G H RAISONI COLLEGE OF ENG., NAGPUR
END SEMESTER EXAMINATION: SUMMER-2021
SESSION 2020-2021 (ONLINE)

DEPARTMENT: ARTIFICIAL INTELLIGENCE
SEM/SEC: 4th A DATE: 11/03/2021
SUBJECT: MLA COURSE CODE: BAIL20
ROLL NO: A-58
NAME: SHIVAM TAWARI
REG NO: 2019AAIE1117028

Q.1. CO1

Q.

1. Clustering:
Clustering is a machine learning technique in which it groups the the unlabelled dataset.

Example: Netflix uses clustering to show the recommedations to the users. It seeks to substantially improve the accuracy of predictions

about how much someone is going to love a movie based on their movie preferences.

2. Classification :
Classification algorithm is a supervised learning technique that is used to identify the category of new observations on the basis of training data. Then it learns from the given dataset or observations and then classifies new observations into a number of classes or groups.

Example: When we want to classify between birds and reptile. Our program classify them into two different classes

3. Supervised Learning :
Supervised Learning is trained well using "labelled" ..training data

Pg.no.2

1. MLA - 11 AM - 1 PM - [illegible]

and basis of that data, machine predict the output.

Example:
Spam filtering can be done using supervised machine learning algorithm. It will classify it into spam or not spam.

4. Unsupervised Learing:
Unsupervised learning is a ML technique in which models are not supervised using training dataset.

Example:
Unsupervised can be used in customers segments in marketing data. Being able to determine segments of customers helps in marketing teams approach these customers segments in unique ways.

5. Hypothesis Space:
The hypothesis space utilized by an ML system is the arrangement of all hypothesis that may be returned by it.

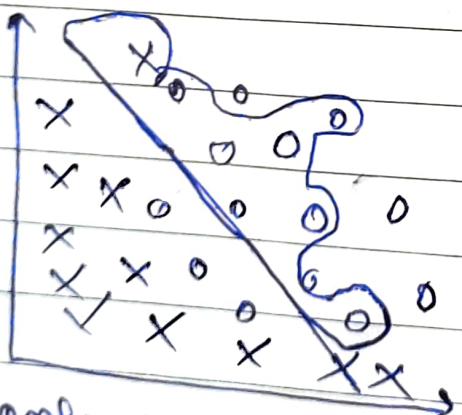
Example:
The input space is in the given example 2^4, is the number of possible inputs. The hypothesis space is $2^{2^4} = 65536$ because for each set of features of the input space two outcomes (0 and 1) are possible.

Q.2.

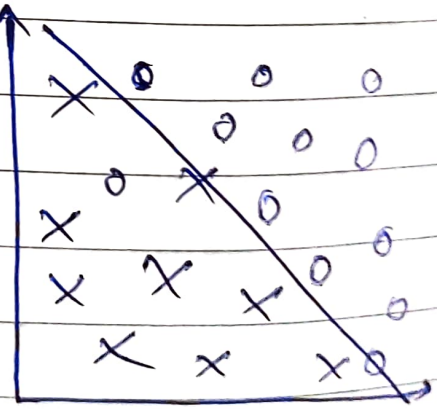
Overfitting	Underfitting
• Fitting the data too well.	Learning too little of the true concept.
• Features are noisy and uncorrelated to concept	Features don't capture concept

• Modeling process very powerful.	There is too much bias in model.
• There is too much search.	Too little search to fit model.
• Good performance on training data.	Poor performance on training data.
• Poor generlicezation to other data.	Poor generlization to othe data.

Example:



Model performing too well!
Overfitting
Accuracy : 100 %.



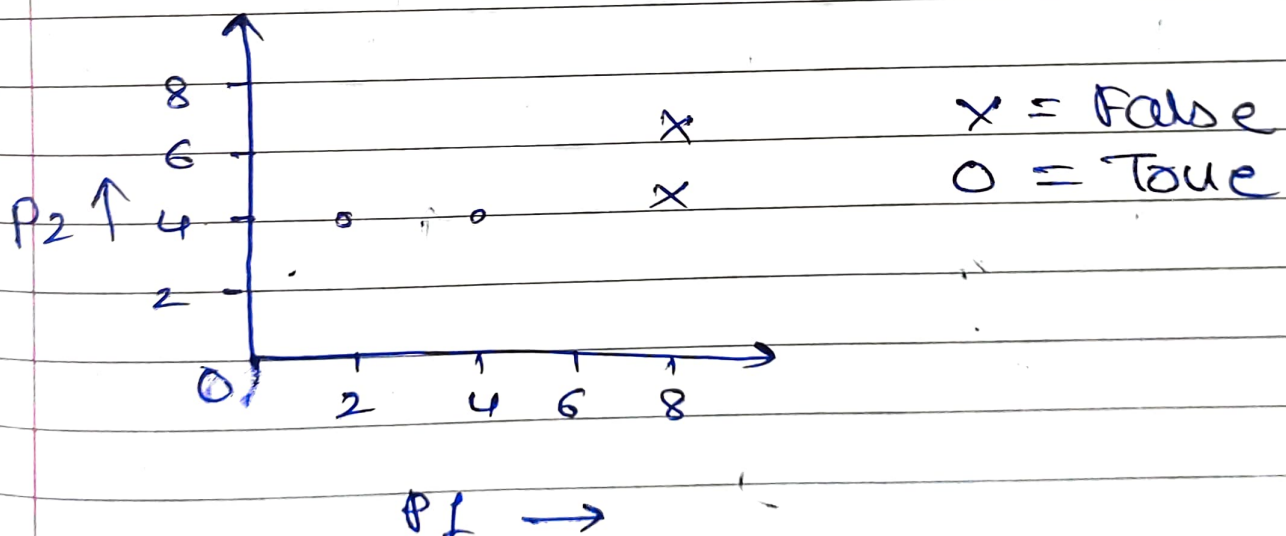
Bad performance!
Underfitting
Accuracy: 50%.

2.
CO2

a. Given ⇒ $K = 3$

P1	P2	Class
7	7	False
7	4	False
3	4	True
1	4	True

Graph:



New point: $P_1 = 3$ and $P_2 = 7$

Comparing new observation with every other observation in dataset.

1. MLA - 11AM - 1PM - BAIL203

Eucledian distance measure is used to select first ~~k~~ = 3

$$dist = \sqrt{(P_1 - P_1')^2 + (P_2 - P_2')^2}$$ observations

P_1	P_2	dist (3, 7)	Class
7	7	$\sqrt{(7-3)^2+(7-7)^2} = 4$	False
7	4	$\sqrt{(7-3)^2+(4-7)^2} = 5$	False
3	4	$\sqrt{(3-3)^2+(4-7)^2} = 3$	True
1	4	$\sqrt{(1-3)^2+(4-7)^2} \approx 3.6$	True

Closest k=3 neighbours are True (3, 4), True (1, 4), False (7, 7)

∴ By majority voting, new observation will be true (3, 7).

Q.3.
CO3.

a. The goal of SVM algorithm is to create the best fit line or decision boundary

that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme vectors that help in creating the hyperplane.

Linear SVM is used for linearly separable data, which means the dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data.

Q3
CO3
b. Let R be rain;

(a) $P_r(R) = \frac{5}{365} = \frac{1}{73}$

(b) When 90% rains;

(b) $P_r(F|R) = \frac{9}{10}$

(c) When rain is 10%;

$P_r(F|\bar{R}) = \frac{1}{10}$

$P_r(R) + P_r(\bar{R}) = 1$

$P_r(\bar{R}) = \frac{72}{73}$

So by Bayes theorem:

$$P_\delta(A|B) = \frac{P_\delta(A)\, P_\delta(B|A)}{\sum P_\delta(A)\, P_\delta(B|A)}$$

Substituting into Bayes theorem:

$$P_\delta(R|F) = \frac{P_\delta(R)\, P_\delta(F|R)}{P_\delta(R)\, P_\delta(F|R) + P_\delta(\bar{R})\, P_\delta(F|R)}$$

$$= \frac{1/73 \times 9/10}{(1/73 \times 9/10) + (72/73 \times 1/10)}$$

$$= \frac{9/730}{9/730 + 72/730}$$

$$= \frac{9}{9+72}$$

$$= \frac{1}{9} \approx 11\%$$

So, there is 11% chance that it will rain on the Friday of Madhuri's wedding.

Q. 4.

CO 4.

Q.

A = x1	x2
1.4	1.65
1.6 -	1.975
-1.4	-1.775
-2.0	-2.525
-3.0	-3.95
2.4	3.075
1.5	2.015
2.3	2.75
-3.2	-4.05
-4.1	-4.85
M = mean : -0.45	: -0.5625

Centered matrix $C = A - M$

$C = X1 + 0.4S$	$X2 + 0.567S$
1.8S	2.217S
2.0S	2.542S
-0.9S	-1.207S
-1.5S	-1.857S
-2.5S	-3.382S
2.8S	3.642S
1.9S	2.592S
2.7S	3.377S
-3.6S	-4.282S

① Covariance :

$$Cov(x, y) = \frac{\Sigma (x_i - \bar{x}) \times (y_i - \bar{y})}{N}$$

$$= 7.98$$

$$COV(x, x) = \frac{\Sigma (x_i - \bar{x})^2}{N}$$

$$= 6.42$$

$$Cov(x, y) = \frac{\Sigma(y_i - \bar{y})^2}{N}$$

$$= 9.95$$

$$Cov(y, x) = Cov(x, y)$$

$$= 7.98$$

$$\therefore \text{Covariance matrix} = \begin{bmatrix} 6.42 & 7.98 \\ 7.98 & 9.95 \end{bmatrix}$$

Eigen values & vectors:

$$(A - \lambda I) V = 0$$

$$|A - \lambda I| = 0$$

$$\begin{vmatrix} 6.42 - \lambda & 7.98 \\ 7.98 & 9.95 - \lambda \end{vmatrix} = 0$$

$$(6.42 - \lambda)(9.95 - \lambda) - (7.98)(7.98) = 0$$

Solving eqn. we get

Eigen values: $\lambda_1 = 0.00746$
$\lambda_2 = 16.36$

Putting λ in eigen decomposition.

$$\begin{bmatrix} 6.42 - \lambda & 7.98 \\ 7.98 & 9.98 - \lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Vectors $\Rightarrow$ $\begin{bmatrix} -0.779 \\ 0.626 \end{bmatrix}$ for $\lambda_1 = 0.007$

$\begin{bmatrix} -0.626 \\ -0.779 \end{bmatrix}$ for $\lambda_2 = 16.36$

1.5.
6.5
4. *kmeans algorithm is an iterative algorithm that divides a group of n datasets into K subgroups /clusters based on the similarity

1. MLA -11AM-PM-B7A-L02 Pg. no. 14

and their mean distance from the centroid of that particular subgroup formed.

Algorithm of K-means:

Step 1: Select the value of K, to decide the number of clusters to be formed.

Step 2: Select random K points which will act as centroids.

Step 3: Assign data point, based on their distance from the randomly selected points, to the nearest centroid which will form the predefined clusters.

Step 4: Place a new centroid of each clusters

Step 5: Repeat step no. 3, which reassign each datapoint to the new closest centroid of each clusters.

Step 6: FINISH.

5. Gaussian Mixture Models assume that there are a certain number of gaussian distributions and each of these distributions represent a cluster, hence, a GMM tends to group the points belonging to a single distribution together.

We have taken unlabelled input data, which means its not categorized and corresponding outputs are not given. Now it is fed to the machine learning model in to train it. Firstly, it will interpret the row data to find the hidden patterns from data and then apply clustering.