

unit-3

Unsupervised learning → It refers to the use of artificial intelligence algorithm to identify patterns in data sets containing data points that are neither classified or labeled.

Clustering → It is the grouping of objects or data points that are similar to each other and dissimilar to objects in other clusters.

Categories of clustering

- 1) Exclusive clustering
- 2) Overlapping clustering eg K-mean
- 3) hierarchical clustering
- 4) probabilistic clustering

Association → when you can find dependencies of one data item to another & map them such that they can help you

eg Bread + Butter, jam + milk

- * K-mean \rightarrow it iterates over again and again unless and until all points within clusters stop changing.

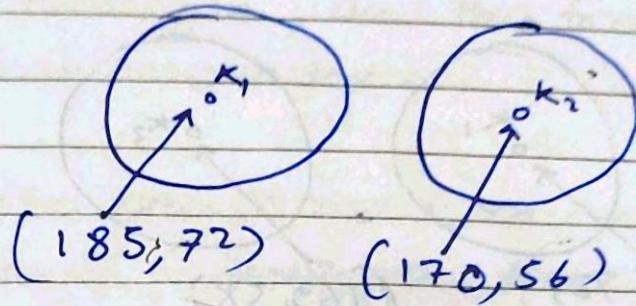
K mean Clustering

Steps:-

1. Initialize K random Centroids.
 - \rightarrow You could pick K random data points and make those your starting point.
 - \rightarrow otherwise you pick K random values for each variable.
2. For every data point, look at which Centroid is nearest to it.
 - \rightarrow using some sort of measurement like Euclidean or cosine distance.
3. Assign the data point to the nearest Centroid.
4. For every Centroid move the Centroid to the average of the points assigned to that Centroid.
5. Repeat the last three steps until the centroid assignment no longer changes.
 - \rightarrow The algorithm is said to have ~~66~~ ⁶⁶ ~~66~~ steps

Once there are no more changes.

	Height	Weight
①	185	72
②	170	56
③	168	60
④	179	68
⑤	182	72
⑥	188	77
⑦	180	71
⑧	180	70
⑨	185	84
⑩	180	88
⑪	180	87
⑫	173	76



Euclidean Distance : - $\sqrt{(x_o - x_c)^2 + (y_o - y_c)^2}$

ED for 3 $\rightarrow k_1 = \sqrt{(168 - 185)^2 + (60 - 72)^2}$
 $= 20.80$

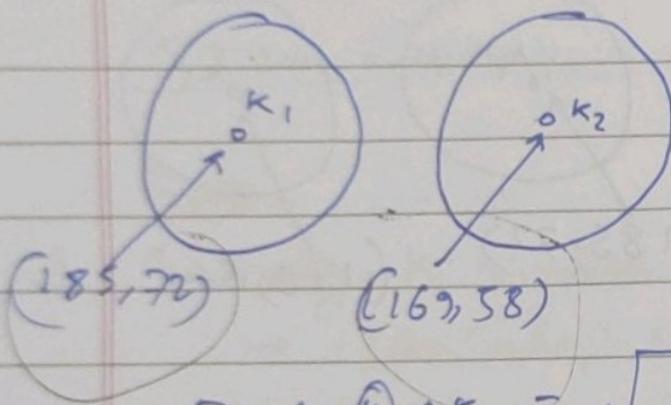
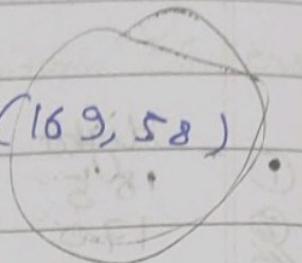
$\rightarrow k_2 = \sqrt{(68 - 170)^2 + (60 - 56)^2}$
 $= 4.48$.

$$4.48 < 20.80$$

so 3 go in k_2 cluster

New centroid calculation :-

$$\text{for } K_2 = \left(\frac{170+168}{2}, \frac{60+56}{2} \right) = (169, 58)$$



$$\text{ED from } K_1 = \sqrt{(179-185)^2 + (68-72)^2} \\ = 6.32$$

$$K_2 = \sqrt{(179-169)^2 + (68-58)^2} \\ = 14.14$$

8

$$6.32 < 14.14$$

so we go in K_1 cluster

$$K_1 = \{1, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

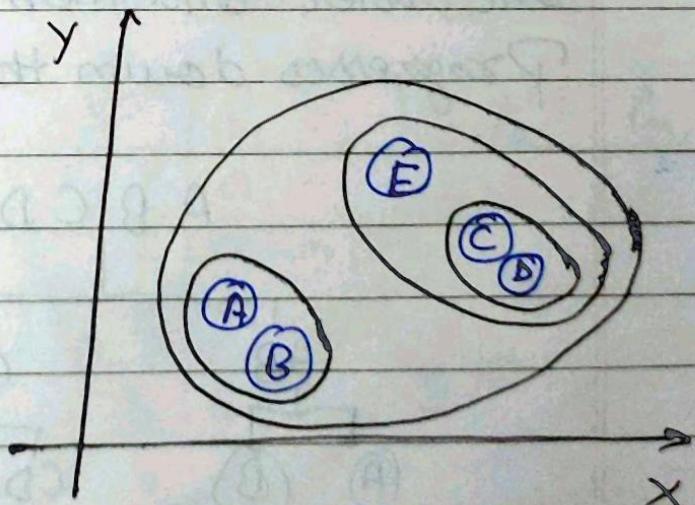
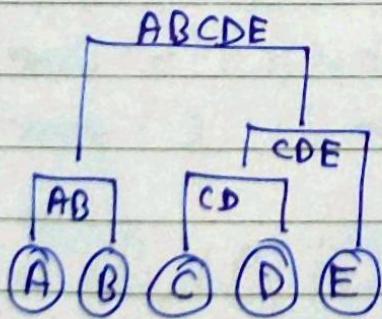
$$K_2 = \{2, 3\}$$

Hierarchical Clustering

The hierarchy of clusters is developed in the form of a tree in this technique, and this tree-shaped structure is known as dendrogram.

Separating data into groups based on some measures of similarity, finding a technique to quantify how they're alike and different, and limiting down the data in what hierarchical clustering is all about.

- Agglomerative
- Divisive



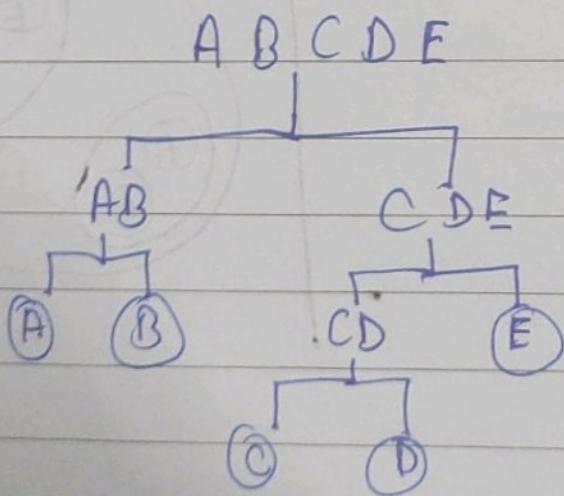
Agglomerative

- Agglomerative Clustering :

Agglomerative Clustering is bottom-up strategy in which each data point is originally a cluster of its own, and as one travels up the hierarchy, more pairs of clusters are combined. In it two nearest clusters are taken and joined to form one single cluster.

Divisive clustering

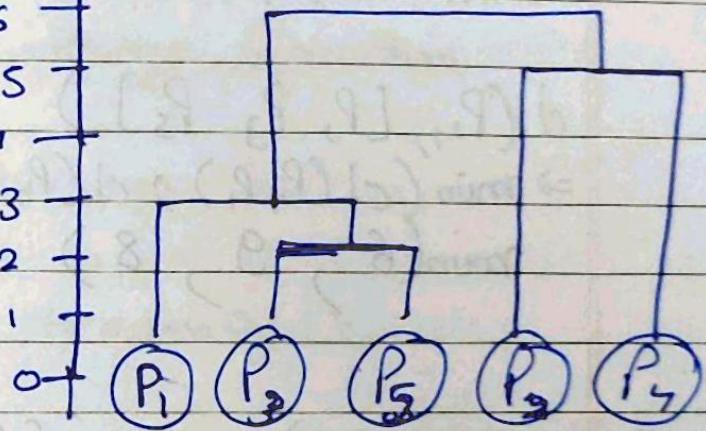
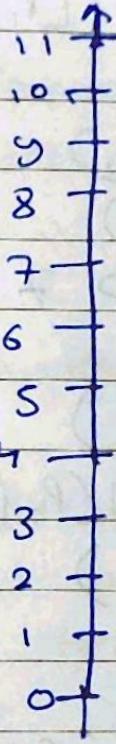
It is a top-down clustering strategy in which all points in the datasets are initially assigned to one cluster and then divided iteratively as one progresses down the hierarchy.



* Agglomerative clustering (Single Linkage)

①

	P_1	P_2	P_3	P_4	P_5
P_1	0				
P_2	9	0			
P_3	3	7	0		
P_4	6	5	9	0	
P_5	11	10	②	8	0



* Find minimum element

in distance matrix. Now combine P_3 & P_5

$$\Rightarrow d(P_1, [P_3, P_5])$$

$$\min(d(P_1, P_3), d(P_1, P_5))$$

$$\min(3, 11) \Rightarrow 3$$

②

	P_1	P_2	$[P_3, P_5]$	P_4
P_1	0			
P_2	9	0		
$[P_3, P_5]$	3	7	0	
P_4	6	5	8	0

$$\Rightarrow d(P_2, [P_3, P_5])$$

$$\min(d(P_2, P_3), d(P_2, P_5))$$

$$\min(7, 10) \Rightarrow 7$$

$$\Rightarrow d(P_4, [P_3, P_5])$$

$$\min(d(P_4, P_3), d(P_4, P_5))$$

$$\Rightarrow \min(9, 8) = 8$$

	(3)	$[P_1, P_2, P_5]$	P_2	P_4
$d(P_2, [P_1, P_3, P_5])$	$[P_1, P_3, P_5]$	0		
$\Rightarrow \min(d(P_2, P_1), d(P_2, P_3), d(P_2, P_5))$	P_2	7	0	
$\Rightarrow \min(9, 7, 10) \Rightarrow 7$	P_4	6	5	0

$$\begin{aligned}
 & d(P_4, [P_1, P_3, P_5]) \\
 & \Rightarrow \min(d(P_4, P_1), d(P_4, P_3), d(P_4, P_5)) \\
 & \min(6, 9, 8) \Rightarrow 6
 \end{aligned}$$

$$d([P_1, P_3, P_5], [P_2, P_4])$$

		$[P_1, P_3, P_5]$	$[P_2, P_4]$	
$\Rightarrow \min(d(P_2, P_1), d(P_2, P_3), d(P_2, P_5), d(P_4, P_1), d(P_4, P_3), d(P_4, P_5))$	$[P_1, P_3, P_5]$	0		
$\Rightarrow \min(9, 7, 10, 6, 9, 8) \Rightarrow 6$	$[P_2, P_4]$	6	0	

Agglomerative clustering (Complete linkage)

Take max distance instead of min and just solve in same way.

Evaluation of clustering Algorithm

Three important factors

- a) Clustering ~~probability~~ tendency
- b) Number of clusters, k
- c) Clustering quality

(a) clustering tendency

- Before evaluating the clustering performance, making sure that the data set we are working has clustering tendency and does not contain uniformly distributed points is very important.
- If the data does not contain clustering tendency, then clusters identified by any state of the art clustering algorithm may be irrelevant.
- Non-uniform distribution of points in data set becomes important in clustering.

o) Number of clusters, k

Some of the clustering algorithms like K-means, require number of clusters, k , as clustering parameter. Getting the optimal number of clusters is very significant in the analysis.

- If k is too high, each point will broadly stay representing a cluster and if k is too low, then data points are incorrectly clustered.
- Finding the optimal number of clusters leads to good quality in clustering.
- There are two major approaches to find optimal number of clusters:
 - 1) Domain Knowledge
 - 2) Data Driven approach

Domain knowledge :- Domain knowledge might give some prior knowledge on finding number of clusters. For example, in case of clustering iris data set, if we have the prior knowledge of species (virginica, versicolor), then $K=3$. Domain knowledge driven K Value gives more relevant insights.

Data driven approach - If the domain knowledge is not available, mathematical methods help in finding out right number of clusters.

c) Clustering quality

- Once clustering is done, how well the clustering has performed can be quantified by a number of metrics. Ideal clustering is characterized by minimal intra cluster distance and maximal inter cluster distance.
- There are majorly two types of measures to assess the clustering performance.

- i) Extrinsic Measures which require ground truth labels. Examples are Adjusted Rand Index, Faulkes - Mallows Scores, Mutual information based Scores, Homogeneity, Completeness and V-measure.
- ii) Intrinsic Measures that does not require ground truth ~~to~~ labels - Some of the clustering performance measures are Coefficients, Calinski - Harabasz Index, Davies - Bouldin Index etc.

Association Rule Learning

Association rule learning (Association rule mining) is a rule-based machine learning method for discovering interesting relations between variables in large databases. It simply how items are associated to each other.

Association analysis is which attempts to find common patterns of items in large data sets - being used in market basket analysis

Association Rule Mining

$$\text{IF } A \Rightarrow B \text{ THEN }$$

An association rule has two parts: an antecedent (if) and a consequent (then).

- An antecedent is an item found within the data.
- A consequent is an item found in combination with the antecedent.

If you have a Then there is Possibility you will get 3

There are three common ways to measure association.

$$\rightarrow \text{Support} = \frac{\text{Freq}(A, B)}{N}$$

This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears.

$$\rightarrow \text{Confidence} = \frac{\text{Freq}(A, B)}{\text{Freq}(A)}$$

This says how likely item B is purchased when item A is purchased, expressed as $\{A \rightarrow B\}$

$$\rightarrow \text{Lift} = \frac{\text{Support}(A, B)}{\text{Support}(A) \times \text{Support}(B)}$$

This says how likely item B is purchased when item A is purchased, while controlling for how popular item B is.

~~Support~~

~~Support~~

~~Support~~

classmate

Date _____
Page _____

T1	A	B	C
T2	A	C	D
T3	B	C	D
T4	A	D	E
T5	B	C	E

Transaction at a local market

Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B, C \Rightarrow A$	1/5	1/3	5/9

Apriori Algorithm

Apriori Algorithm is a classical algorithm in data mining. It is used for mining frequent itemsets and relevant association rules.

Sg

Min Support = 50%

Threshold Confidence = 70%

T ID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

(I)

Items	Support
1	$\frac{2}{4} \rightarrow 50\%$
2	$\frac{3}{4} \rightarrow 75\%$
3	$\frac{3}{4} \rightarrow 75\%$
x 4	$\frac{1}{4} \rightarrow 25\%$
5	$\frac{3}{4} \rightarrow 75\%$

Itemset 4 have 25% support which is less than 50% so we do not take it in new table

(II)

Items	Support
x {1, 2}	$\frac{1}{4} \rightarrow 25\%$
{1, 3}	$\frac{2}{4} \rightarrow 50\%$
x {1, 3}	$\frac{1}{4} \rightarrow 25\%$
{2, 3}	$\frac{3}{4} \rightarrow 75\%$
{3, 5}	$\frac{3}{4} \rightarrow 75\%$
{3, 5}	$\frac{3}{4} \rightarrow 75\%$

Check {1, 2} Come in item how many time EK Bath

1, 3, 3, 5

11 → 15
26 then

~~classmate~~
Date _____
Page _____

② Itemset

X {1, 3, 53}
{2, 3, 53}
X {1, 2, 3, 53}

Support

$$\frac{1}{4} = 25\% \\ \frac{2}{4} = 50\% \\ \frac{1}{5} = 25\%$$

{2, 3, 53} | 2

Itemset

(2¹3) → 5

(3¹5) → 2

(2¹5) → 3

2 → (3¹5)

5 → (2¹3)

3 → (2¹5)

Support

$$2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2$$

Confidence

$$\frac{3}{2} = 100\%$$

$$\frac{2}{2} = 100\%$$

$$\frac{3}{3} = 66\%$$

$$\frac{2}{3} = 66\%$$

$$\frac{2}{3} = 66\%$$

$$\frac{2}{3} = 66\%$$

Confidence

S(A ∪ B)

S(A)

{2, 3, 5})

$$\text{eg } (2^1 3) \rightarrow 5 = \frac{S((2^1 3) \cup 5)}{S(2^1 3)} = \frac{2}{2} = 100\%$$

→ {2, 3, 5})

Eliminate whose Confidence less than 70%

PCA problem!

Q1 Given the following data use PCA to reduce the dimension from 2 to 1.

Feature	Example 1	Example 2	Example 3	Example 4
x	4	8	13	7
y	11	4	8	14

Ans Step 1 No. of features = 2
No. of samples = 4

Step 2 : Computation of mean of variables

$$\bar{x} = \frac{4+8+13+7}{4} = 8$$

$$\bar{y} = \frac{11+4+5+14}{4} = 8.5$$

Step 3 : Computation of Covariance matrix

Ordered pair or
 $(x_1, x_1), (x_1, y_1), (y_1, x_1), (y_1, y_1)$

n^2 ordered pair

1 Covariance of all ordered pair

$$\begin{aligned} \text{Cov}(x, x) &= \frac{1}{N-1} \sum_{k=1}^N (x_{i,k} - \bar{x}_i)(x_{j,k} - \bar{x}_j) \\ &= \frac{1}{4-1} \sum_{k=1}^4 (x_{i,k} - \bar{x}_i)^2 \\ &= \frac{1}{3} [(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2] \\ &= 14 \end{aligned}$$

$$\boxed{\text{Cov}(x, x) = \frac{1}{N-1} \sum_{k=1}^N (x_{i,k} - \bar{x}_i)^2}$$

$$\begin{aligned} \text{Cov}(x, y) &= \frac{1}{4-1} [(4-8)(11-8.5) + (8-8)(7-8.5) \\ &\quad + (13-8)(5-8.5) + (7-8)(11-8.5)] \\ &= -11 \end{aligned}$$

$$\text{Cov}(y, x) = \text{Cov}(x, y) = -11$$

$$\begin{aligned} \text{Cov}(y, y) &= \frac{1}{4-1} [(11-8.5)^2 + (7-8.5)^2 + (5-8.5)^2 + (11-8.5)^2] \\ &= 23 \end{aligned}$$

(2)

Covariance matrix

 $n \times n \quad 2 \times 2$

$$S = \begin{bmatrix} \text{cov}(x, x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{cov}(y, y) \end{bmatrix}$$

$$= \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

(3)

Eigen value, Eigen ~~vector~~ vector, Normalized, eigen vector

i) Eigen value

$$\det(S - \lambda I) = 0$$

$$\det \begin{pmatrix} 14 - \lambda & -11 \\ -11 & 23 - \lambda \end{pmatrix} = 0$$

$$(14 - \lambda)(23 - \lambda) - (-11)(-11) = 0$$

$$\lambda^2 - 37\lambda + 201 = 0$$

$$\lambda_1 = 30.3849, \lambda_2 = 6.6151$$

$$\lambda_1 > \lambda_2$$

$\lambda_1 = 30.3849 \Rightarrow$ First Principal component

$$\lambda_2 = 6.6151$$

Eigen vector of λ_1

$$(S - \lambda_1 I) U_1 = 0$$

$$\begin{bmatrix} 14 - \lambda_1 & -11 \\ -11 & 23 - \lambda_1 \end{bmatrix} \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \end{bmatrix} = 0$$

~~$$(14 - \lambda_1) \bar{U}_1 - 11 \bar{U}_2 = 0$$~~

$$\begin{bmatrix} (14 - \lambda_1) \bar{U}_1 - 11 \bar{U}_2 \\ -11 \bar{U}_1 + (23 - \lambda_1) \bar{U}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\frac{\bar{U}_1}{11} = \frac{\bar{U}_2}{14 - \lambda_1} = t$$

$$(14 - \lambda_1) \bar{U}_1 - 11 \bar{U}_2 = 0$$

$$(14 - \lambda_1) \bar{U}_1 = 11 \bar{U}_2$$

$$\bar{U}_1 = 11$$

$$\bar{U}_2 = 14 - \lambda_1$$

$$\frac{\bar{U}_1}{11} = \frac{\bar{U}_2}{14 - \lambda_1} = t$$

Eigen vector U_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14 - \lambda_1 \end{bmatrix}$

$$= \begin{bmatrix} 11 \\ 14 - 30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$

③ Normalize the eigen vector U_1 ,

$$e_1 = \begin{bmatrix} \frac{11}{\sqrt{11^2 + (-16.3849)^2}} \\ \frac{-16.3849}{\sqrt{11^2 + (-16.3849)^2}} \end{bmatrix} = \begin{bmatrix} 0.5535 \\ -0.8303 \end{bmatrix}$$

1

$$\lambda e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

(5) Derive new dataset

$$\begin{array}{c} e_{x_1} & e_{x_2} & e_{x_3} & e_{x_4} \\ P_{11} & P_{12} & P_{13} & P_{14} \\ PC & & & \\ \text{Contra} & & & \end{array}$$

$$P_{11} = e_1^T \begin{bmatrix} 4 & -8 \\ -11 & -8.5 \end{bmatrix}$$

$$= [0.5574 \quad -0.8303] \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= -4.3052$$

$$P_{12} = [0.5574, -0.8303] \begin{bmatrix} 8 & -8 \\ -5 & -8.5 \end{bmatrix}$$

$$= 3.7361$$

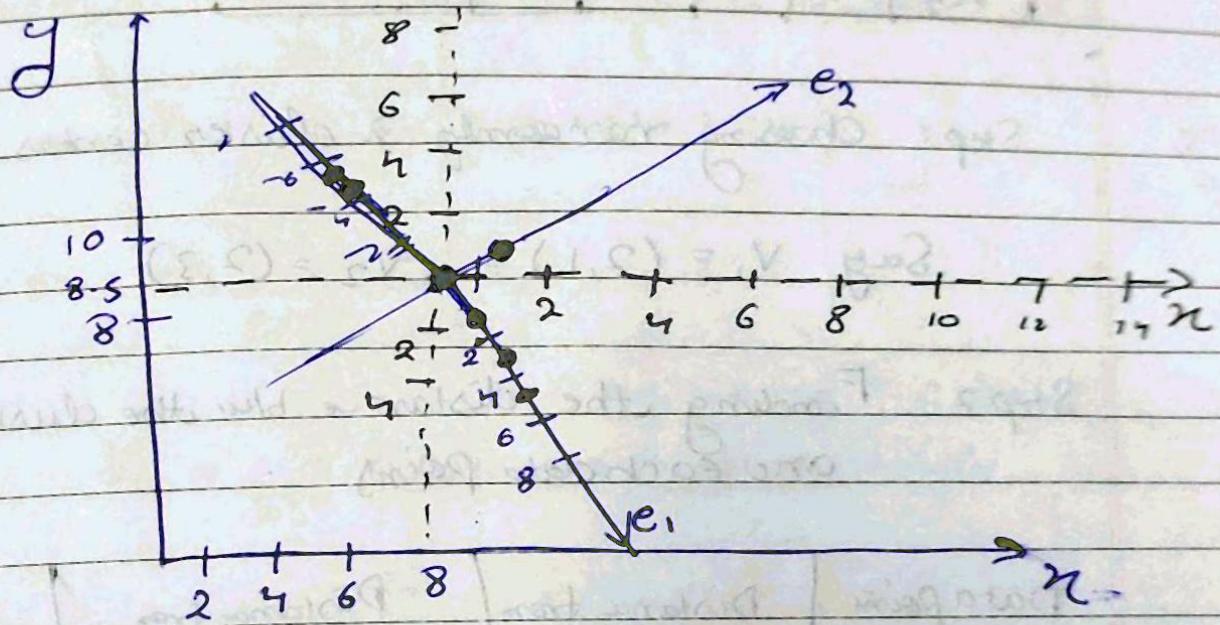
$$P_{13} = 5.6928$$

$$P_{14} = -5.1238$$

~~PC₁~~

Σx_1	Σx_2	Σz_3	Σx_4
PG1	-4.3052	3.7361	5.6528

Coordinate System for Principal Components



Q) use K-means Clustering algorithm to divide the following data into two clusters.

x_1	1	2	2	3	4	5
x_2	1	1	3	2	3	5

Step 1: Choosing randomly 2 cluster centers

$$\text{Say } v_1 = (2, 1) \quad v_2 = (2, 3)$$

Step 2: Finding the distance b/w the cluster centers and each data point

Data point	Distance from $v_1(2,1)$	Distance from $v_2(2,3)$	Assigned center
$a_1(1,1)$	1	2.27	v_1
$a_2(2,1)$	0	2	v_1
$a_3(2,3)$	2	0	v_2
$a_4(3,2)$	1.41	1.41	v_2
$a_5(4,3)$	2.83	2	v_2
$a_6(5,5)$	5	3.61	v_2

Step 3

cluster 1 of $v_1 \rightarrow \{a_1, a_2, a_3\}$
 cluster 2 of $v_2 \rightarrow \{a_3, a_5, a_6\}$

Step 4: Recalculate the cluster centers

$$v_1 = \frac{1}{3} [(1, 1) + (2, 1) + (3, 2)] \\ = \frac{1}{3} (6, 4) \\ = (2, 1.33)$$

$$v_2 = \frac{1}{3} [a_3 + a_5 + a_6] \\ = \frac{1}{3} [(2, 3) + (5, 3) + (5, 5)] \\ = \frac{1}{3} (11, 11) = (3.67, 3.67)$$

Step 5: Repeat from step 2 until we get same cluster center or same cluster elements as in previous iteration

Distance table:-

Data Points	Distance from (V ₁ , (2, 1.33))	Distance from V ₂ (3.57, 3.67)	Assigned Center
a ₁ (1, 1)	1.05	3.78	V ₁
a ₂ (2, 1)	0.33	3.15	V ₁
a ₃ (2, 3)	1.67	1.8	V ₁
a ₄ (3, 2)	1.204	1.8	V ₁
a ₅ (4, 3)	2.605	0.75	V ₂
a ₆ (5, 5)	4.74	1.88	V ₂

$$\text{Cluster 1 of } V_1 = \{a_1, a_2, a_3, a_4\}$$

$$\text{Cluster 2 of } V_2 = \{a_5, a_6\}$$

Recalculating the cluster center

$$V_1 = \frac{1}{4} [a_1 + a_2 + a_3 + a_4]$$

$$= \frac{1}{4} [(1, 1) + (2, 1) + (2, 3) + (3, 2)]$$

$$= \frac{1}{4} (8, 7) = (2, 1.75)$$

$$V_2 = \frac{1}{2} [a_5 + a_6]$$

$$= \frac{1}{2} [(4, 3) + (5, 5)]$$

$$= \frac{1}{2} [9, 8] = (4.5, 4)$$

So clusters elements and centers are not same as in the previous

Distance table

Data points	Distance from $v_1 (2, 1.75)$	Distance from $v_2 (7.5, 5)$	Assigned centers
$a_1 (1, 1)$	1.25	4.61	v_1
$a_2 (2, 1)$	0.75	3.9	v_1
$a_3 (2, 3)$	1.25	2.69	v_1
$a_4 (3, 2)$	1.03	2.5	v_1
$a_5 (4, 3)$	2.36	1.12	v_2
$a_6 (5, 5)$	4.42	1.12	v_2

Cluster 1 of v_1 :- $\{a_1, a_2, a_3, a_4\}$

cluster 2 of v_2 :- $\{a_5, a_6\}$

Cluster centers are same as in the previous iteration

Cluster 1 :- $\{(1, 1), (2, 1), (2, 3), (3, 2)\}$

Cluster 2 :- $\{(4, 3); (5, 5)\}$

Semi - Supervised Learning

Semi-Supervised Learning is a type of Machine Learning algorithm that represents the intermediate ground between Supervised and unsupervised learning algorithm. It uses a combination of labeled and unlabeled datasets during the training period.

Disadvantage of Supervised learning

- ~~The best~~ it requires hand-labeling by ML specialist or data scientists, and it also requires a high cost to process.

Disadvantage of unsupervised learning

- It also has a limited spectrum for its applications.
- ⇒ To overcome these drawbacks Semi-Supervised learning introduced.

Assumption

- **Continuity Assumption**:- The objects near each other tend to share the same group or label. This assumption is also used to supervised learning - and the classes are separated by the decision boundaries. But in semi-supervised the decision boundaries are added with the smoothness assumption in low-density boundaries.
- **Cluster assumption**:- data are divided into different discrete clusters. Further the points in the same cluster share the output label.
- **Manifold assumption**:- This assumption helps to use euclidean distances and densities, and thus data lie on a manifold of fewer dimension than input space.

working of Semi-Supervised learning.

Semi-Supervised learning uses - ~~Ground~~ labels to train the model with less labeled training data than supervised learning. The process can combine various neural network models and training ways. The whole working of Semi-Supervised learning is explained in the below points.

- Firstly it trains the models with less amount of training data similar to the Supervised learning models. The training continues until the model gives accurate results.
- The algorithm uses the unlabeled datasets with ~~Ground~~ labels in the next step. And now the result may not be accurate.
- Now the labels from labeled training data and ~~Ground~~ labels data are linked together.
- The input data is labeled training data and unlabeled training data are also linked.
- In the end, again train the model with the new combined input as did in the first step. It will reduce errors and improve the accuracy of the model.

Real world application of SVM

- 1) Speech analysis
- 2) web content Classification
- 3) Protein Sequence classification
- 4) Text document classifier.

Logistic Regression

Time Clicked on Ad Categorical Data Dependent Variable

Time	Clicked on Ad	Categorical Data	Dependent Variable
68.95	No		
80.23	No		
69.45	No		
74.75	No		
50.5	Yes		
55.5	Yes		
80.0	No		
70.5	No		

Target

Categorical

X Continuous / Numerical

Linear regression does not able to predict the dependent variable using independent variable.

unit -3K-medoid

It meas least dissimilar element or most similar one

There are some cons in K-means clustering
i.e. if this an object with an extremely large value may substantially distort the distribution of objects in clusters/groups. Hence it is sensitive to Outliers.

i	x	y
x_1	2	5
x_2	3	4
x_3	3	8
x_4	4	7
x_5	6	2
x_6	6	4
x_7	7	3
x_8	7	4
x_9	8	5
x_{10}	7	6

Step 1: We select two random representative objects:

$C_1(3, 4)$, $C_2(7, 4)$

i	x	y	C_1	Distance/Cost	C_1
x_1	2	6	3 4	$ 2-3 + 6-4 $	3
x_2	3	8	3 4	$0+4$	4
x_3	4	7	3 4	$1+3$	7
x_5	6	2	3 4	$3+2$	5
x_6	6	4	3 4	$3+0$	3
x_7	7	3	3 4	$4+1$	5
x_9	8	5	3 4	$5+1$	6
x_{10}	7	6	3 4	$4+2$	6

manhattan Distance = $|a-c| + |b-d|$

	C_2				
x_1	2	6	7	4	$ 2-7 + 6-4 $
x_3	3	8	7	4	$4+4$
x_4	4	7	7	4	$3+5$
x_5	6	2	3	4	$1+2$
x_6	6	4	7	4	$1+0$
x_7	7	3	7	4	$0+1$
x_9	8	5	7	4	$1+1$
x_{10}	7	6	7	4	$0+2$

Compare $\text{cost}(c_1)$ and $\text{cost}(c_2)$ for every i & select the minimum one

Step 2: Then clusters are

cluster 1: $\{(2,6), (3,8), (4,7), (3,4)\}$

cluster 2: $\{(7,4), (6,2), (6,4), (7,3), (8,5), (7,6)\}$

Calculate total cost

$$\text{cost}(n, c) = \sum_{i=1}^d |x_i - c_i|$$

Total cost = $\{\text{cost}((3,4), (2,6)), \text{cost}((3,4), (3,8)),$
 $\text{cost}((3,4), (4,7)), \text{cost}((7,4), (8,5)),$
 $\text{cost}((7,4), (6,2)), \text{cost}((7,4), (6,4)),$
 $\text{cost}((7,4), (7,3)), \text{cost}((3,4), (7,6))\}$

$$= (3+4+4) + (3+1+1+2+2) \\ = 20$$

Step 3 Select one of non-medoids = 0

Let's $0' = (7,3)$ i.e. x_7

So new measured's are $c(3, 4) \& o'(7, 3)$

i	x_i	y_i	o'	Distance/cost +	$C,$
x_1	2	6	7 3	$ 2-7 + 6-3 $	8
x_3	3	8	7 3	$4+5$	9
x_4	4	7	7 3	$3+4$	7
x_5	6	2	7 3	$1+1$	(2)
x_6	6	4	7 3	$1+1$	(2)
x_8	7	4	7 3	$0+1$	(1)
x_9	8	5	7 3	$1+2$	(3)
x_{10}	7	6	7 3	$0+3$	(3)

i	x_i	y_i	o'	Distance (cost +)	G
x_1	2	6	3 4	$ 2-3 + 6-4 $	(3)
x_3	3	8	2 4	$0+4$	(4)
x_4	4	7	3 4	$1+3$	(4)
x_5	6	2	3 4	$3+2$	5
x_6	6	4	3 4	$3+0$	3
x_8	7	4	3 4	$4+0$	4
x_9	8	5	3 4	$5+1$	6
x_{10}	7	6	3 4	$4+2$	6

Compare the cost of C_i and cost O_i for every i & Select the minimum one

Cluster 1 : $\{(3, 4), (2, 6), (3, 8), (7, 2)\}$

Cluster 2 : $\{(7, 3), (6, 2), (5, 4), (7, 7), (8, 5), (3, 6)\}$

$$\begin{aligned}\text{Current total cost} &= (3+4+7) + (2+2+1+3+5) \\ &= 11 + 11 \\ &= 22\end{aligned}$$

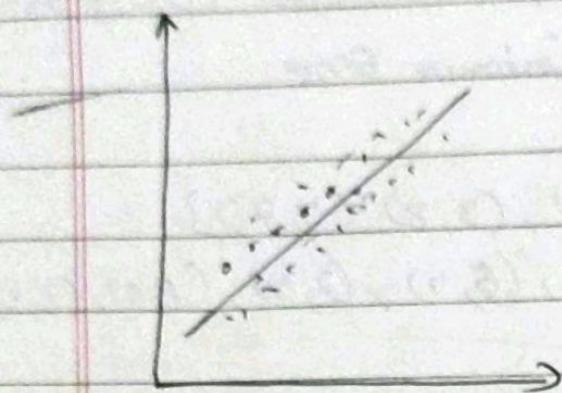
Step 4) So cost of Swapping medoid from C_2 to O_2 's

$$\begin{aligned}S &= \text{Current total cost} - \text{past total cost} \\ &= 22 - 20 \\ &= 2 > 0\end{aligned}$$

So moving O_2 would be bad idea. So previous choice was ~~poor~~ good.

If we get -ive value then we have to calculate again till we don't get +ve value.

Principal Component Analysis



- PCA is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning.
- It is a statistical process that converts the observation of correlated features into a set of linearly uncorrelated feature with the help of orthogonal transformation. These new transformed features are called the Principal Components.
- PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.