Supervised => Labelled Data => Target

Historical Data => Train a model => train-test-split

$h_\theta(x) = \theta^T X = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$

House Price Prediction

Remove Overfitting => Bias

Labelled Data => Accuracy.

Train the model

Features | Target value => MEDV => Price

Predict <= Model <= Error metrics based on dependent data

Gradient Descent

Unsupervised :) No labelled Data :)

↓

Group Data Based on Similarity.

↓

$(x_1 - x_2)^2$, Correlation coefficient

↓

Clustering

↙ ↘

Exclusive

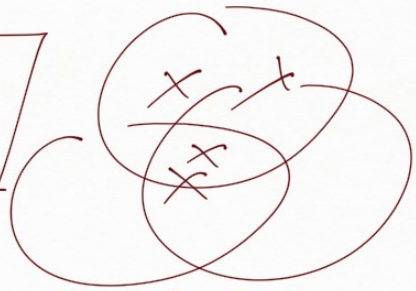
K-Means

Hierarchical.

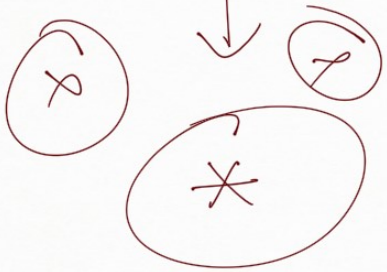
Overlapping

Fuzzy

C-Means

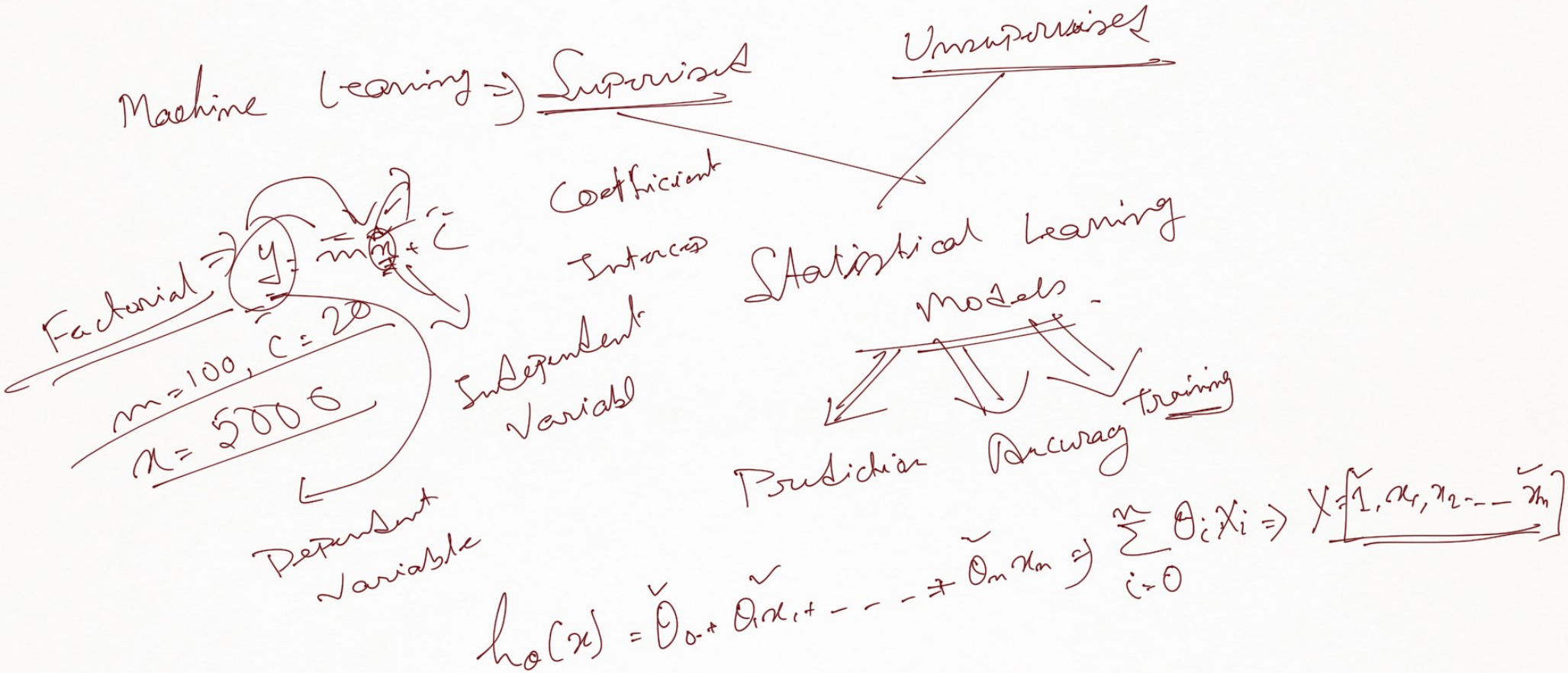


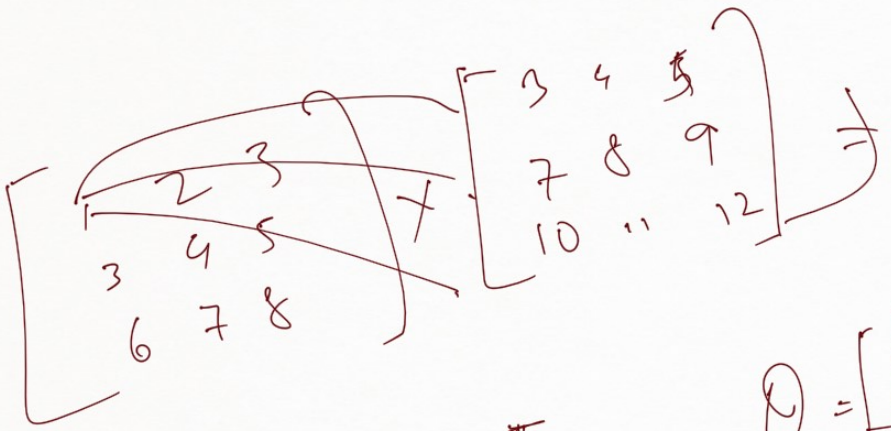
Clustering → Exclusive Clustering



Overlapping Clustering

Fuzzy Clustering


Machine Learning
Supervised
Unsupervised
Coefficient
Intercept
Factorial
y = mx + c
m = 100, c = 20
x = 5000
Independent Variable
Dependent Variable
Statistical Learning
models
Prediction
Accuracy
training
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_m x_m \Rightarrow \sum_{i=0} \theta_i x_i \Rightarrow X = [1, x_1, x_2 \cdots x_m]$



$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \times \begin{bmatrix} 3 & 4 & 5 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} \Rightarrow$$

$3\times2 \quad 2\times3 \quad \overset{3\times3}{\underset{3\times3}{}}$

$$h_\theta(x) = \theta^T X \Rightarrow \quad \theta = \underbrace{[\theta_0 \ \theta_1 \ \cdots \ \theta_n]}_{1\times n} \quad n\times1 \ \ 1\times n$$

$$X = [1 \ x_1 \ x_2 \ \cdots \ x_n]_{1\times n}$$

Transpose $\Rightarrow$ Exchange Rows & Columns

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

$x \Rightarrow$ Features / Columns of your dataset

House Price ⇒ Area, No×, Zip ⇒

Cost Function ⇒ $\frac{\partial J(\theta)}{\partial \theta}$ = ?

MSE ⇒ $\sum_{i=1}^{n} (h_\theta(x_i) - y_i)^2$

$\theta_0 \; \theta_1 \; \theta_2 \; \ldots \; \theta_n$

$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$ (α: learning Rate, $\frac{\partial J(\theta)}{\partial \theta}$: slope of)

$\theta_0 = \theta_0 - \alpha \frac{\partial (J\theta)}{\partial \theta_0}$

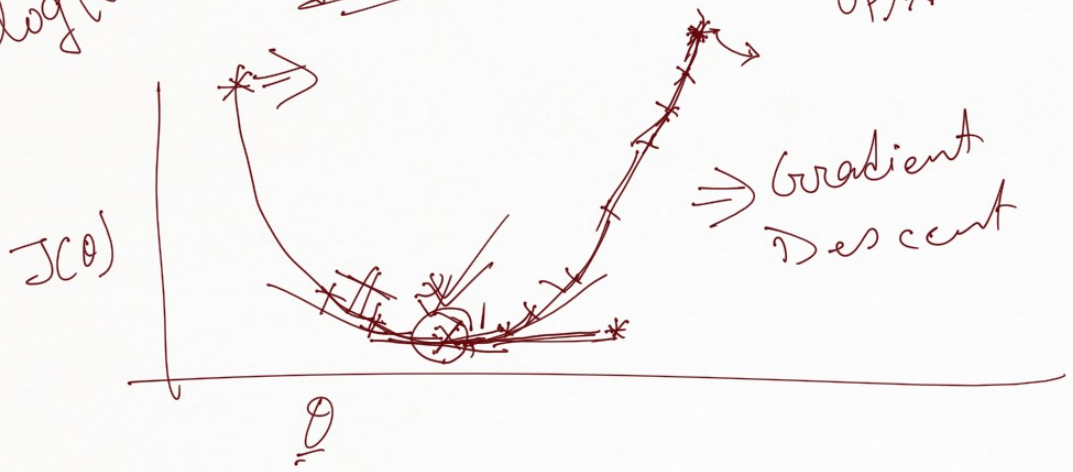
$\theta_1 = \theta_1 - \alpha \frac{\partial J(\theta)}{\partial \theta_1}$

$-y \log(\hat{y}) - (1-y) \log(1-\hat{y})$

Optimum θ ⇒ Finding min

$J(\theta)$ ⇓ Derivative

$\frac{\partial J(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}$ ( )



⇒ Gradient Descent

Learning Techniques => Association Rules. $\frac{e^{\log(x)}}{1 - e^{\log(x)}}$

Data mining

$\frac{1 \Rightarrow A \Rightarrow B}{2 \Rightarrow B}$

Gini Index
Entropy

Bread => Butter

China Drops => US will Bomb on Taiwan so in A => B
water

Market Basket Analysis.

Apriori Algorithm

Rice => Pulses

Fruit => Vegetable

milk => Chocolate

$A \Rightarrow B$

Analysis Technique => Generate Associations / relations =>

Retail Store =>

Order id	Products
1	[n-items]
2	

Market Basket Analysis

{
Rice => Pasta
Rice => Oil
Bread => Milk
{Bread, Butter} => Milk
}

Apriori Algorithm → Finding out association

Support / Confidence / Lift

Frequency →

Frequency of a item / Total transaction

$$S(\text{Item}) = \frac{F(\text{Item})}{N} = \frac{7}{10}$$

Milk ⇒ 7

Egg ⇒ 3

$\{Milk, Egg\} \Rightarrow 3$ → A, B

$\{Milk, Egg\} = 3$

$S\{Milk, Egg\} = \frac{3}{10}$

A ⇒ B ⇒

Antecedent / consequent

Milk = 7 } 10

Egg = 3 }

$\{Milk, Egg\} = 3$

$$C(A,B) = \frac{S(A,B)}{S(A)}$$

$$= \frac{3}{10} / \frac{7}{10} = \frac{3}{7}$$

$$= \frac{3}{10} \times \frac{10}{7} = \frac{3}{7}$$

$$\text{Lift } d(A,B) = \frac{C(A,B)}{S(B)}$$

$$= \frac{S(A,B)}{S(A) \times S(B)}$$

$$= \frac{3/7}{3/10} = \frac{3}{7} \times \frac{10}{3} = \frac{10}{7}$$

$$\simeq 1.4$$

$$= \frac{3}{10} \times \frac{10}{7} = \frac{3}{7} \times \frac{10}{7} = \frac{10}{7} > 1$$

Lift > 1

$$\frac{1}{10} \times \frac{100}{21} = \frac{10}{21} < 1$$

K => Number of items

Apriori Steps =>

1. K=1. Calculate Support. Reject all item whose support < threshold

2. K=2, Calculate Support, Confidence, Lift. =>

3. Repeat Steps 2 by incrementing K till we have no more itemset remaining. =>

Threshold for support, Confidence, Lift. =>

0.5 0.5 2

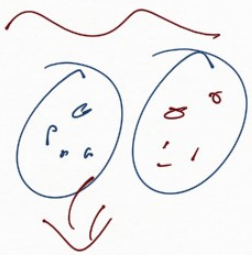
{Milk, Bread, Egg}

Milk, Bread

Bread, Milk

Milk, Egg

K-Means Clustering :)

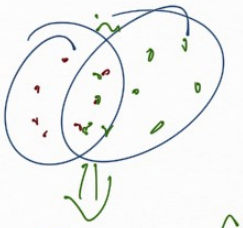


Exclusive Clustering

↓↓

Customer Segmentation

Recommendation Engine



↓↓

Fuzzy clustering / Overlapping Clustering.

Find out similarity of items.

K-means ⇒ Hierarchical clustering ⇒

$\{1, 2, 3 \ldots n\} \Rightarrow N$

$\Downarrow$

K: different clusters ⇒ $\{c_1, c_2, c_3 \ldots c_K\}$

$\sum_{i=1}^{K} |C_i| = N$ where $|C_i| \Rightarrow$ Number of elements in cluster C_i

$\sum_{i=1}^{K} Var(C_i) \Rightarrow Var(C_i) \Rightarrow$ Inter cluster variance

$\Rightarrow W(C_i) \Rightarrow$ Group similar items

minimize $\Rightarrow \sum_{i=1}^{k} w(C_i)$ $\Rightarrow$ Euclidean Distance

$w(C_i) \Rightarrow$ Inter cluster Variance

Cost function $\Rightarrow \sum_{K=1}^{K} \sum_{i,i' \in C_K} (x_{iK} - x_{i'K})^2$

$\Downarrow$

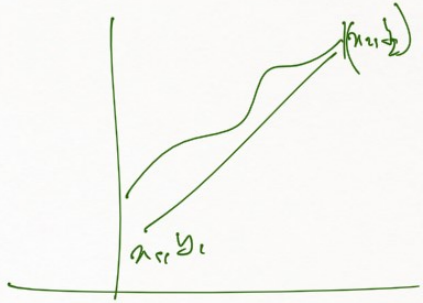
$\{(x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2\} + \cdots$

$\{(x_4 - x_5)^2 + (x_5 - x_6)^2\} -$

(x_1, y_1) (x_2, y_2)

$E.D. = \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2}$

$= \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$

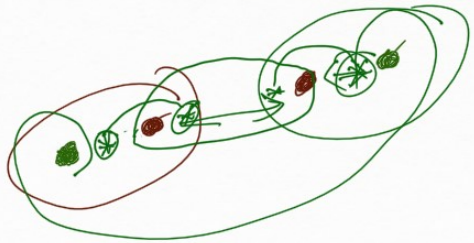


Similar Objects =)

K-Means algo =) Random Selection of Clusters =)

- =) Intra-Clusters Variance
- =) Shift points betw. Cluster Centroids

mem of All points in a Cluster =)



$$\sum_{k=1}^{K} \sum_{i,j \in C_k} \left(x_{ik} - x_{jk}\right)^2$$

where $K \Rightarrow$ Number of clusters $\Rightarrow 2, 3, 4 = \_\_$

$i, j \in C_k =$ All points in cluster C_k

$$\sum_{k=2}^{K} \sum_{i,j \in C_1} \left\{ \left(x_{11} - x_{21}\right)^2 + \cdots \right\}_{i,j \in C_k} \Big\}_{|C_k| = N}$$

$i, j \in \{1, N\}$

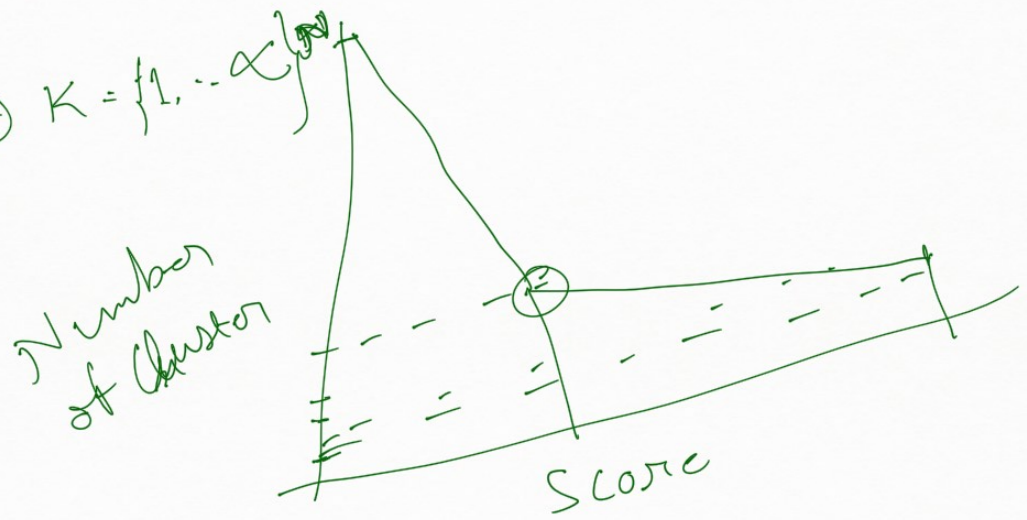
Steps of K-Means

1) Guess 3-Random centroids
2) Calculate variance of each point from th centroids.
3) Assign point to the closest centroid
4) Recalculate centroid.
5) Repeat 2-4 till no shifting occurs.

Elbow Method =)

Total variance in all clusters => For each K =)

$K \in R \Rightarrow K = \{1, \ldots \infty\}$



Silhoutte Score = $\frac{a - b}{\max(a, b)}$

a :) distance of a point from nearest cluster.

b :) Intra cluster diff.

$-1 \leq$ Silhoutte Score ≤ 1

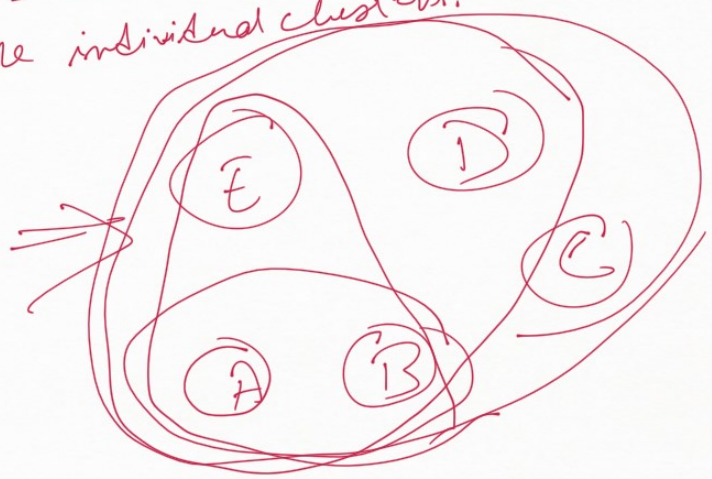
+1 :) clusters are very far

0 :) clusters are very close

-1 :) Decision boundary is not clear

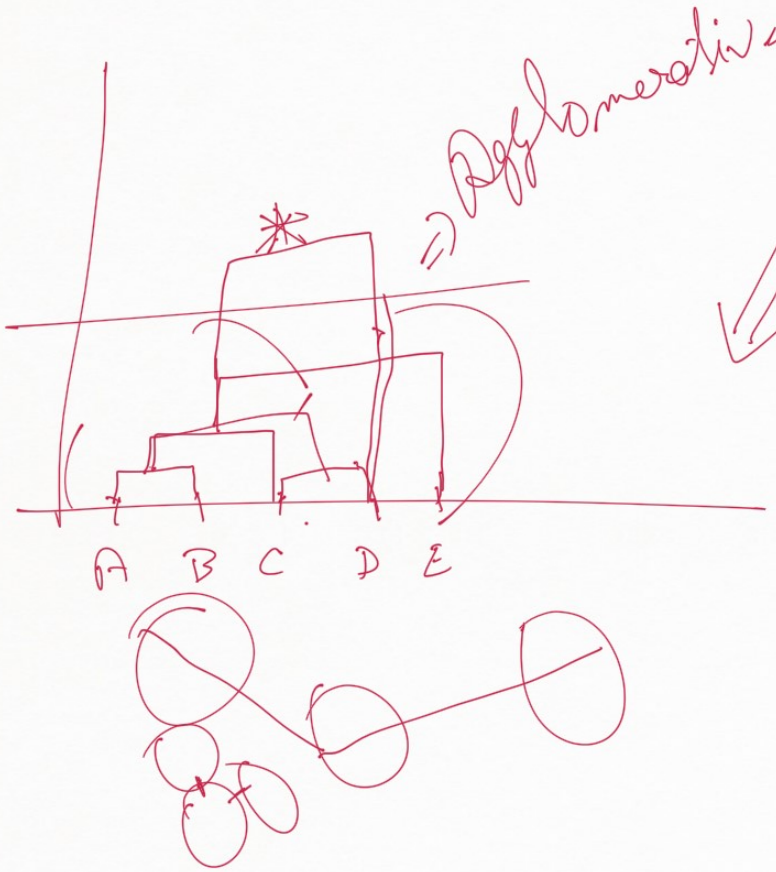
Hierarchical =>

cluster => Entire space is 1 cluster.
Individual points are individual clusters.



Bottom-Up => Agglomerative

Top-Down => Divisive.



⇒ Agglomerative.

1) Which points to combine for one cluster
2) Which clusters to fuse together

Linkage method

1) Single ⇒ Considers minm. distance betn. clusters
2) Complete ⇒ Considers maxm. "
3) Group Average
4) Centroid
5) Ward