Beginner class

what you want do?

T - Task → classification, Regression] supervised Learning

→ Clustering - unsupervised

- pattern recognization
- feature Selection
- dimensionlity reduction

how to evaluate?

performance :-

- Measuring Algo's Accuracy
- scikit-learn rich set of performance Measure

Experience :-

→ From Historical Data

Supervised

Petal width	Petal height	color	flower
-	-	-	-
	-	-	-
-	-	-	-

features

Instance / Record / sample

Features Selection

- Select More weight feature.
- In ML - you have do features selection
- In DL - NN do it by itself.

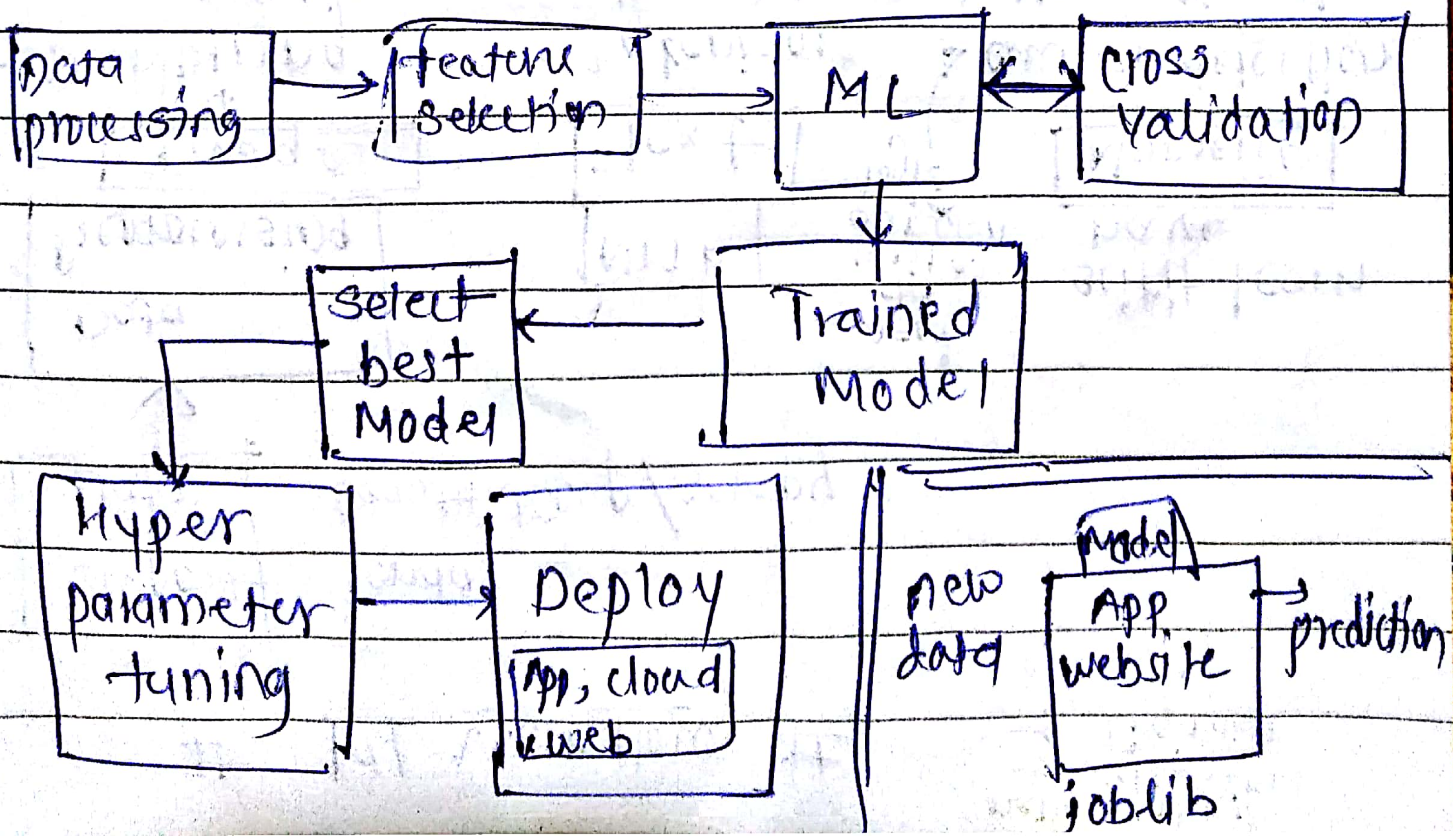
Redundant features:

- features ~~do~~ which add more meaning for prediction

Classification	Regression	Clustering
SVM	Linear Reg	
	GLM	
DA		
LDA, QDA	SVR.	
Naive	Ensemble Methods	
Logistic		

Bike sharing ~~dim~~ demand - Kaggle

Import Data



ML Workflow

Data Analysis
— Scientist
— Engineer

Import Data
Pandas
BeautifulSoup / scrapy

Data Preprocessing

Image
- Debluring
- Edge delinition
- Edge Enhancement
- histogram Equalization
- object detection
- Denoising
- Segmentation
- CLAHE (Adaptive histogram)

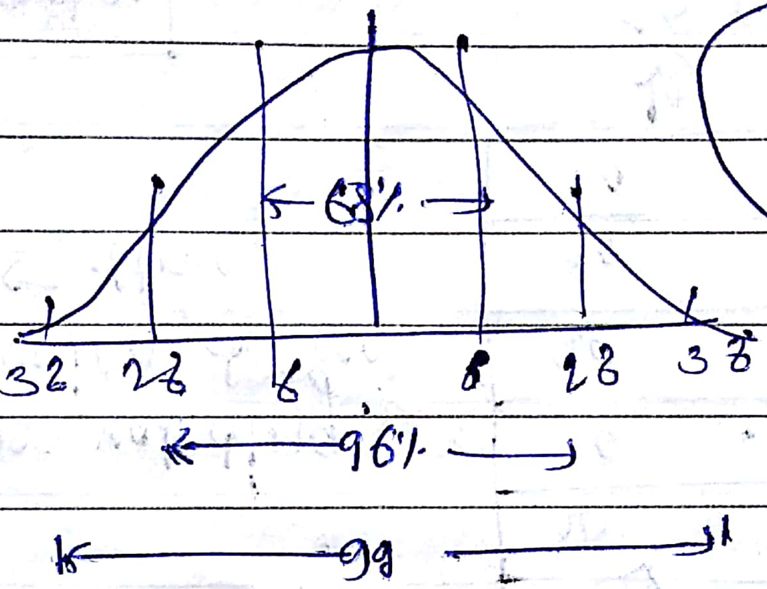
NLTK
Text
- Tokenizers
- stop words
- TF - IDF
- t-SNE (dimensionality Reduction)
- POS Tagging
- stemming

need to write our own logic

scikit-learn have
Numeric
- outlier detection
- Normalization
- standardization
- Binarization
- One hot Encoding
- Imputer

Thumb rule

- If data points beyond the 2 or 3 std. ~~variance~~ deviation from Mean it is Outliers



Gaussian distribution

Data preprocessing

- feature selection
- Extraction
- Engineering
- Dimensionality Reduction

(PCA, t-SNE)

When manipulate Existing ~~on~~ features to create new features

Extracting features from Data.

Feature selection → ML Algo ↔ Cross Validation

Train: Training, validation

Test: testing

Train 100: 20 → holdback → validation set; 20, 20, 20, 20 → Training

⇓

20 → Training

20 → holdback

20, 20, 20 → train

⇓ so on

cross validing traning dataset 5 time

Learning through Semester → Train

UT - Virq-Mock → cross validation

final Examination → Test

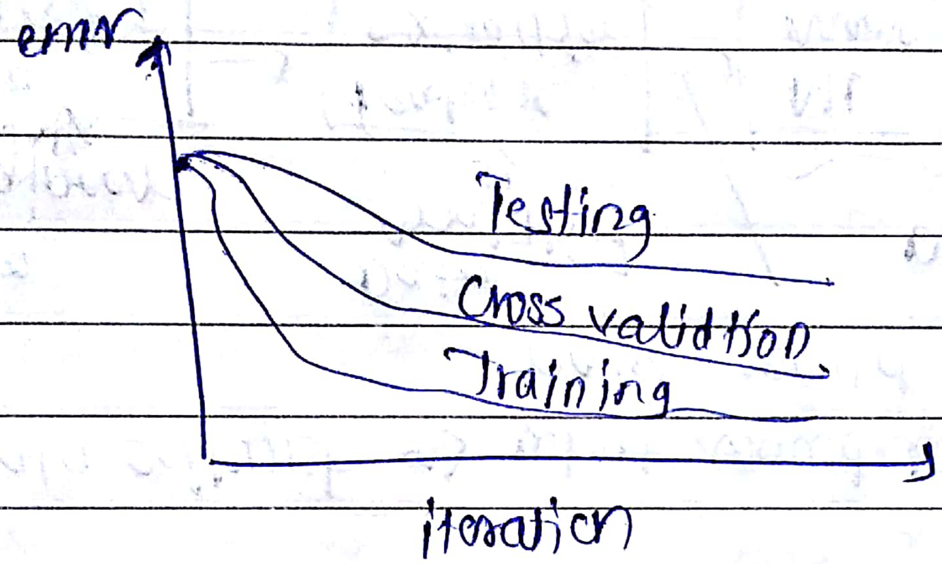
Cross validtion

→ K - ~~flod~~ ~~vo~~ cross validtion
fold

→

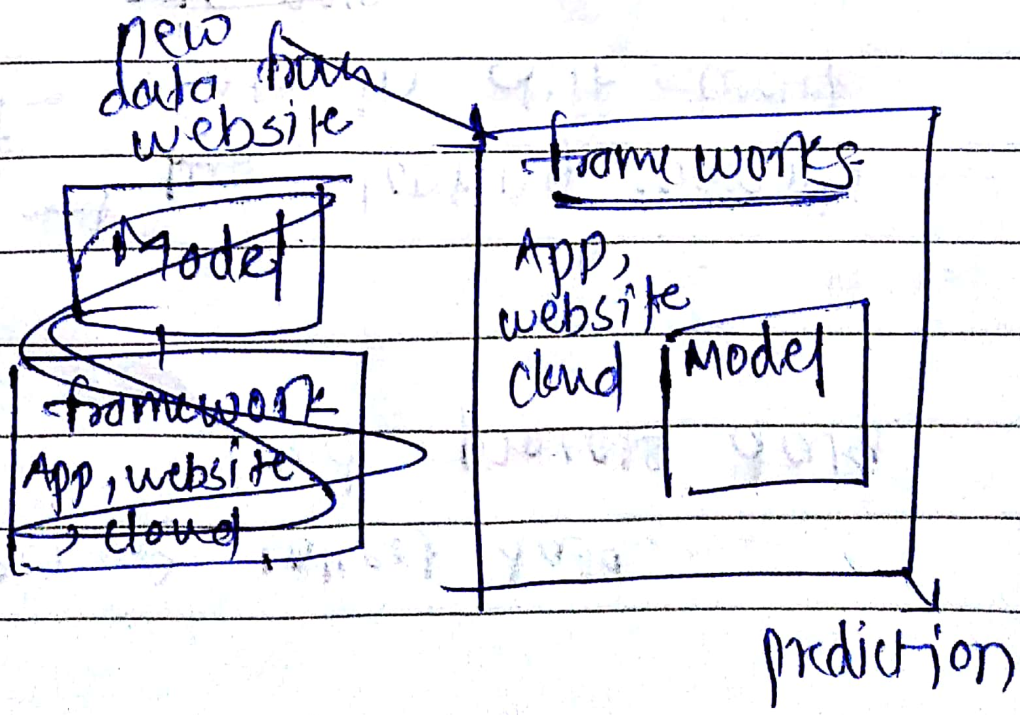
→ shuffle split cross validtion

- LOOCV (leave one out cross validation



Hyper Parameter Tuning :-

Deploy

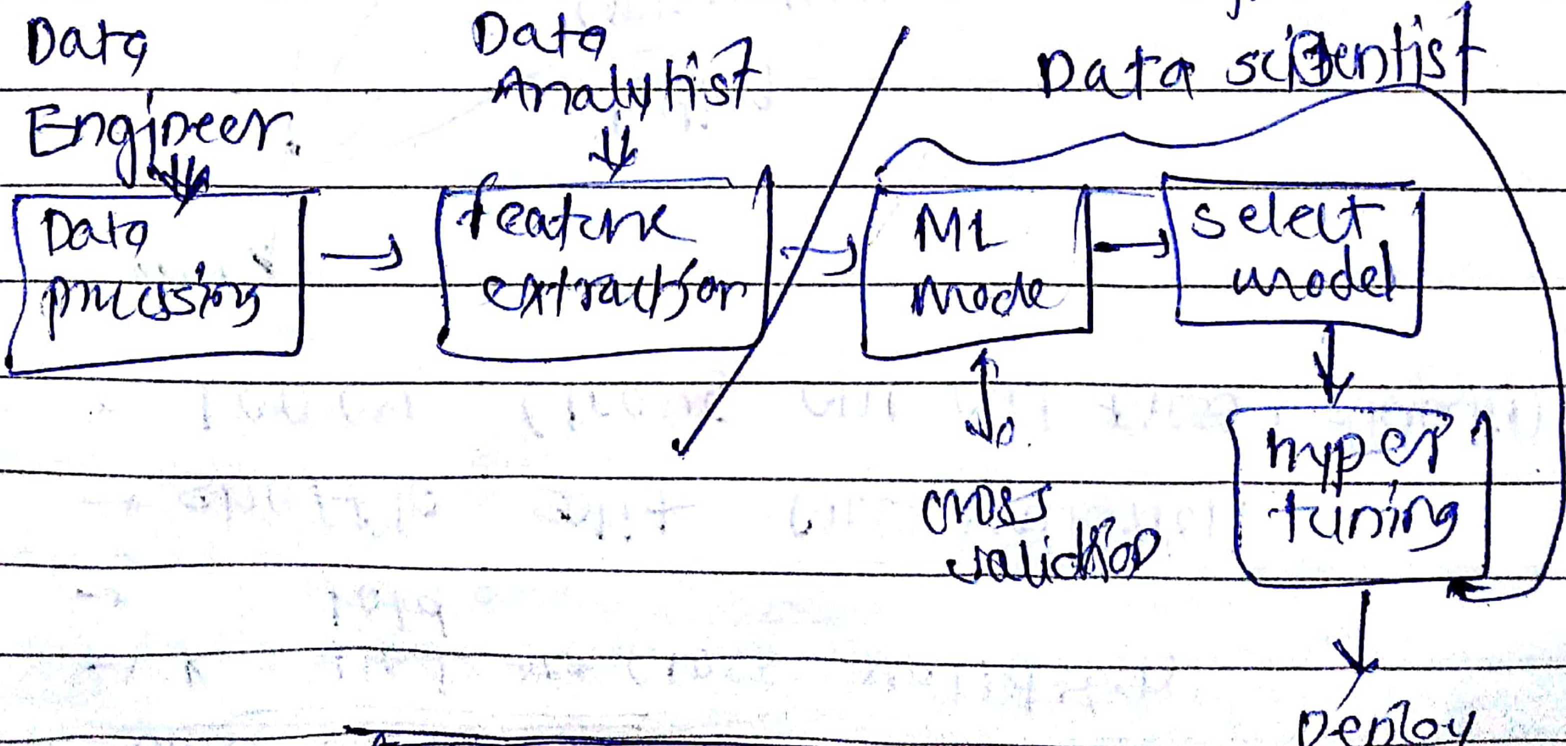


Data Engineer ⇒ collect data
and process data.

Data ~~Analysti~~ Analytist ⇒ ~~How~~ looking towards
data in diff. way
~~in new~~

- just Analysing data.
- what happened
- how happend

Data scitenist ⇒ what would happen
how would it happen



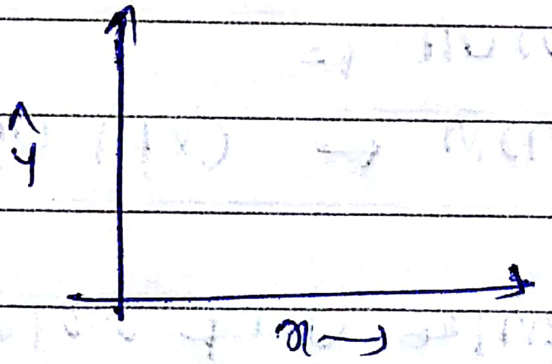
full stack Data ~~scitn~~ scientist

Model, Algo, Data

$y = f(x)$

$$\boxed{\text{Model} = \text{Algo}(\text{Data})}$$

Manul Compution prediction



in NN
- no. of layer
- no. of neuron
- weight
- Bias
- final Matrix

$\hat{y} = w_0 + w_1 x$

learnt model $(\hat{w}_0, \hat{w}_1)$

learnt Model

$\hat{w}_0$ = Best value for w_0
= optimal point

x	y
1	4.8
3	11.3
5	17.2
7	?

objective f^n

$\Downarrow$

$f(x) = 2x^2 - 5x + 3$

Close form solution - Can solve by manually

close form solution NOT optimization.

- min $(f(x))$ → well defined problem
  ⇒ unconstraint problem.

- min $f(x)$ s.t. $1 \leq x \leq 10$ & $2x^2 \geq 5$

optimization problem
most of time ~~a~~ have
Constraint

Ques. $\boxed{\underset{a}{\text{argmin}}\ f(a)}$?

⇒ find argument a for which f(a) is minium

Optimization problem required

objective function

constriant (optional)

Problem formulation of LR :-

$\hat{y}$ = predicted

y = actual value.

(1) Error or loss f^n $= \sum_{i=1}^{N} (\hat{y}_i - y_i)$

N = total no. of instances.

$\boxed{\text{But its wrong objective } f^n}$

- giving false impression of minization of error

Now . P.T.O

(2) $$L = \sum_{i=1}^{N} |\hat{y}_i - y_i|$$

necessory cond^n — sufficient const

$$\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial^2 f}{\partial x^2} = +ve \quad \text{minima}$$

$$\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial^2 f}{\partial x^2} = -ve \quad \text{maxima.}$$

But it's not differeniable

absolute function not differentiable

∴ ~~not~~ it not optimise objective fn.

(3) $$L = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

RSS - Residual ~~Square~~ sum of square
SSR - sum of square of residual.

f^n - parabolic

still its' not better solution

(iv) if increasing Monotonously learning rate is not increases

So need to normalization objective function.

$$\boxed{L = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}$$

→ its Most logical objective function for optimization

$$\hat{w}_1 = \frac{\overline{(xy)} - \bar{x}\,\bar{y}}{\overline{(x^2)} - (\bar{x})^2}$$

$$= \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}$$

x^2	x	y	xy
1	1	4.8	1×4.8 ( )
9	3	11.3	3×11.3 ( )
25	5	17.2	5×17.2 ( )

$$\overline{xy} = \frac{\Sigma(\ )}{3}$$

$$\boxed{\hat{y} = \hat{w}_0 + \hat{w}_1 x}$$

$$\boxed{\hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}}$$

$L = 3.1$

$\bar{y} = 1.8$

need to check.

(OLS)
ordinary least sequare

Analytical solution

Numerical solution
Iterative Method
Gradient desent

$$\frac{\partial L}{\partial w_0} \text{ slop} = \text{gradient of } L \text{ w.r.t } w$$

Total update ① → Direction of update from gradient.

② → step size

- update rule of gradient desent

$$w_0^{k+1} = w_0^{k} - \alpha \left(\frac{\partial L}{\partial w_0} \right) \times 1 \quad \rightarrow \Sigma(\hat{y}-y)^2$$

$$w_1^{k+1} = w_1^{k} - \alpha \left(\frac{\partial L}{\partial w_1} \right) \times x_i$$

⇒ This is batch gradient desent / Vanilla gradient desent

improvization :-

① ~~stoca~~ ~~stheca~~ stochastic G.D.

② Mini-Batch G.D.