ML Algo

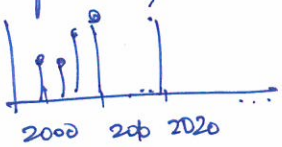
Supervised ML
- Labled Data
- Training on Data
- Eg: Regression.

Unsupervised ML
- No labeled data.
- No training phase
- find the relationship b/w features in the given data.

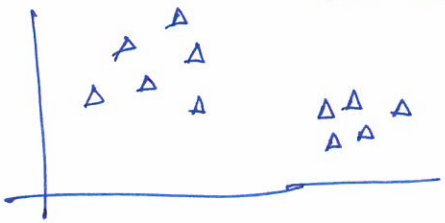
Eg: Clustering.

3-Tasks

1. Classification → fruit → apple, orange, grapes
2. Prediction → predicting price of House



3. Clustering



ML Algo

Linear Algo

$y \rightarrow$ linear combination of i/p x_i

Linear Regression.

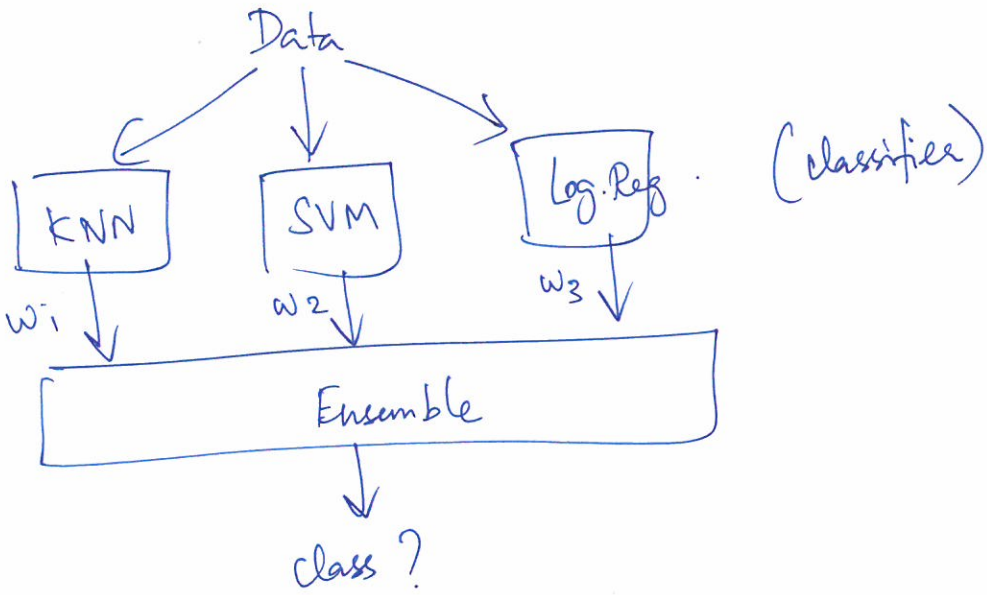
Non-linear Algo

X

K-NN Algo.

Ensemble Algo:

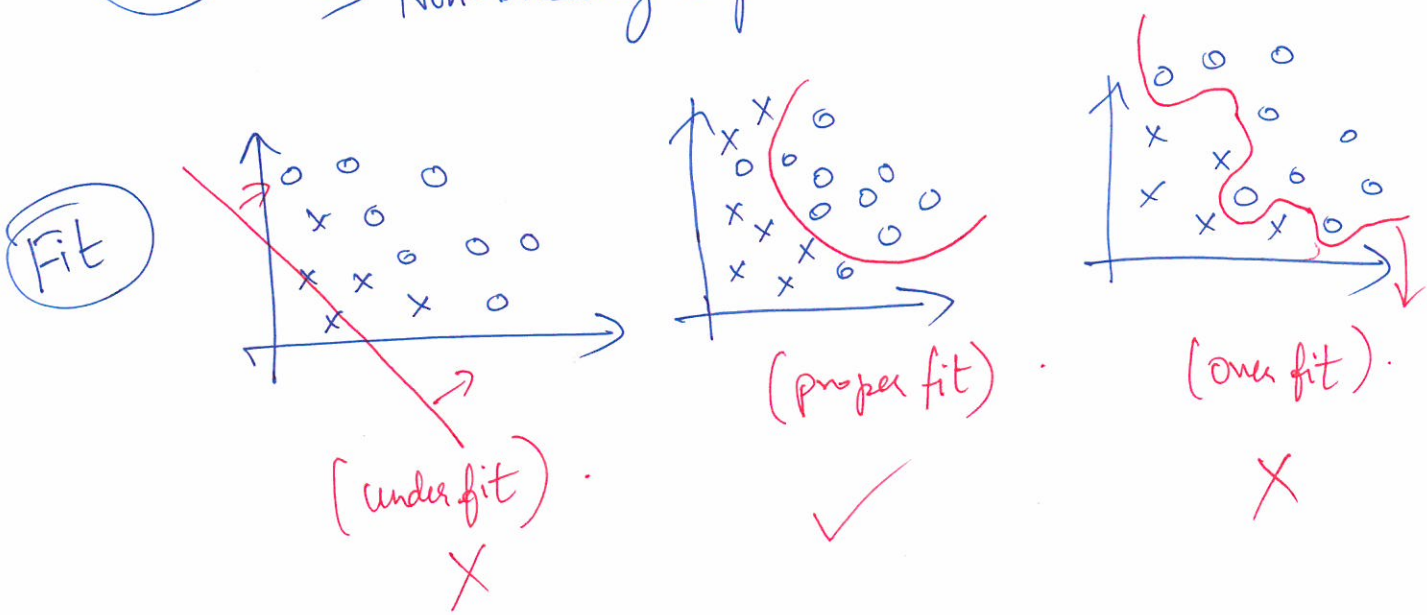
Collection

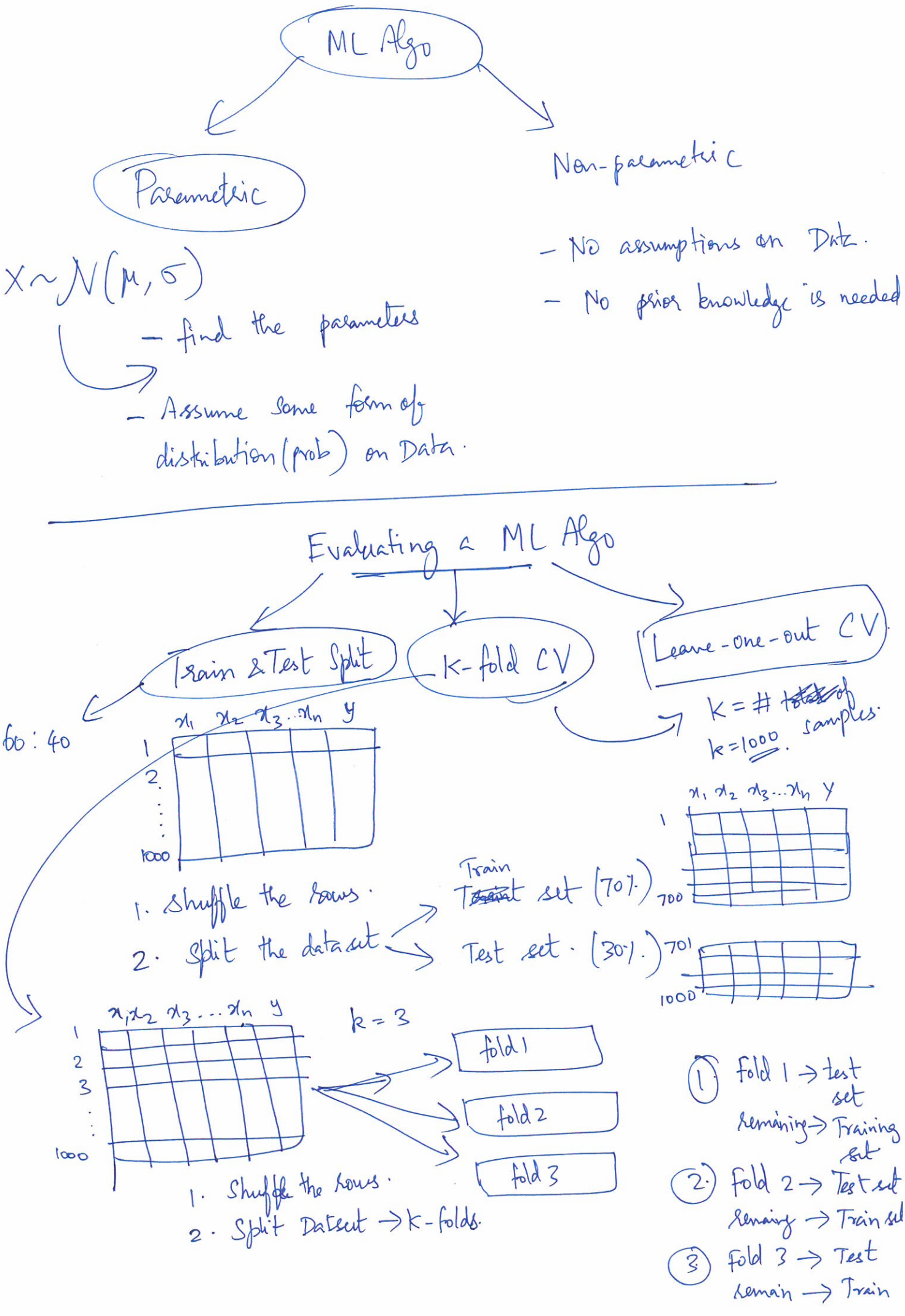


Linearly Separable

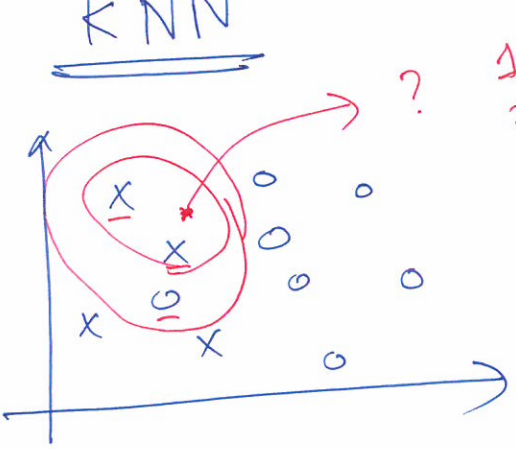
Data

Non-linearly separable

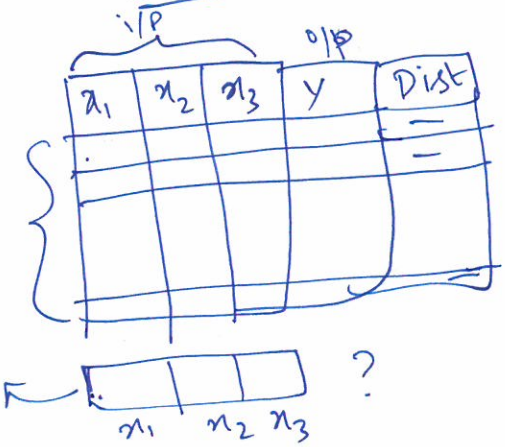


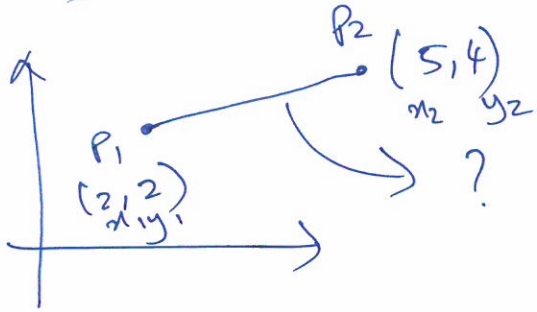

ML Algo
Parametric
Non-parametric
$X \sim N(\mu, \sigma)$
- find the parameters
- Assume some form of distribution (prob) on Data.
- No assumptions on Data.
- No prior knowledge is needed
Evaluating a ML Algo
Train & Test Split
K-fold CV
Leave-one-out CV
60 : 40
x_1 x_2 x_3 ... x_n y
1
2
1000
1. Shuffle the rows.
2. Split the dataset
Train set (70%)
Test set. (30%)
1
700
701
1000
K = # samples.
k = 1000.
x_1, x_2 x_3 ... x_n y
1
2
3
1000
k = 3
fold 1
fold 2
fold 3
1. Shuffle the rows.
2. Split Dataset → k-folds.
(1) Fold 1 → test set
Remaining → Training set
(2) Fold 2 → Test set
Remaining → Train set
(3) Fold 3 → Test
Remain → Train

KNN

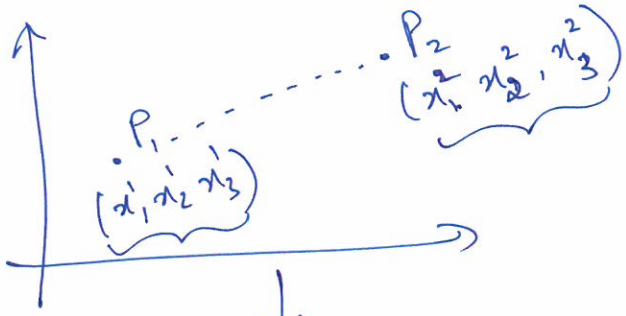


Distance:





$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$



$$\text{Euclidean Dist} = \sqrt{\sum_{i=1}^{n}(x_i' - x_i^2)^2}$$

$P_1(x_1', x_2', \ldots x_n')$

$P_2(x_1^2, x_2^2, \ldots x_n^2)$

1. Compute E.D b/w Query pt & all the pt in Data Set
2. Sort Dataset based on Dist.
3. Find the classes for k-neighbors.
4. Assign output (class) for query pt → Counts of Neighbors.

$$k = \frac{\sqrt{\#\text{Data point}}}{2}$$