MACHINE LEARNING

☆ Learning System :- It is a system which provides the compilation of data about resources for learning

- When we talk about system; Every system has its own Input / output / process, this is how the system works

- Resource :- Devices required for ~~the~~ Input, process, output is known as resources.

- Management :- It manages the input devices, processing pattern, output devices along with that management is going to manage the information flow, material flow and money flow, these are the three different processes for three different flows comes into the ~~pitura~~ picture when deal with any kind of system

☆ LMS (Learning Management System) :-
- It enables you to create, manage and deliver something that encourages the same way like a word processor.
CRUD :- Create, Read, Update, Delete, Any software that allows you to do these things, this means ~~it~~ it ~~was~~ allows you to go for simulation. Eg: Modle

☆ Simulation :- It means some sort of CRUD operation.

☆ ~~Modle~~ :-
- Modle allows as to create programs, exams, MCQ, upload assignments etc.

PAGE No. 2
DATE / /

* Designing a learning System :- (In machine Learning)
  - According to Tom (Michial ← IDK the spelling), computer program is said to be learning from experience with ~~respect~~ respect to task and the performance measure is the performance at task 't' which is ~~denoted by~~ measured by p, & is proved in 'th experience 'e'. ~~For eg: In an e~~
  - For Eg :- In an email spam is detected which is one of the example.

* Goals of Machine Learning
  - Machine learning is the subfate of Artificial Intelligency & when we are going to deal with intelligency, Probability density function, Basically computer is based upon mathematics, so this terminology learning, machine learning, set theory, group theory, statistics, probability, membership function, all these terminologies comes from the mathematics.

- Its main goal and usage is to build new ~~or~~ and/or leverage existing algorithms to learn from data, in order to build generalizable models that give accurate predictions, or to find patterns, particularly with new and unseen similar data.

2nd Lecture : 18th Jan 2022

Imp Q.
* What is Machine Learning?
* What is Learning System?
* → Learning System is the system which provides the compilation of data about resources for learning. And when we talk about system, every system has its own IPU ie some sort of input is there or some sort of output is there

and have same pattern in there & finally it will generate the output. This is how the system is going to work

→ When we deal with the compilation of the data about resources for learning it mean that for creating & storing learning resources you need some sort of access to that learning resources. For us, in our system, resources are nothing but some sort of software or some sort of hardware or some sort of internet connector or network instrument, switches, hub, keyboard - storage devices. All the devices used for I/O are considered as Resource

★ Define Management:

→ Management is going to manage the input devices, processing pattern, output devices along with that, it will also manage the information flow, material flow, & money flow, these are the three different processes.

★ Three Aspects of Machine Learning

1. Information flow
2. Money flow
3. Material flow
4. Resources

task (t)
performance measure (p)
experience (e)

★ Designing A Learning System using ML:-

→ According to Tom Michael, one of the scientist/author, computer program is said to be learning from experience with respect to some task (t) & does the performance measure (p) which improves with experience (e), Okay so,

→ Eg:- Email - Spam detection

★ How do we calculate performance measure (P)?

→ For eg ; I have 10 different input devices and once all those 10 input devices is given to the system, and the program is able to identify those 10 different resources categorized as laptop, mobile, or desktop, or notepad or laptop. If it is identified properly then we can say that the performance is 100%.

→ If the program is only be able to detect 3 then only 30%.

★ 25th Jan, 2020

★ Is this machine learning the best way to achive your goals.

→ Example :- There is one interview going on, and there is only one vacancy, almost 100 applications people have received with that almost 90 applications are related with ML & remaining 10 are from other field. So is this course good? As machine learning will transform the word because the businesss that don't use machine learning will be out completed by those that do and machine learning will empower you to reach your goals so it will save money and time even lives okay so this is the expectations of what people can achive as far as machine learning is concerned.

★ How the machine learning helps to transform the world?

⇒ Basically it helps to transform the world so it is helpful to save the time & money & at the same time it helps to save the lives.

Eg : Image Recognition : If suppose you are given one sort of

brain
neuro image, and by looking into the image you should be able to say whether it is a left brain or right brain.
- Machine learing helps in translation of spoken words into text.

★ Next application of this is the statistical heritage regarding the statistics, quantitative data that we are going to be able to handle with the help of statistical techniques and if you would like to deal with linguistic terminology like high speed, low category, you have too much, you made a higher end form very tall, so if you want to classify these things then this kind of classification is approximately better to achive with the help of machine learning as compared to statistics then some sort of learning associations are there, classification or data will be classified into no. of classes, the way students are classified into different specialization, where somebody may go for machine learning, ~~xxxx~~ web technology, this is what we call as classification.

- Prediction :- for eg : In share market, We earn a profit ratio from respective from demand account. Share market is up & down income source. Its your prediction, sometime the profit percentage can increase and some day it may decrease this is how prediction works.

- Again there are couple of applications while extracting the data,
  - ~~Regression~~
  - Extraction of data
  - Regression of data

End of applications of ML

★ Developing a learning system requires (elements)
- task (t)
- performance measure (p)
- Training Experience (e)

The objective is to find the unknown target function.
- Basically when we deal with system, we have experience, in mind, have task infront of you & now we need to measure performance.
- When all these things comes into pic, how to train the system is a challenge

φ How to train the system is a challenge :-
- For this first decide what do you mean by task, what is the task, how do you go with it, & what to do exactly with the task. So task means we need to do some sort of work that is to be carried out with the help of the system. To carry, perform the process, design the process ~~and~~ inorder to implement that process what is required, how you get the results. When we come up with the task, the performance of the task is to be measured once it is implemented.

★ Neural Network or functioning of target :- This way be divided into 4 different aspects.
- ~~Re~~ Regression : It talks about variable count, independence.
-

★ Training Data :- It is an extremely large data set that is used to teach a machine learning model,

★ Target function :- Solving a problem, solving an algorithm to find the data Goal that you are trying to achieve, gaining more knowledge as we can. So instead of setting up the target quantitatively machine learning helps you helps you to setup target quantitatively.

★ How to setup Linguistic Target

★ Training Data & Training Information :- The raw items what we have collected ie. as data. Once it is processed it becomes an information, So processed data is called as informations and training data is an extremely large data set that is used to teach a machine learning model.

★ Training data under the supervised model and training data under the unsupervised model. So supervised models are the models where there is a need of a tutor (instructor). It is said to be unsupervised model when the tutor/instructor is not providing any kind of information that you needed in the sense you study on your own.

→ So for supervised machine learning models the training data is a load data & the training data is an initial set of data used to help the program to understand how to apply technologies like neural network to learn & produce sophisticated results. Consider model is nothing but one sort of box, where we are providing some input, some sort of data

and that box is going to process that data based on certain technologies. There are two aspects
- Aspect of Domain
- Horizontal technologies.

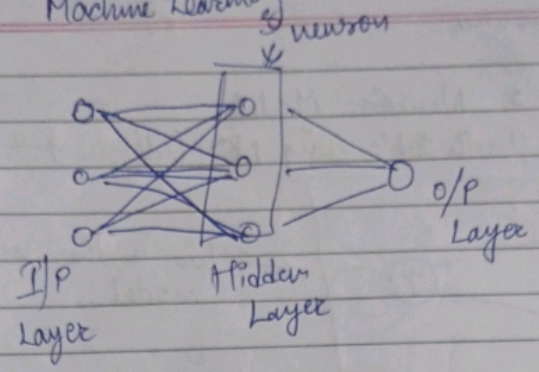
- So horizontal way technologies are there & domains are there vertically. If we do this in reverse this says every time technologys are went to grow.

- So we need to train the data that is the initial set of data to help program in order to understand & how to apply technologies which are associated with programming languages, & algorithms, neural networks to produce sophisticated results.

★ 27th Jan 2022 - ~~Approxim~~ Approximation:-

~~→ Approximation when the form of function is not known and~~
→ Some cases the form of function is not known & in some cases its known.
→ Approximation - when the function is unknown

Machine Learning 29/3/2022



3-5 morethan 3-5 neuron :- network goes to complex network

Deep learning :- Which complexity has complex network

- Learning Rate of the network (τ) ($\alpha = 0.3$) (~~value of~~ eg)?
- weight :- helps to build the system
- Bias - helps to strengthen the weights ; denoted by b.

Eg :-

S = {0.5, -0, +0, 0, 5, -5, +7}

S = {0, 0, 0, 1, 0, 1}

f(0, 1) [output in the form of]

Such function is known as Activation function

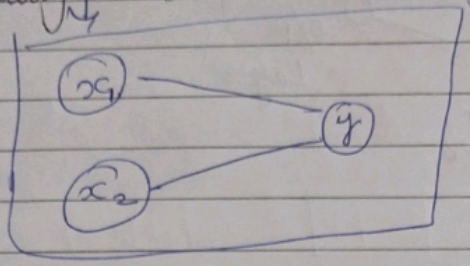
Ans :- If ($y_{in} \leq 0$) return 0
else

If the activation function is written by User function

Mcc ← Pronounce.

★ Mc Culloche Pitts Neuron Model.

[Generating AND Truth Table using MC Culloche Pitts Neuron Model]



← This is the main model.

Truth Table or AND Gate.

	x_1	x_2	y
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	1

$j = 1 \rightarrow 4$

★ Multiplication of two binary numbers

0 . 0 = 0

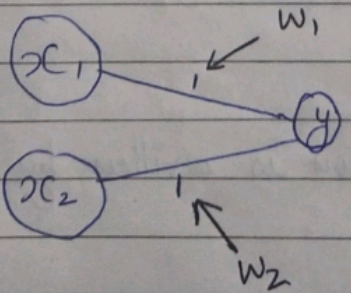
0 . 1 = 0

1 . 0 = 0

1 . 1 = 1

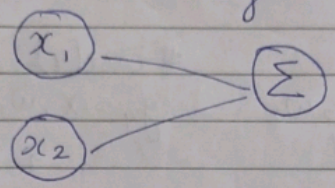
T → T

Weight = $w_1 = 1$ & $w_2 = 1$

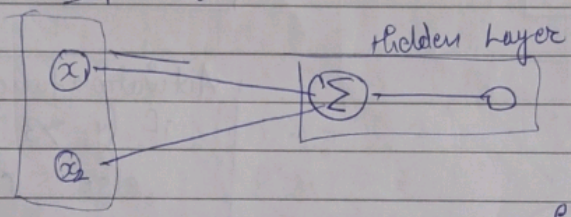


classmate
Date
Page 11

$x_1 * w_1$
$x_2 * w_2$ | $= (x_1 * w_1) + (x_2 * w_2)$
$= \sum_{i=1} x_i w_i$
$= y$



Input Layer



$e = t - a$

$y_1 = x_1 w_1 + x_2 w_2$
$= 0.1 + 0.1$
$= 0 + 0$
$y_1 = f(y_1) = f(0) = 0$ — Compiment

$e =$
$t =$ Targeted
$a =$

- If in system there is an error exists we are computating accuracy
- If there is no error then we are computating ~~Acc~~ precission

#case2 $y_2 = x_1 w_1 + x_2 w_2$
$= 0.1 + 1$
$= 6 + 1$
$= 1$

$f(1) = 0$

If ($x_1 = 1$ & $x_2 = 1$) return 1
else

case(3)

$y_3 = x_1 w_1 + x_2 w_2$
$= 1.1 + 0.1$
$= 1 + 0$
$= 1$

$f(1) = 0$

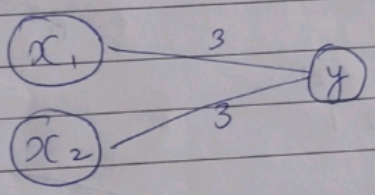
case(4)

$y_4 = x_1 w_1 + x_2 w_2$
$= 1.1 + 1.1$
$= 1 + 1$
$= 2$

$f(2) = 1$

Activation function
if ($y_i > 3$) 1
else 0

Example ②



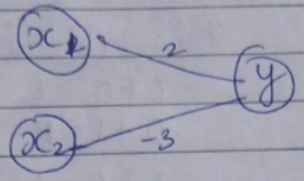
x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Output is generated by using rules of binary multiplication

$$y = \sum_{i=1}^{2} x_i w_i$$

case 1 → $y_1 = (0.3 + 0.3) = 0 + 0 = 0 = 0$
case 2 → $y_2 = (0.3 + 1.3) = 0 + 3 = 3 = 0$
case 3 → $y_3 = (1.3 + 0.3) = 3 + 0 = 3 = 0$
case 4 → $y_4 = (1.3 + 1.3) = 3 + 3 = 6 = 1$

Example 3



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Output is granted by using rules of binary multiplication

Case 1 → y_1 = (0.2 + 0.2) = 0 = 0
Case 2 → y_2 = (0.2 + 1.2) = 2 = 2
Case 3 → y_3 = (1.2 + 0.2) = 2 + 0 = 2
Case 4 → y_4 = (1.2 + 1.2) = 2 + 2 = 4

Case 1 → y_1 = [0.2 + 0.-3] = 0 - 0 = 0
Case 2 → y_2 = [0.2 + 1.(-3)] = 0 - 3 = -3
Case 3 → y_3 = [1.2 + 0.(-3)] = 2 - 0 = 2
Case 4 → y_4 = [1.2 + 1*-3] = 2 - 3 = -1

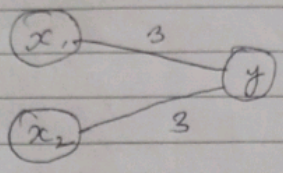
Activation function :-

if (y_i > 3) 1
else 0

29/03/2022 - Tuesday!

★ Generate OR Truth Table using Mc Culloche Pitts Neuron Model:-

#1



Binary Addition Rules

0 + 0 = 0
0 + 1 = 1
1 + 0 = 1
1 + 1 = 0 with carry 1

x_1	x_2	$y = x_1 + x_2$
0	0	0
0	1	1
1	0	1
1	1	1

OR → If one of the ~~input/p~~ I/P signal is high then you will get high(1) of p [$w_1 = w_2 = 3$]

$$y = \sum_{i=3}^{2} x_i w_i$$

Case 1 → $y_1 = (0.3 + 0.3) = 0 + 0 = 0$
Case 2 → $y_2 = (0.3 + 1.3) = 0 + 3 = 3$
Case 3 → $y_3 = (1.3 + 0.3) = 3 + 0 = 3$
Case 4 → $y_4 = (1.3 + 1.3) = 3 + 3 = 6$

Threshold = 6

[if $y_i \geq 3$, return 1
else
return 0]

Realize

~~Resizing~~ NOT function using Mc Culloch Pitts Neuron Model:-

X	Y
0	1
1	0

Fourth table

Not Gate

(x) ——— (y)

Case 1 → $y_1 = (0.1 + 1(0.1)) = 0 + 1 = 1$

Case 2 → $y_2 = (1.1 + 0.1) = 1 + 0 = 1$

Threshold = 1

Shouldn't have two inputs in NOT !!

If $y_i \geq 1$, return 1
else
return 0

Case 1 → $y_1 = 0.1 = 0$

Case 2 → $y_2 = 1.1 = 1$

If $(y_i \geq 1)$ return 0
else 1

REALIZE output

Realize the ~~[illegible]~~ of AND NOT function using Mc Culloch Pitts Neuron Model

$w_1 = 1$ & $w_2 = -1$

AND Truth Table:-

X	Y	$y_i = x_1 * x_2$
0	0	0
0	1	0
1	0	0
1	1	1

AND function

case 1 ⇒ $y_1 = [0.1 + (1)(-1)] = 0 - 1 = -1$

case 2 ⇒ $y_2 = [0.1 + (1)(-1)] = 0 - 1 = -1$

case 3 ⇒ $y_3 = [1.1 + (0)(-1)] = 1 - 0 = 1$

case 4 ⇒ $y_4 = [1.1 + 1(-1)] = 1 - 1 = 0$

Threshold = 1

* NOT Truth Table

x	y
0	1
1	0

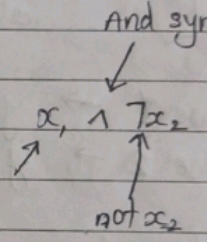
ANDNOT [Correct Answer from Sir]

And Truth Table

x_1	x_2	y	Y
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

Y = not gate

x_1	x_2	$\neg x_2$	$x_1 \wedge \neg x_2$
0	0	1	0
0	1	0	0
1	0	1	1
1	1	0	0



$w_1 = 1$ & $w_2 = -1$

Next page done again

Case 1 $\Rightarrow$ $y_i = x_i w_i$ $\Rightarrow$ $y_i = [(0.1) + (1.-1)] = 0-1 = -1$

$\Rightarrow$ $y_2 = [(0.0) + (0.-1)] = 0-0 = 0$

$\Rightarrow$ $y_3 = [(1.1) + (1.-1)] = 1-1 = 0$

$\Rightarrow$ $y_4 = [(1.1) + (0.-1)] = 1-0 = 1$

$y_i = 0$
$y_i = f(y_i)$
$y_j = f(0)$
$y_j = 0$

$y_1 = 0.0 + 1.(-1)$
$y_1 = 0 - 1$
$= -1$
$= f(-1)$
$= 0$

same for y_2 & y_4.

But in Table :- [On board by sir]

x_1	x_2	$\overline{x_2}$	$x_1 \wedge \overline{x_2}$	actual o/p	e = t - a
0	0	1	0	0	0
0	1	0	0	0	0
1	0	1	1	0	1
1	1	0	0	1	-1

☆ for the following noisy versions of the ____ pattern, Identify the response of network by segregating ~~the input~~ into ____

$w_1 = 0$, $w_2 = -2$, $w_3 = 2$

vector			$x_1 w_1 + x_2 w_2 + x_3 w_3$	Response
x_1	x_2	x_3		
0	-1	0	$0 + (-1)(-2) + 0 = 2$	Correct
0	-1	1	$0 + (-1)(-2) + (1)(2) = 4$	Correct
0	1	-1	$0 + (1)(-2) + (-1)(2) = -4$	Incorrect
0	0	1	$0 - 0 + 2 = 2$	Correct
0	0	-1	$0 - 0 - 2 = -2$	Incorrect
0	1	0	$0 - 2 + 0 = -2$	Incorrect
1	0	1	$0 - 2 + 2 = 0$	Infi
1	0	-1	$0 - 0 - 2 = -2$	Incorrect
1	-1	0	$0 + 2 + 0 = 2$	Correct
1	0	0	$0 - 0 + 0 = 0$	Infinite
1	1	0	$0 - 2 + 0 = -2$	Incorrect
1	1	1	$0 - 2 + 2 = 0$	Infinite

$$\begin{bmatrix} w_1 = 0 \\ w_2 = -2 \\ w_3 = 2 \end{bmatrix}$$

if $(x_1w_1 + x_2w_2 + x_3w_3) > 0$) → Correct =
< 0) → Incorrect = 5
$= 0$) → Indefinite = 2

31st March 2022 [Discussion Lecture about Case Study]

5 marks

1. Case study - Any domain :- Agriculture, Finance, Insurance, Medical, Corporate, Health etc
2. Define its problem (Any problem)

Agriculture

For eg :- Optimum allocation of water for the consumer
- Define the problem (3-4 lines)
- Propose the solution : BPD (Business Process Diagram)

info + diagram → format

- Actor
- Use cases (continuous Activity)
- 0 - 0 (one to one)
- 0 - M (One to Many)
- M - M (Many to many)

PRN 36. pdf

~~Title~~ Title

3. Articulate the problem :- (Synthesize a problem) (Having knowledge about problem)

Water for consumer

↓

1) Farmer - farming
2) Household - family, Home, Bunglow
3) Industry

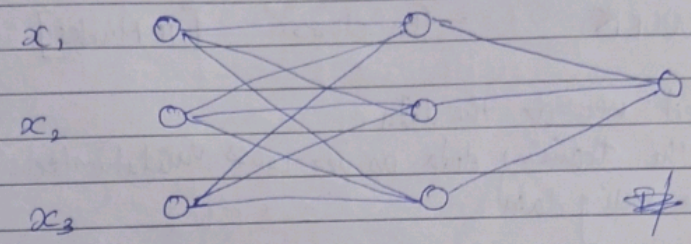
4) Electricity Generation

☆ LAB WORK : In class [10 Marks]

1. To collect ~~new~~ the data
2. Decide the training data as per cross validation techniques
3. Decide the testing data.
4. Accuracy.

PAGE No. 20
DATE / /

★ Perceptron Learning Rule:-



I/P layer Hidden Layer output Layer

Input / process layer.

★ Perceptron Learning Rule :— (meaning)
To strengthen input system by adding bias

- Output of perceptron can be in binary number.
- Weights are from 0 to 1 :- 0.1, 0.3, 0.8, 0
-

★ Supervised vs Unsupervised :- (difference)
- Supervised needs a trainer or teacher
- Unsupervised doesn't need a trainer

★ Decision matters
1 — Vgood
0 — Vbad

★ Delta Learning Rule :-

$\Delta w = w + \delta w$ $c = t - a$ now $\Delta w = w + \delta w$

$= 1 + 0.3$ $= 1 - 0.7$ $= 1 + 0.3$

$= 1.3$ $c = 0.3$ $\Delta w = 1.3$

old w new w

Shall change in weight with respect to error.

Ans) Combination of Red :- {R, B,

Probability of Red balls = 0.3
Probability of Black balls = 0.3
Probability of Yellow balls = 0.2
Probability of Green balls = 0.2

Unity of all the four balls will be 0.3 + 0.3 + 0.2 + 0.2
= 1.

- ☆ Dimensions

- Navye Bias Classifier depends on Dimensions
- Dimension is nothing but some sort of entity ~~or~~ which holds set of features or columns or framework or attributes.
- 1 Dimension can have n no. of attributes.
- In dimension, no. of ~~dim~~ attributes or features should be more or less? which one is much better?
  → Less (to get rid of ~~dom~~ complexity of data that will be stored)

2003012i036

- ☆ Size Database

Lakh x 128 ← No. of records with respect to the record.

IDE 3
Binary Search Tree

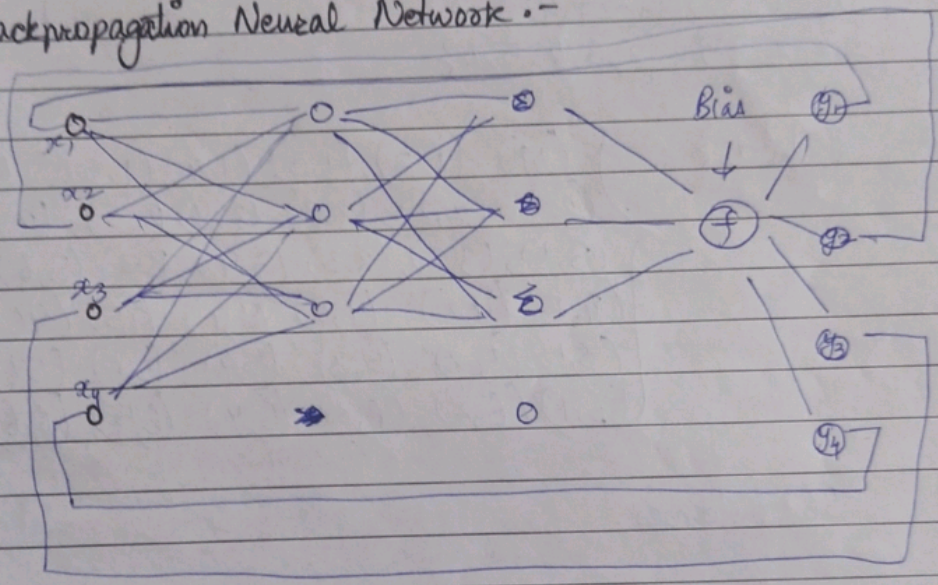
No: of Records ~~of~~

- ☆ High-D [High Dimension]

★ Random Computation :-

★ Perceptions :-
- Single Layer Perception
- Multilayer Perception
- O/P layer Perception.

★ Backpropagation Neural Network :-



★ # 7th April 2022

★ Naive Bayes Classifier and Classification Algorithm (or Classification)
a) Structured data → csv, .mdb
b) Unstructured data → .txt
c) Semistructured data → tags, .html, .XML

[Probability & Statistic Stuff]

50 25%
100
4/1 0.25%

PAGE NO. 23
DATE / /

★ Probability :-
- Dice Example
- 52 Cards Example

★ Dice Example :-

dice 1 = {1, 2, 3, 4, 5, 6}
dice 2 = {1, 2, 3, 4, 5, 6}

$\{d_1, d_2\}$ = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

36 outcomes.

★ Balls = {$\underbrace{R_1, R_2, R_3}_{\text{Red}}, \underbrace{B_1, B_2, B_3}_{\text{Black}}, \underbrace{Y_1, Y_2}_{\text{Yellow}}, \underbrace{G_1, G_2}_{\text{Green}}$}

Calculate the probability of each color and show that total probability is alway unity :-

→

PAGE No.
DATE / /

☆ Class Test
Name :- Harshika Jadhav
PRN :- 20030121036
Batch 2020-23
Machine Learning Group-1

Φ Discuss any 10 Machine Learning Applications?

- Before we move forward to these applications, lets first know what machine learning basically is? Machine learning is a data analysis model. It is a branch of practical intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

Here are the 10 most used Machine learning Applications

1. Speech Recognition :-
- While using Google, we get the "Voice Search" option, which comes under speech recognition, and is a popular machine learning application.
- Speech recognition is the process of converting voice commands into text, also known as "Speech into text", or "Computer attention to speech.". Currently machine learning algorithms are widely used in various speech recognition applications. Google Assistant, Siri, Cortana, and Alexa, use speech recognition technology to follow more voice instructions

2. Products Recommendations :-
- Machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix, etc., to ~~recom~~ recommend the product to the user. Whenever

we search for a particular product on Amazon, we start to find an ad for the same product while the internet is accessing the same browser and this is due to machine learning.

- Google understands user interest using a variety of machine learning algorithms and purposes the product as per the customers interest.
- Similarly, when we use Netflix, we get some recommendations for entertainment series, movies, etc., and this is done with the help of machine learning.

3. Image Recognition :-

- Image recognition is one of the most common machine learning applications. Used to identify objects, people, places, digital images, etc. The most popular use case for photo recognition and face detection is, Suggestion to tag a friend automatically.
- Facebook gives us a feature to suggest tagging a friend automatically. Whenever we upload a photo with our facebook friends, we automatically receive a branded tag proposal, and the technology for this face-to-face & face recognition & visual algorithm.
- It is based on a Facebook project called "DeepFace", which is responsible for facial recognition & photo identifications.

4. Traffic Prediction :-

- If we want to visit a new place, we take the help of Google Maps, which shows us the right route with a short route & predicts traffic conditions.

- Predicts traffic conditions such as whether traffic is cleared, slow, or overcrowded with the help of two methods:
  - Realtime car location builds Google Map app & sensors.
  - The average time it took was past days at the same time.
- ~~The average time it took was past~~ do
- Everyone who uses Google Maps helps this app improve. It takes information from the user and returns it to his or her website to improve performance.

5. Self Driving Cars :-
- One of the most enjoyable programs for automotive learning is self driving cars. Machine learning plays an important role in self driving. Tesla, a well known car-manufacturer, specializes in self-driving cars. It uses an unregulated learning method to train car models to find people & things while driving.

6. Email Spam & Malware filtering :-
- Whenever we receive a new email, it is automatically filtered as important, normal, and spam. We regularly receive important emails in our inbox with important markers & spam emails in our inbox, and the technology for this machine learning.

7. Stock Market Trading :-
- Machine learning is widely used in stock market trading. In the stock market, there is always a risk of rising and falling stocks, so in this short-term machine learning machine

the neural memory is used to predict trends in the stock market.

8. Medical Diagnosis :-
- In medical science, machine learning is used to diagnose disease. With this, medical technology is growing very fast & it is able to create 3D models that can predict the exact location of lesions in the brain.
- It helps to detect brain tumors and other brain-related diseases early.

9. Virtual Personal Assistant :-
- We have various personal assistants such as Google Assistant, Alexa, Cortana, Siri. As the name suggests, they help us acquire knowledge through our voice commands. These assistants can help us in a variety of ways with our voice instructions such as play music, call someone, open email, schedule appointments, etc.
- These visual assistants use machine learning algorithms as an integral part.
- These assistants record our voice commands, send them to the server in the cloud, and extract them using ML algorithms and perform accordingly.

10. Online Fraud Detection :-
- Machine learning makes online activities safer and more secure from fraud. Whenever we do something online, there may be various ways in which fraudulent activity can take place, such as fake accounts, fake IDs, and money

laundering. So to find out, the feed forward Neural network assists us to determining whether if is a real job or a fraudulent activity.

- for each actual transaction, the output is converted to specific hash values, and these values are the input of the next round. for every ~~do~~ real deal, there is a pattern that undergoes a change in fraudulent activity which is why it recognizes & more online activities much safer.

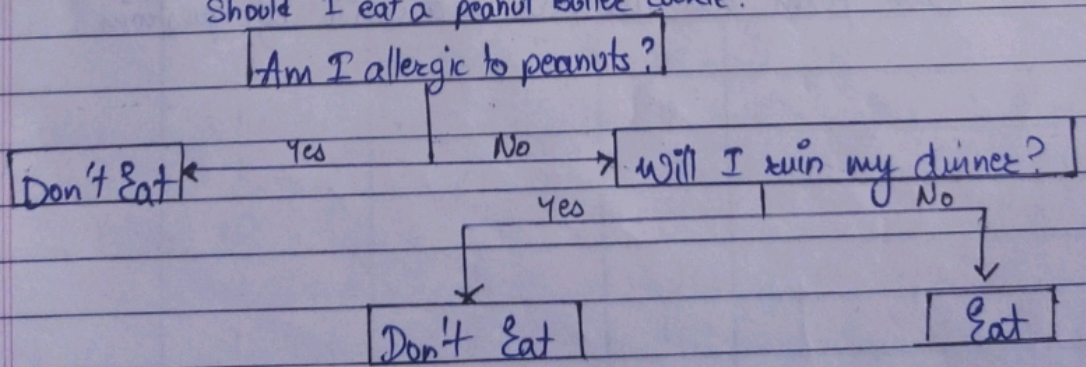
PAGE No. 1
DATE / /

Name :- Harshika Avinash Jadhav Batch 2020-23
PRN :- 20030121036 Machine Learning Group I

Q Give the demonstration of Entropy and Information gain to draw the decision tree. Assume a sample data set.

- Before building a decision tree algorithm the first step is to answer this question, for this we need to understand the way to apply a few key concepts from information theory.
- Lets examine this method by ~~toots~~ taking the following steps :-
  1. Take a very brief look at what a Decision Tree is.
  2. Define and examine the formula for Entropy.
  3. Discuss what a Bit is in information theory.
  4. Define Information Gain and use entropy to calculate it.
  5. Write some basic Python functions using the above concepts.

- Decision Tree : ~~Dess~~ Decision tree algorithm is a supervised learning algorithm for partiton or regression problems. Our ultimate goal is to use historical data to predict the outcome. Unlike linear regression, decision tree can pick up indirect interactions between variables in data.



- In this example, the decision tree may take note of the fact that you should eat a cookie only if certain conditions are met. This is the ultimate goal of the decision tree. We want to continue to make decisions (divisions) until certain conditions are met. Once we are together we can use it to divide or make a prediction. This model is very basic using only two variables (allergies, wasting dinner). But, if you have a database with thousands of variables/columns how do you determine which variables/columns work best to seperate them? A popular way to solve this problem, especially if you are using the ~~ID3~~ ID3 algorithm, is to use entropy and information gain.

- The Task

Suppose we had some data and wanted to use it to create online questions that predicted something about the interviewee. After looking at the relationship to the data we decided to use the decision tree algorithm. The goal of the questionaire will be ~~the~~ to speculate that the respondent is some one of the questions. We have a small fiction database with fifteen entries. Each form has answers to a series of questions.

name	age	apple_pie?	potato_salad?	sushi?	midwest?
John	32	0	1	1	1
Jonnah	25	1	1	0	1
Jeff	33	1	1	0	1
Erik	26	0	0	1	0
Lisa	30	1	1	1	1
Rose	25	1	0	0	1
Lily	42	1	1	0	1
Oscar	38	1	1	0	1
Mike	36	1	1	0	0
Sarah	29	1	0	1	0
David	35	1	0	0	1
Nolan	28	1	1	1	0
Joe	20	0	1	0	1
Andrew	38	1	0	0	1
Antonio	31	0	0	1	0

- Entropy.

To get started we will use information called entropy. In data science, entropy is used as a measure of how a column is "mixed". Specifically, entropy is used to measure disruption.

midwest?
1
1
1
0
1
1
1
1

0
0
1
0
1
1
0

our target column, "midwest?"

There are ten people living in the middle west and five non-residents. If someone were to ask you how mixed a column is, you might say it is mixed with most (2/3) people from the middle west, Entropy gives us a way to measure the response to a "mixed type". When (1) and 10)s are very mixed in a column, the entropy increases. If "in the middle of the west?" have equal numbers of (1)s and (0)s our entropy will be 1. If "in the middle of the west?" only includes (1)s entropy will be 0.

We can use the following formula to calculate entropy:-

$$-\sum_{i=1}^{c} p(x_i) \log_b p(x_i)$$

the formula for entropy

Lets calculate the entropy for the "midwest?" column

1. We need to iterate through each unique value in a single column and assign it to i. For this example, we have 2 cases(c) in the "midwest?" column, either (0) or (1).

2. We then compute the probability of that value occurring in the data. For the case of (1), the probability is 10/15. For the case of (0), the probability is 5/15.

3. We take the probability of each case and multiply it by the logarithm base 2 of the probability. 2 is the most common base because entropy is measured in bits

(more on that later). The full explanation of why 2 is used is out of the scope of this post, but a user on stack exchange offers a good explanation. For this case of (1), we get $10/15 * \log 2(10/15)$. For the case of (0), we get $5/15 * \log 2(5/15)$.

4. Next, we take our product from each case above and sum it together. For this example, $10/15 * \log 2(10/15) + 5/15 * \log 2(5/15)$

5. Finally, we negate the total sum from above, $(10/15 * \log 2(10/15) + 5/15 * \log 2(5/15))$.

Once we put all the steps all together we get :-

$$-(10/15 . \log_2(10/15) + 5/15 . \log_2(5/15))$$
$$-(-.389975 + -.528308)$$
$$-(-.918278)$$
$$.918278$$

Our final entropy is .918278

- Use entropy to make Decisions :-
- Our goal is to find the best / most diverse columns on which to build a decision tree. Finally, we want to keep separating variables/columns until our mixed target column is no longer mixed
- For example, let's look at the entropy of "midwest?" column after we split our database into "potato_salad?" column.

~~Spl~~ ~~pf~~ splitting "potato_salad?"

name	midwest?
John	1
Jonnah	1
Jeff	1
Lisa	1
Rose	1
Lily	1
Oscar	1
David	1
Joe	1
Andrew	1

name	midwest?
Erik	0
Mike	0
Sarah	0
Nolan	0
Antonio	0

The above dataset is split in two sections. On the left side, everyone who likes potato salad. On the right side everyone who doesn't. We fill focus on the left side which now has seven people from the midwest and two people who aren't. By using the formula for ~~#~~ entropy on the left split midwest column the new entropy is .764204. ~~#~~ Our goal is to lower the entropy and we went from .918278 to .764204. If we look at the right column our entropy went up as there are an equal amount of (1)s and (0)s. The formula for information gain will change the entropy on both sides.

- Information Gain :-

- Earlier we established we want splits that lower the entropy of our target column. When we split on "potato-salad?" we saw the entropy in the "midwest?" went down on the left side. Now we need to understand the total entropy lowered when we look at both sides of the split.

Information gain will use the following formula :-

$$IG(T, A) = \text{Entropy}(T) - \sum_{v \in A} \frac{|T_v|}{T} \cdot \text{Entropy}(T_v)$$

T = Target, our "midwest?" column
A = the variable (column), "potato-salad?"
v = each value in A, each value in the "potato-salad?" column.

1. First, we'll calculate the original entropy for (T) before the split, 918278.
2. Then, for each unique value (v) in variable (A), we compute the number of rows in which (A) takes on the value (v), & divide it by the total number of rows. For the "potato-salad?" column we get 9/15 for the unique value of (1) & 6/15 for the unique value of (0).
3. Next, we multiply the results by the entropy of the rows where (A) is (v). For the left split (split on 1 for "potato-salad?") we get 9/15 * .764204. For the right side of the split, we get 6/15 * 1.

4.

4. We add all of these subset products together,

$9/14 * .764204 + 6/15 = .8585224$

5. We then subtract from the overall entropy to get information gain $.918278 - .8585224 = .059754$

Our information gain = ~~0.059~~ .059754