Assignment 1

Tuesday, 6 December 2022 12:42 AM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE4220003

Questions solved in this Assignment.

→ T1 Question Paper Solution

Q1 a) Linear Regression can be used to predict the stock prices of NIFTY. LR is a type of supervised learning technique.

b) Parameters used to predict the prices can be

① opening price of a stock
② highest price possible at any time.
③ lowest " " " " ".
④ Closing " of stock.

c) Yes, there is a limitation with Linear Regression. It is limited to linear relationships only. By its nature linear Regression only looks at linear relationship between dependent & independent variable & ignores any non-linear relationship.

Q2 There are 5 stages in machine learning project life cycle.

① Project Initiation

- this step involves 3 major things
  - Idea
  - Requirements
  - Data Aquisition.
- Idea is developed & analysed so that the project requirements are clear & data can be aquired which basically is the back-bone of any ML project.

② Data Exploration: here in this phase the data aquired is explored in depth to understand the nature of the data, the kind of relationship its attributes hold etc.

③ data preprocessing & feature selection

- this phase involves
  - Removing irrelevant features from dataset.
  - handling NULL or missing values.
  - Reducing the size of dataset.
  - transform categorical variable to numerical.
  - Normalising data points.

④ Model Development

now the actual model is developed while choosing the algo we consider:

- Data size
- Type of problem
- Availability of packages.

⑤ Model Evaluation

- In this phase the model is evaluated based on its prediction accuracy by multiple metrics, such as precision, recall, F1-score. etc.

Q3

X	Y	distance from 6.5
1	23	5.5
1.2	17	5.3
3.2	12	3.3
4	27	2.5
5.1	8	1.4
6.5	?	0

If $K = 5$ then all are selected. as NN.

then prediction is $\frac{23+17+12+27+8}{5}$

$= 17.4$

Q4 Advantages → ① No training period.
② & also due to that data can be added to the dataset at any time
③ Very easy to implement.

Dis-Advantages → ① Does not work well with large datasets.
② Does not work well with v. large datasets.
③ Need feature scaling.
④ Sensitive to missing noisy data & outliers.

→ best way to use KNN so it works best is by choosing the right value of k.

Assignment 2

Tuesday, 6 December 2022 2:22 AM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE4220003

Questions solved in this Assignment.

① → 2 Examples of KNN
② → Write basic M/c Learning/AI Theory

Ques1 KNN example 1

	Height	Weight	Class
①	162	66	NR
②	171	62	NR
③	166	51	UW
④	150	110	OW
⑤	157	67	NR
⑥	163	61	UW
⑦	170	101	OW

Query → 165, 91, ? using Euclidian dist.
k = 5

d① $\sqrt{(165-162)^2 + (91-66)^2} = \sqrt{9+625} = 25.17$
d② $\sqrt{(165-171)^2 + (91-62)^2} = 29.61$
d③ $\sqrt{(165-166)^2 + (91-51)^2} = 40.01$
d④ $\sqrt{(165-150)^2 + (91-110)^2} = 25$
d⑤ $\sqrt{(165-157)^2 + (91-67)^2} = 25.29$
d⑥ $\sqrt{(165-163)^2 + (91-61)^2} = 30.06$
d⑦ $\sqrt{(165-170)^2 + (91-101)^2} = 11.18$

4 nearest neighbours = d1, d4, d5, d7, d2
NR, OW, NR, OW, NR

⇒ 165, 91, NR is answer.

Ques2 KNN example 2

	Blood Sugar level	Age	D/ND
d1	100	31	ND
d2	125	35	D
d3	112	38	D
d4	176	41	D
d5	79	15	ND
d6	92	19	ND

Week2 3 ((140, 35), ?) Euclidian

d1 $\sqrt{(140-100)^2 + (35-31)^2} = 40.19$
d2 $\sqrt{(140-125)^2 + (35-35)^2} = 15$
d3 $\sqrt{(140-112)^2 + (35-38)^2} = 28.16$
d4 $\sqrt{(140-176)^2 + (35-41)^2} = 36.49$
d5 $\sqrt{(140-79)^2 + (35-15)^2} = 64.19$
d6 $\sqrt{(140-92)^2 + (35-19)^2} = 50.59$

3 NN are d2, d3, d4
D, D, D

∴ (140, 35, D) is the answer.

- **Basics of AI & Machine Learning. Theory.**

Artificial Intelligence is m/c ability to observe, think & react like human beings. It's grounded in the idea that human intelligence can be broken down into precise ability which computers can be programmed to mimic. AI is an umbrella term that encompasses a wide variety of concept & technologies including machine learning.

AI consist of many subfields that uses technique to mimic specific behaviours we associate with human intelligence. for example, humans can speak, hear, read & write language & glean meaning from it. the field of speech recognition & NLP mimic these abilities by converting raw audio signals into text & processing that text to extract meaning from it. other subfields are also building intelligent systems that replicate human behaviours such as:

→ Robotics.
→ Pattern Recognition.
→ Computer Vision.

Assignment 3

Tuesday, 6 December 2022 3:04 AM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE4220003

Questions solved in this Assignment.

1. 1 Example of Either SVD or PCA
2. " " " K-means
3. " " " Naive Bayes

Q1) SVD Example

$$A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

$$AA' = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \times \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix}$$

$$\begin{bmatrix} 16+0 & 12+0 \\ 12+0 & 9+25 \end{bmatrix}$$

$$\begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix}$$

→ finding Eigen vector for AA'

$$|AA' - \lambda I| = 0$$

$$\begin{vmatrix} 16-\lambda & 12 \\ 12 & 34-\lambda \end{vmatrix} = 0$$

$$(544 - 50\lambda + \lambda^2) - 144 = 0$$

$$\lambda^2 - 50\lambda + 400 = 0$$

$$(\lambda - 10)(\lambda - 40) = 0$$

$$\therefore \lambda = 10, 40$$

- Eigen Vector for $\lambda = 40$

$$A.A' - \lambda I = \begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix} - 40\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -24 & 12 \\ 12 & -6 \end{bmatrix}$$

now reducing the matrix.

$R_1 \leftarrow R_1 \div -24$

$R_2 \leftarrow R_2 - 12 \times R_1$

$$= \begin{bmatrix} 1 & -.5 \\ 0 & 0 \end{bmatrix}$$

System associated with Eigen value $\lambda = 40$

$$(A.A' - 40I)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -.5 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 - \frac{1}{2}x_2 = 0$$

$$x_1 = \frac{1}{2}x_2$$

Eigen Vector

$$v = \begin{bmatrix} .5x_2 \\ x_2 \end{bmatrix} \quad \text{If } x_2 = 1$$

then $v_1 = \begin{bmatrix} .5 \\ 1 \end{bmatrix}$

- Eigen Vector for $\lambda = 10$

$$|A.A' - \lambda I| = \begin{bmatrix} 16 & 12 \\ 12 & 32 \end{bmatrix} - 10\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 12 \\ 12 & 24 \end{bmatrix}$$

now reducing the matrix.

Interchange $R_1 \leftrightarrow R_2$

$R_1 / 12 \rightarrow R_1$
$R_2 - 6 \times R_1 \rightarrow R_2$

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

$$(A.A' - 10I)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 + 2x_2 = 0$$

$$x_1 = -2x_2$$

$$v = \begin{bmatrix} -2x_2 \\ x_2 \end{bmatrix}$$

let $x_2 = 1$

$$v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

- for v_1, $L = \sqrt{.5^2 + 1^2} = 1.11$

normalising $u_1 = \left[\frac{0.5}{1.11}, \frac{1}{1.11}\right]$

$$= (.4472, .8944)$$

- for v_2, $L = \sqrt{(-2)^2 + 1^2} = 2.23$

normalising $u_2 = \left(\frac{-2}{2.23}, \frac{1}{2.23}\right)$

$$= (-0.8944, .4472)$$

Solution

$$\Sigma \begin{bmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{bmatrix} = \begin{bmatrix} 6.32 & 0 \\ 0 & 3.16 \end{bmatrix}$$

$$V = [v_1, v_2] = \begin{bmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

U is found using formula $u_i = \frac{1}{\sigma_i} A.v_i$

$$\therefore U = \begin{bmatrix} -.447 & .894 \\ -.894 & -.447 \end{bmatrix}$$

• k-means Clustering

Marks of students in T1

$\{4, 7, 8, 6, 14, 15, 10, 9\}$

Marks of student in T2

$\{18, 11, 5, 2, 7, 9, 12, 15\}$

$K = 2$ 2 cluster centers randomly

$C_1 = (8,5)$, $C_2 = (15,9)$

$ED_1 = \sqrt{(4-8)^2 + (18-5)^2} = 13.60$ (C1)
$\sqrt{(4-15)^2 + (18-9)^2} = 14.21$

$ED_2 = \sqrt{(7-8)^2 + (18-9)^2} = 6.08$ (C_1)
$\sqrt{(7-15)^2 + (11-9)^2} = 8.24$

$ED_3 = 0$

$ED_4 = \sqrt{(6-8)^2 + (2-5)^2} = 3.60$ (C_1)
$\sqrt{(6-15)^2 + (2-9)^2} = 11.40$

$ED_5 = \sqrt{(14-8)^2 + (7-5)} = 6.32$ (C_2)
$\sqrt{(14-15)^2 + (7-9)^2} = 2.23$

$ED_6 = 0$

$ED_7 = \sqrt{(10-8)^2 + (12-5)^2} = 7.28$ (C_1)
$\sqrt{(10-15)^2 + (12-5)^2} = 8.60$

$ED_8 = \sqrt{(9-8)^2 + (15-5)^2} = 10.04$ (C_2)
$\sqrt{(9-15)^2 + (15-9)^2} = 8.48$

$$\begin{bmatrix} C_1 = (4,18), (7,11), (6,2), (10,12) \\ C_2 = (14,7), (9,15) \end{bmatrix}$$ Answer

Naive Bayes Classification

	Purple	white	Support	total
Brinjal	100	350	50	500
Coconut	250	350	400	1000
others	50	100	200	350
total	400	800	650	1850

Item = {Purple, white support} = x

$$P(A/B) = \frac{P(B/A).P(A)}{P(B)}$$

Brinjal $P(x \mid Brinjal) = 0.2 \times 0.7 \times 0.1 = 0.014$

$$P(Purple \mid Brinjal) = \frac{\frac{100}{400} \cdot \frac{400}{1850}}{\frac{500}{1850}} = 0.2$$

$$P(white \mid Brinjal) = \frac{\frac{350}{800} \cdot \frac{800}{1850}}{500/1850} = 0.7$$

$$P(Support \mid Brinjal) = \frac{\frac{50}{650} \cdot \frac{650}{1850}}{\frac{500}{1850}} = 0.1$$

$$P(Purple \mid coconut) = \frac{\frac{250}{400} \cdot \frac{400}{1850}}{\frac{1000}{1850}} = 0.25$$

$$P(white \mid coconut) = \frac{\frac{350}{800} \cdot \frac{800}{1850}}{\frac{1000}{1850}} = 0.35$$

$$P(Support \mid coconut) = \frac{\frac{400}{650} \cdot \frac{650}{1850}}{\frac{1000}{1850}} = 0.4$$

$$P(Purple \mid others) = \frac{\frac{50}{400} \cdot \frac{400}{1850}}{350/1850} = 0.14$$

$$P(white \mid others) = \frac{\frac{100}{800} \cdot \frac{800}{1850}}{350/1850} = 0.28$$

$$P(Support \mid others) = \frac{\frac{200}{650} \cdot \frac{650}{1850}}{350/1850} = 0.57$$

Max Probability = 0.03 which is of coconut

Assignment 4

Tuesday, 6 December 2022   1:53 PM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE4220003

Questions solved in this Assignment.

① → 2 Examples of Neural Network

2 Questions on Neural Networks

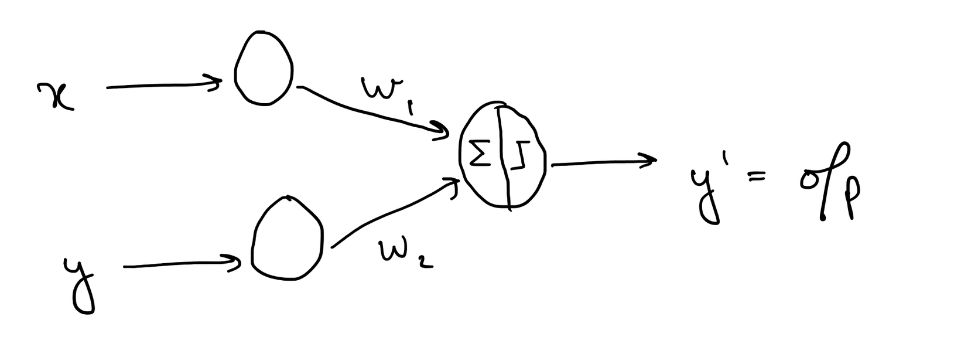
Q. Create a neural Network that can solve the problem of OR Gate.

Sol^n Initially w_1 & $w_2 = 0$

$$f(a) = \begin{cases} 0, & \text{if } a > 1 \\ 1, & \text{if } a \leq 1 \end{cases}$$

Truth table

x	y	o/p
0	0	0
0	1	1
1	0	1
1	1	1



→ $y' = f(w_1 x + w_2 y + b)$ $b = 1$

$x = 0 \quad y = 0$

$y' = f(0 + 0 + 1)$

$= f(0 + 1)$

$= f(1) \Rightarrow y' = 0$

→ $x = 0 \quad y = 1$

$y' = f(0 + 1 + 1)$

$= f(2) \Rightarrow y' = 1$

→ $x = 1 \quad y = 0$

$y' = f(1 + 0 + 1)$

$= f(2) \Rightarrow y' = 1$

→ $x = 1 \quad y = 1$

$y' = f(1 + 1 + 1)$

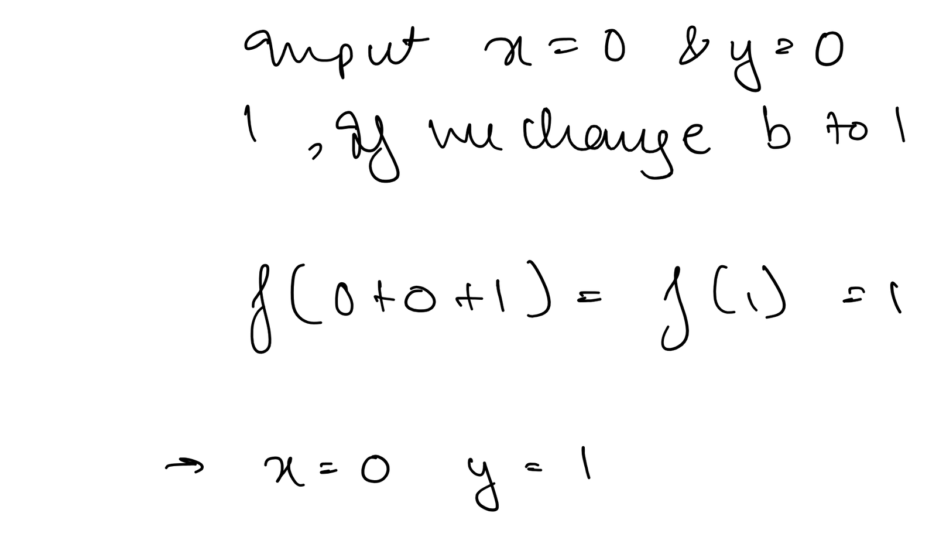
$= f(3) \Rightarrow y' = 1$

Q. Create a neural Network that can solve the problem of NAND Gate.

Sol^n Initially w_1 & $w_2 = 1$ & $b = -1$

Truth table

x	y	o/p
0	0	1
0	1	1
1	0	1
1	1	0



→ $y' = f(w_1 x + w_2 y + b)$ $b = -1$

$y' = f(0 + 0 - 1)$

$= f(-1) \Rightarrow y' = 1$

→ from perceptron rule $w_1 x + b \leq 0$, then $y' = 0$ then this row is incorrect as o/p should be 1 for NAND Gate.

→ So we want values that will make the input $x = 0$ & $y = 0$ to give y' value of 1, if we change b to 1 we have,

$f(0 + 0 + 1) = f(1) = 1$ this works.

→ $x = 0 \quad y = 1$

$y' = f(0 + 1 + 1) = 2$
$\Rightarrow y' = 1$

correct.

→ $x = 1 \quad y = 0$

$y' = f(1 + 0 + 1) = 2$

$\Rightarrow y' = 1$

correct.

→ $x = 1 \quad y = 1$

$y' = f(1 + 1 + 1) = 3$
$\Rightarrow y' = 1$

this is not correct.

∴ change values of w_1 & w_2 to -1
& b to 2
we get

→ $x = 1 \quad y = 1$

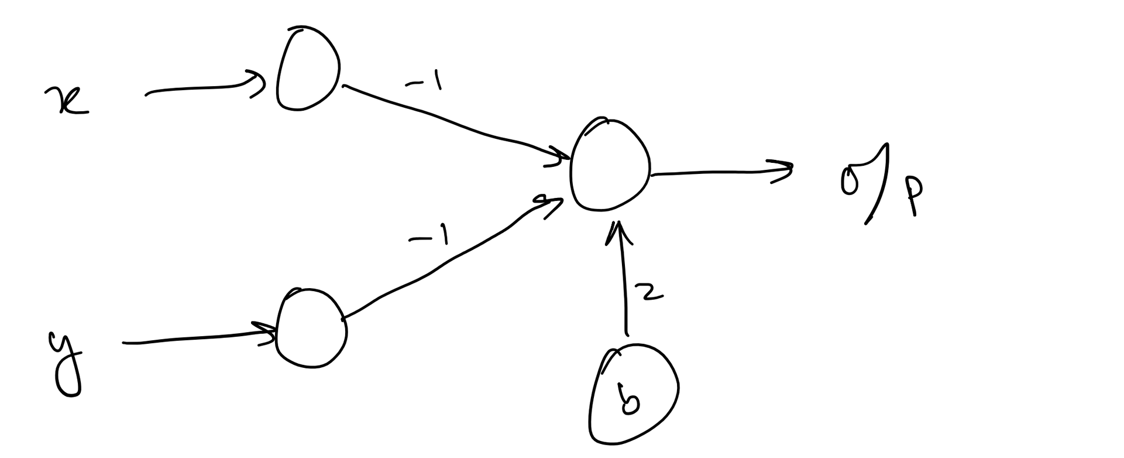
$y' = f((1 \times -1) + (1 \times -1) + 2)$

$\Rightarrow f(-1 - 1 + 2) = f(0)$

$\Rightarrow y' = 0$

correct.

∴ final NAND Gate neural N/w looks like this



Assignment 5

Tuesday, 6 December 2022 1:55 PM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE/22/0003

Questions solved in this Assignment.

① → Explore few feature selection & extraction Algorithms.

- **MFCC (mel-frequency cepstrum coefficients)**

In audio processing MFC is a short-term power spectrum of a sound, based on LCT of of a log power spectrum on a non-linear mel scale of frequency.

MFCC are coefficients that collectively make up MFC. They are derived from a type of cepstral representation of a audio clip. the difference b/w the MFC & MFCC is the spacing of frequency bands which approximates the human auditory system response. frequency wraping can also allow for better representation of sound.

- **SVD - Singular Value Decomposition.**

SVD is basically a matrix factorisation technique in which a matrix is decomposed into 3 matrices. It has some intresting Algebric properties & conveys imp. geometrical & theoritical insights about linear transformation

$$A \rightarrow U \Sigma V^T$$

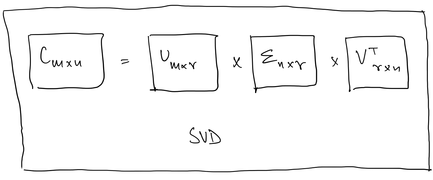
here,

A → Initial linear matrix to be decomposed.

U → m×n matrix of orthogonal eigen vectors of AA^T

V^T → transpose of n×n matrix containing orthogonal eigen vectors of A^TA

Σ → a n×n diagonal matrix of singular values which are the sq. root of the eigen value of A^TA



- **PCA Principle Component Analysis.**

- A statistical procedure which converts a set of observations of possibly co-relleated into a set of linearly uncorellated variable called principle components using an orthogonal transformation.

- The aim is to perform dimentionality reduction while preserving as much of the randomness in the high dimentional space as possible.

- It performs a co-ordinate rotation that aligns the axis with the direction of max. variance.

- the main limitation of PCA is that it does not consider the class separability since it does not take into account the class label of the feature vector.

- **LDA (Linear Discriminant Analysis)**

- The objective of LDA is to perform Dimentionality reduction while preserving as much of the class discriminatory as possible.

- PCA ignores class labels & focuses on finding the principle components that maximise the variance in the given data thus it is an unsupervised Algorithm, on the other hand LDA is supervised Algorithm. that intends to find linear Discriminants that represent those axis which maximize separation b/w different classes.