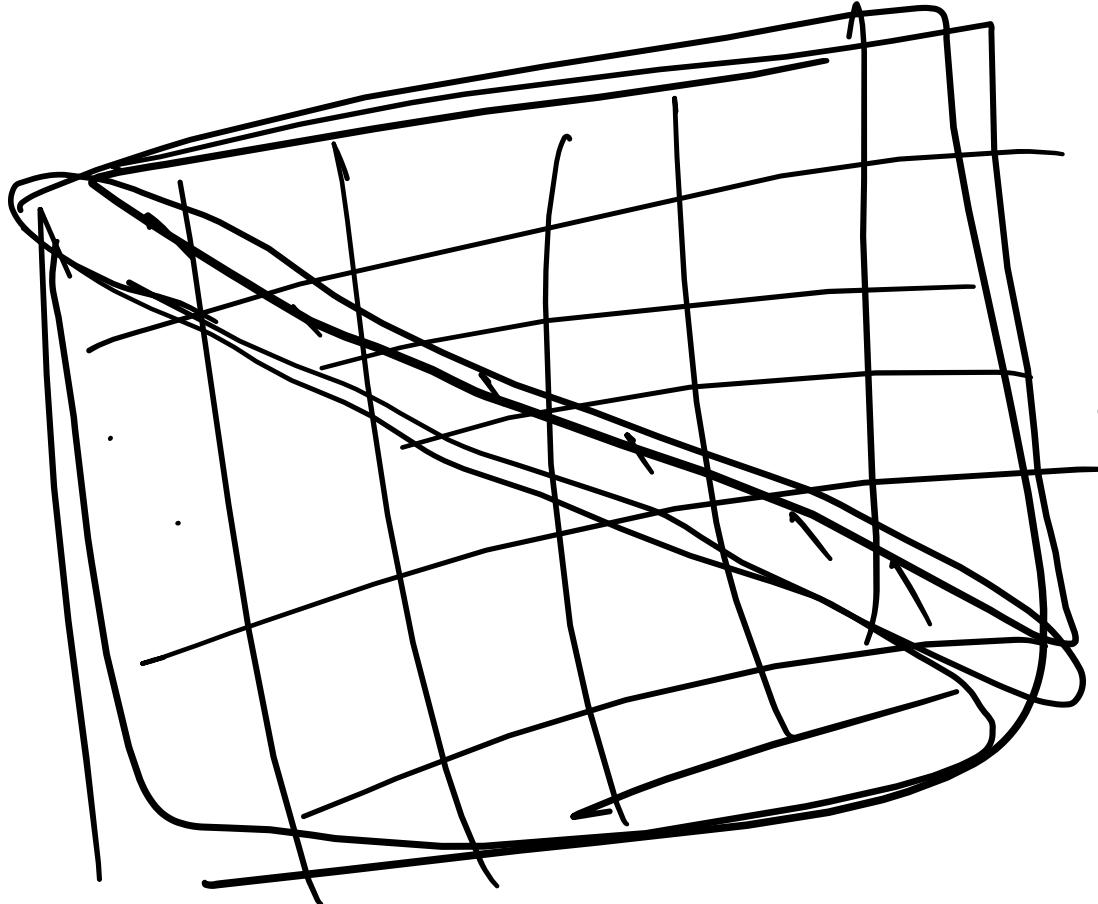
 Recommended







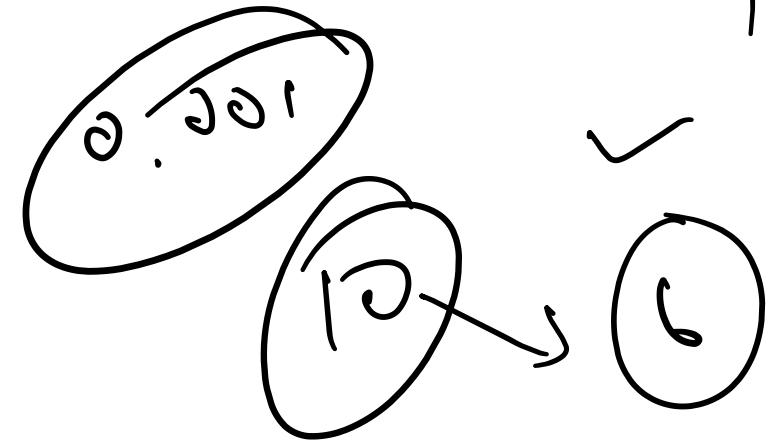
$F_1, F_2, F_3 \dots$



Select_kbest

df. Corr ()

x^1
ANOVA
 $\chi^2 - p$



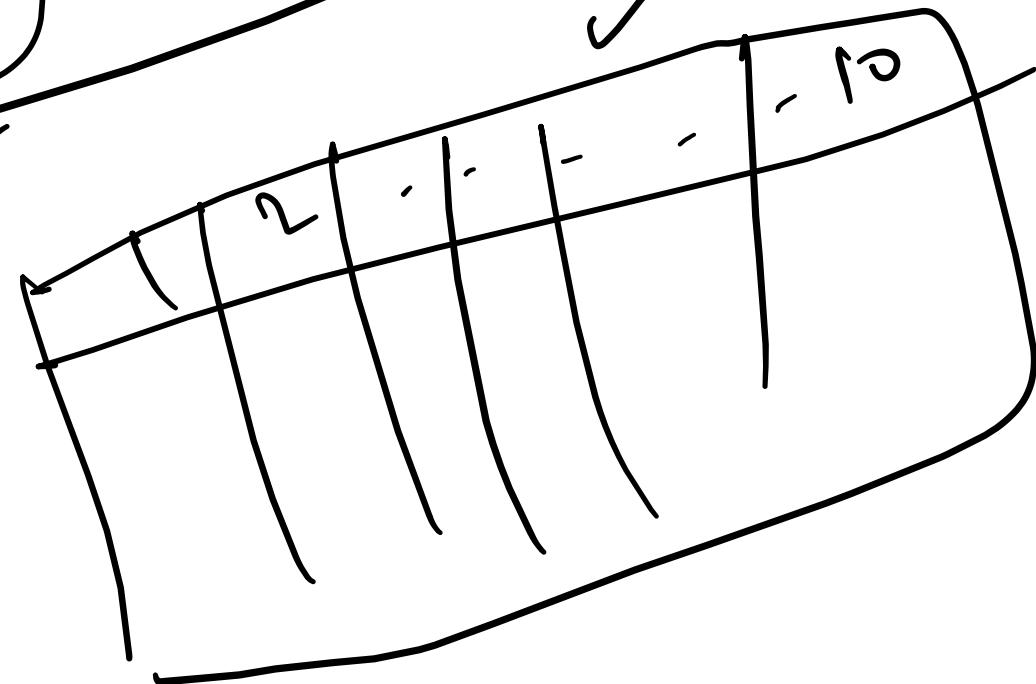
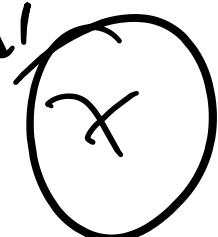
Forwd:

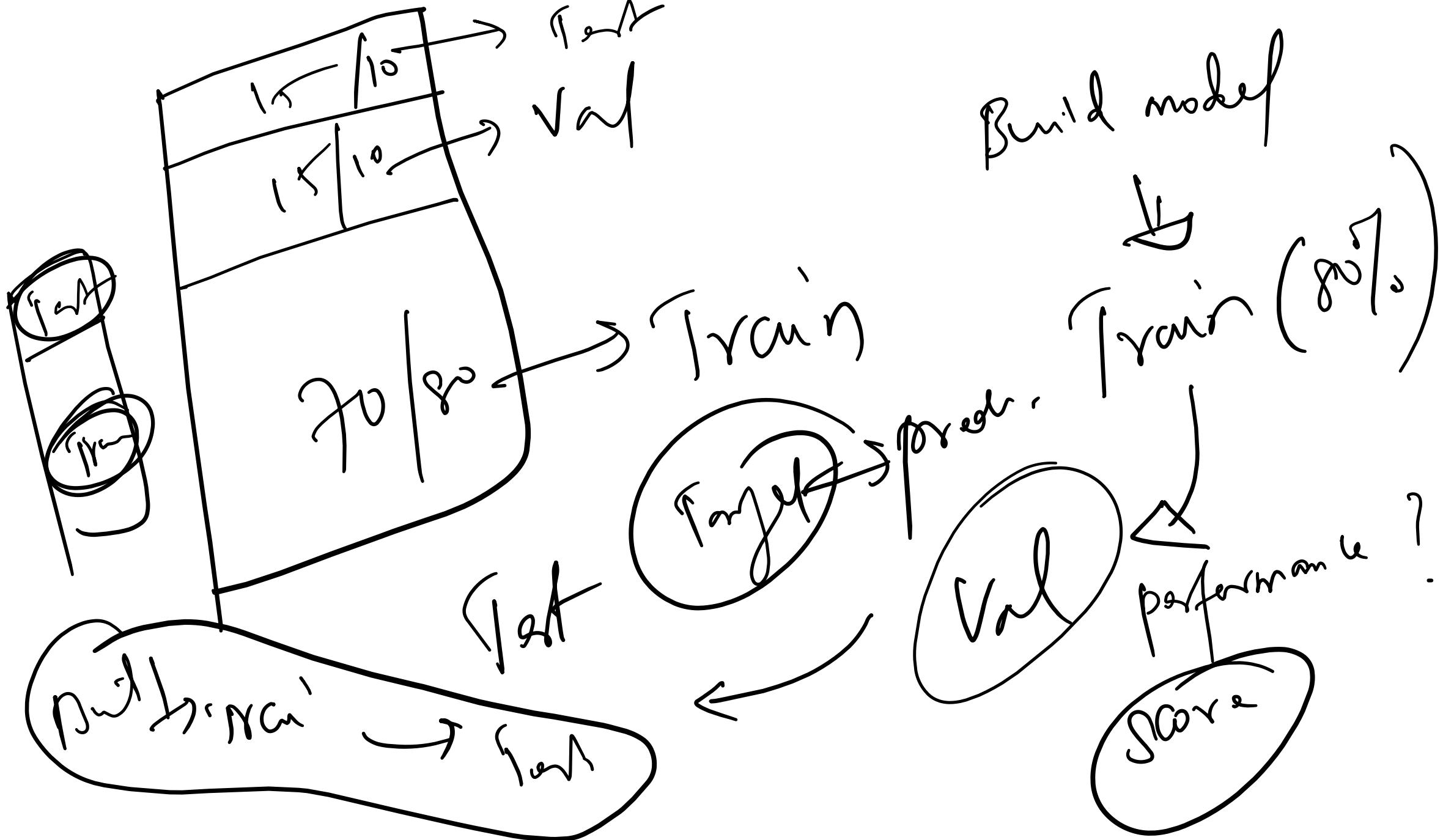
$\rightarrow 10 \rightarrow$ Features



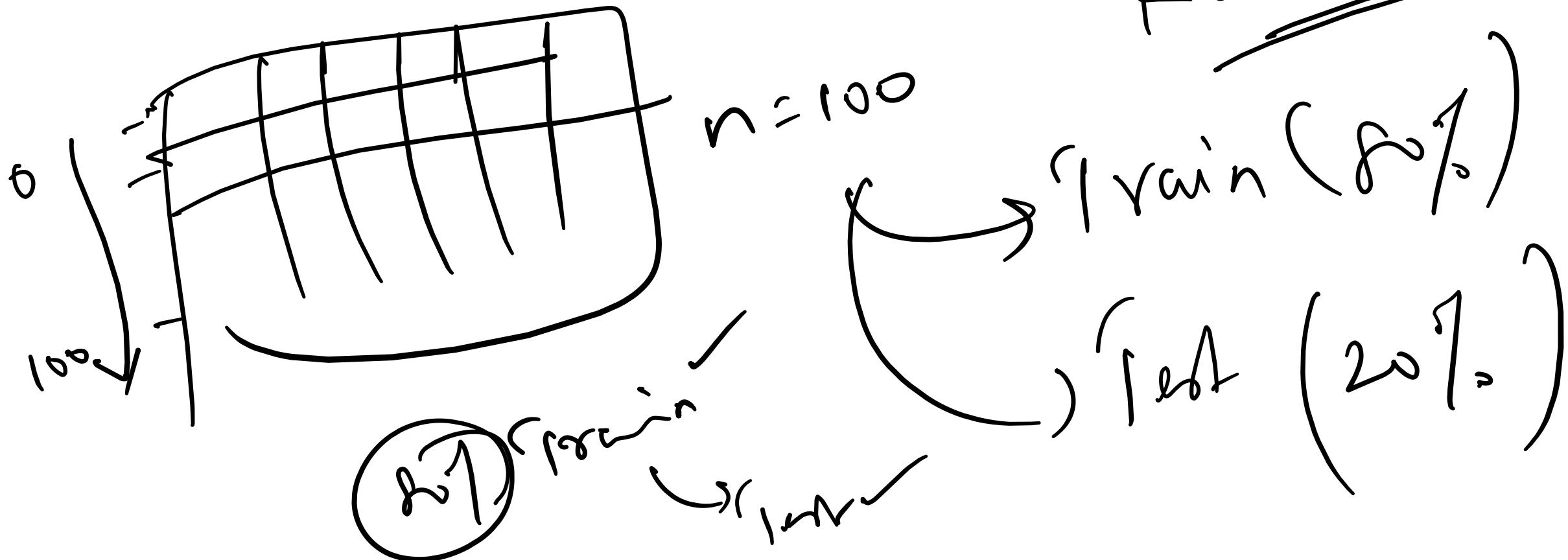
10

Bcwd





While we split the data, it should be,
shuffled



Deployment:

↳ GCP

↳ Heroku