Name : Soham M. Pandit   Class : BE - A
Roll No : 34553   Batch : B3

Data Science and Vizualization CIE - II

Q1) A] Demonstrate how the K-means algorithm assigns data points to clusters with an example dataset. Also state the advantages and disadvantages of K-Means clustering.

→ Suppose we have 6 data points :
(1,1), (1.5,2), (3,4), (5,7), (3.5,5), (4.5,5)

We want to form K=2 clusters

- Step 1 : Initialization

Randomly pick 2 points as initial centroids say, (1,1) and (5,7)

- Step 2 : Assignment

Each point is assigned to the nearest centroid
Cluster 1 (closer to (1,1)) : (1,1), (1.5,2), (3,4)

Cluster 2 (closer to (5,7)) : (5,7), (3.5,5), (4.5,5)

- Step 3 : Update centroids

Mean of Cluster 1 $= ((1+1.5+3)/3, (1+2+4)/3)$
$= (1.88, 2.33)$

Mean of Cluster 2 $= ((5 + 3.5 + 4.5)/3,$
$(7 + 5 + 5)/3)$
$= (4.33, 5.67)$

- Step 4 : Repeat Assignment

Reassign points to new centroids → clusters stabilize after a few iterations

- Final Clusters :

Cluster 1 : $(1,1)$, $(1.5, 2)$, $(3,4)$
Cluster 2 : $(5,7)$, $(3.5, 5)$, $(4.5, 5)$

Advantages :
- Simple and fast, easy to implement and computationally efficient.
- Scales well to large datasets.
- Works well when clusters are spherical and well-seperated.

Disadvantages :
- Requires K (no. of clusters) to be specified in advance.
- Sensitive to initialization of centroids.
- Outliers can distort cluster centroids.

Page No.
Date

Q1) B] Illustrate how association rules are formed from frequent item sets with real-world example.

→ Lets take an example of Market Basket Analysis in which a supermarket wants to find relationships between items customers buy together.

- Step 1 : Transaction dataset

Transaction ID	Items Bought
T_1	{ Milk , Bread }
T_2	{ Milk , Diaper , Beer , Bread }
T_3	{ Milk , Diaper , Beer , Cola }
T_4	{ Bread , Milk , Diaper , Beer }
T_5	{ Bread , Milk , Diaper , Cola }

- Step 2 : Find frequent Itemsets
{ Milk , Bread } → 4 times
{ Milk , Diaper } → 4 times
{ Diaper , Beer } → 3 times
{ Milk , Diaper , Bread } → 3 times

- Step 3 : Generate Association Rules
From frequent itemset { Milk , Diaper } we can generate rules :

1) Milk → Diaper
- Support = 4 / 5 = 80%
- Confidence = (Milk + Diaper) / (Milk) = 4 / 5 = 80%

2) Diaper → Milk
- Support = 80%
- Confidence = (Milk + Diaper) / (Diaper) = 100%

3) Diaper → Beer
- Confidence = 3 / 4 = 75%

4) Beer → Diaper
- Confidence = 3 / 3 = 100%

Hence, we can interprete that 100% of the time when customers buy diapers, they also buy milk.

Q2) A] Explain decision tree and describe its main components with an example.

→ A decision tree is a supervised learning model used for classification and regression. It makes decisions by splitting data into branches based on feature values until a final outcome is reached.

Main components :

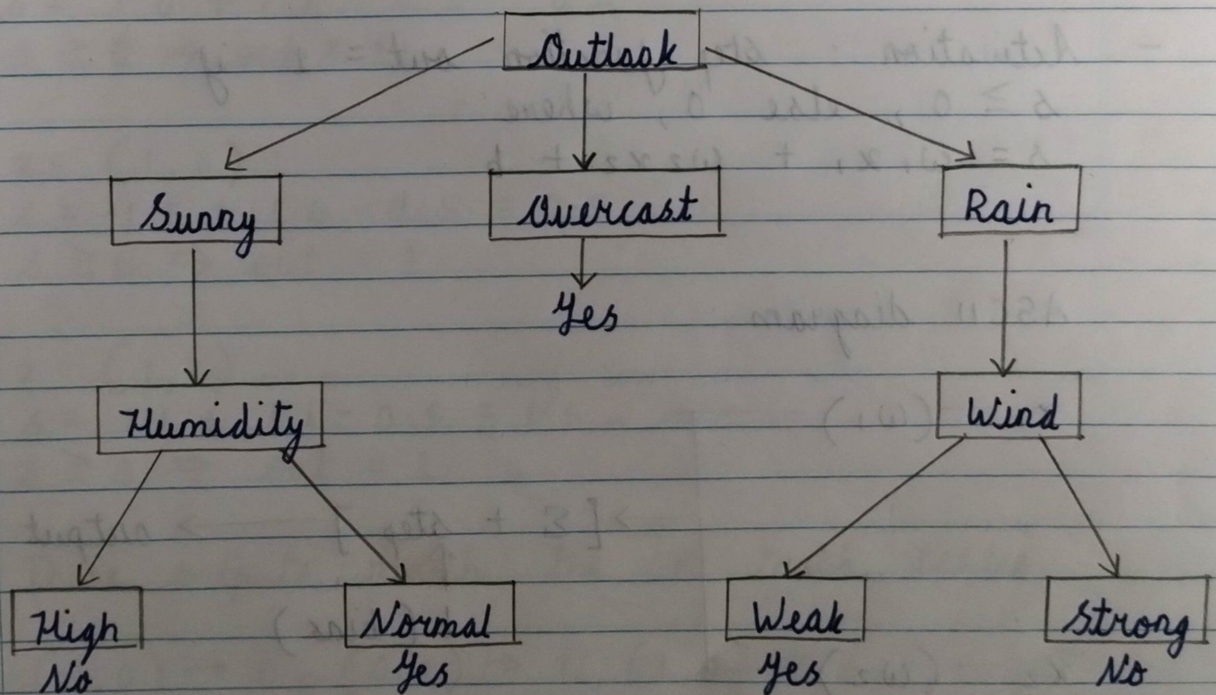
1) Root Node - The starting point (represents the entire dataset)

2) Decision Nodes - Internal nodes where data is split based on an attribute (eg. age > 30)

3) Branches - Edges that represent outcomes of a decision.

4) Leaf Nodes - Final outputs

Eg: Predicting if a person will play Tennis based on weather.



Q2) B] Illustrate step by step how a feed-forward neural network can be used to implement OR gate.

→ A single-layer feed-forward neuron with 2 inputs and a step activation can implement the OR function because OR is linearly seperable.

- Inputs : x_1, x_2
- Weights : ω_1, ω_2
- Bias : b (or equivalently an input $x_0 = 1$ with weight $w_0 = b$)
- Activation : step function out = 1 if $s \geq 0$, else 0, where $s = \omega_1 x_1 + \omega_2 x_2 + b$

ASCII diagram :

```
x1 —(w1)————┐
            │
            ├——→ [Σ + step] ————→ output (OR)
            │         ↑
            │     b (bias)
            │
x2 —(w2)————┘
```

Page No.
Date

- Choose weights and bias

Pick simple values :
$\omega_1 = 1$, $\omega_2 = 1$, $b = -0.5$

Activation : output = 1 if $s = x_1 + x_2 - 0.5 \geq 0$,
else 0

- Compute $s = \omega_1 x_1 + \omega_2 x_2 + b$, then apply step :

1) $x = (0,0)$:
$s = 1 \cdot 0 + 1 \cdot 0 - 0.5 = -0.5$
$s < 0 \Rightarrow$ out $= 0$

2) $x = (0,1)$:
$s = 1 \cdot 0 + 1 \cdot 1 - 0.5 = 0.5$
$s \geq 0 \Rightarrow$ out $= 1$

3) $x = (1,0)$:
$s = 1 \cdot 1 + 1 \cdot 0 - 0.5 = 0.5$
$s \geq 0 \Rightarrow$ out $= 1$

4) $x = (1,1)$:
$s = 1 \cdot 1 + 1 \cdot 1 - 0.5 = 1.5$
$s \geq 0 \Rightarrow$ out $= 1$

These outputs match the OR truth tables :

$(0,0) \rightarrow 0$, $(0,1) \rightarrow 1$, $(1;0) \rightarrow 1$, $(1,1) \rightarrow 1$