ASSIGNMENT - 01

1. Explain types and subtypes of machine learning with examples.

Types of machine learning -

1. Supervised learning : A training set of examples with the correct responses (targets) are provided and, based on this training set, the algorithm generalizes to respond correctly to all possible inputs. This is called learning from examples.

Types of supervised learning -

i) Regression : This is a type of problem where continuous response values are predicted.

example : 1. predicting the price of house in a city.
2. value of stock.

ii) Classification : This is a type of problem where we predict the categorical response values where the data can be separated into specific "classes".

example : 1. a mail is spam or not.
2. will it rain today or not.

2. Unsupervised learning : Correct responses are not provided, instead the algorithm tries to identify similarities between the inputs that have something in common are categorized together.
Then, statistical approach to unsupervised learning is known as density estimation.

i) Clustering : In this type of problem, similar things are grouped together.

example : given news articles can be clustered into different types of news.

3. Reinforcement learning : This is somewhere between supervised and unsupervised learning.
The algorithm gets told when the answer is wrong, but does not get told how to correct it.
It has to explore and try out different possibilities until it works out to get the answer right.

2. Explain entropy, information gain and gini index with their formula. Also define their role in constructing decision tree.

1. Entropy: Entropy is a ~~me~~ commonly used measure in information theory, that characterizes the (im)purity of an arbitrary collection of examples.

- Given a collection S, containing positive & negative examples of some target concept, the entropy S relative to this classification is,

- $$\text{Entropy}(S) = -p_{\oplus}\log_2 p_{\oplus} - p_{\ominus}\log_2 p_{\ominus}$$

$p_{\oplus} \rightarrow$ proportion of positive examples in S
$p_{\ominus} \rightarrow$ negative examples in S.

- One interpretation of entropy from information theory is that it specifies the minimum number of bits of information needed to encode classification of an arbitrary member of S.

- $$\text{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$, in general.

(c-wise classification).

2. **Information gain:**

Information gain is simply the expected reduction in entropy caused by partitioning the examples according to this attribute.

The information gain, gain (S,A) of an attribute A, relative to a collection of examples S, is,

$$\text{gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v).$$

where,

values(A) → set of all possible values for attribute A.

S_v → subset of S for which A has value v.

3. **Gini index:**

The impurity (or purity) measure using in building decision tree in classification and Regression tree (CART) is Gini index.

$$\text{Gini}(p) = 1 - \sum_{i=1}^{j} p_i^2$$

$i \in \{1, 2, \ldots j\}$

p_i → fraction of items labeled with class i in the set.

3. How logistic regression is a classification technique? What is the significance of sigmoid function?

- Logistic regression is one of the basic and ~~p~~ algorithm of solve a classification problem.
- A problem is identified as classification algorithm when independent variables are continuous in nature and dependent variables are in categorical form i.e in classes like positive and negative class
- Unlike linear regression, which outputs continuous number values, logistic regression transforms it's output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes.

Sigmoid function-

In order to map predicted values to probabilities, we use the sigmoid function. The function maps any real value into another value between 0 and 1.

$$S(z) = \frac{1}{1+e^{-z}}$$

It gives the output as a conditional probabilities of the predictions.

MANJARA CHARITABLE TRUST
RAJIV GANDHI INSTITUTE OF TECHNOLOGY, MUMBAI
JUHU VERSOVA LINK ROAD, VERSOVA, ANDHERI (WEST), MUMBAI - 53.

4. Perform linear regression using least square method on the following tes data. If X=16, find Y.

X(input)	Y(output)	X^2	XY
3	25	9	925
6	33	36	198
8	37	64	296
12	45	144	540
15	53	225	795
20	57	400	1140
22	67	484	1474
86	317	1362	5368

$n=7$

By least square method,

$$b1 = \frac{\Sigma XY - (\Sigma X \Sigma Y)/N}{\Sigma X^2 - (\Sigma X)^2/N}$$

$$b1 = \frac{5368 - (86 \times 317)/7}{1362 - (86)^2/7}$$

$$b1 = 4.824$$

$$b0 = \frac{\Sigma Y - b1\,\Sigma X}{n} = \frac{317 - 4.824 \times 86}{7}$$

$$b0 = -13.98$$

$$y = -13.98 + 4.824x$$

when $x = 16$, $y = -13.98 + 4.824 \times 16$

$$y = 63.14$$

5. Apply logistic regression algorithm to classify following data. Also find accuracy.

X1	X2	Y
1.64	2.63	0
3.4	4.1	0
7.2	2.7	1
6.5	1.8	1
7.6	3.5	1

$b0 = -0.4$
$b1 = 0.8$
$b2 = -1.1$

i) For $x1 = 1.64, \; x2 = 2.63$

$$y = \frac{1}{1 + e^{-(b0 + b1X1 + b2X2)}}$$

$$y = 0.1212$$

ii) For $x1 = 3.4, \; x2 = 4.1$

$$y = \frac{1}{1 + e^{-(0.4 + 0.8 \times 3.4 - 1.1 \times 4.1)}}$$

$$y = 0.10$$

iii) For $x1 = 7.2, \; x2 = 2.7$

$$y = 0.916$$

iv) For $x1 = 6.5, \; x2 = 1.8$

$$y = 0.944$$

v) For $x1 = 7.6, \; x2 = 3.5$

$$y = 0.861$$

Y	class
0.1212	0
0.10	0
0.916	1
0.944	1
0.861	1

Accuracy = (correct prediction/no of predictions) * 100
= (5/5) * 100
= 100 %

6. Consider following data. Find root element of the tree using ID3 algorithm.

No.	FORM	COLOR	SIZE	CLASS
1.	circle	red	small	+
2.	circle	red	big	+
3.	triangle	yellow	small	−
4.	triangle	red	big	−
5.	circle	yellow	small	−
6.	circle	yellow	big	−

no of + : 2 , no. of − : 4 , total = 6

$E(table) = -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}$

=

for attribute form,

~~circle : total = 2, + = 2, - : 0~~

~~red : total = 2, + =~~

triangle : total = 2, + : 0, - : 2

circle : total = 4, + : 2, - : 2

$$E(s = circle) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}$$

$=$

$$E(s = triangle) = -1\log_2 1$$

$=$

$$I(form) = \frac{4}{6}\times \quad + \frac{2}{6}\times$$

$=$

Gain $= E(S) - I(S)$

$=$

for attribute colour,

red (3) : + : 2, - : 1

yellow(3) : + : 0, - : 3

$$E(colour = red) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$$

$=$

$$E(colour = yellow) = -\frac{3}{3}\log_2\frac{3}{3}$$

$=$

Scanned by CamScanner

$I(colour) = \frac{3}{6} \times \quad + \frac{3}{6} \times$

$=$

$gain(colour) =$

$=$

For attribute size,

small (3) : + : 1 , − : 2
big (3) : + : 1 , − : 2

$E(size = small) = -\frac{1}{3} \times \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}$

$=$

$E(size = big) = -\frac{1}{3} \times \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}$

$=$

$I(size) = \frac{3}{6} \times \quad + \frac{3}{6} \times$

$=$

~~E(siz~~ gain(size) =

$=$

7) Consider following data, using CART algorithm find gini index of target or class attribute, also find gini index of Reputation using splitting subset. (risk = target /class)

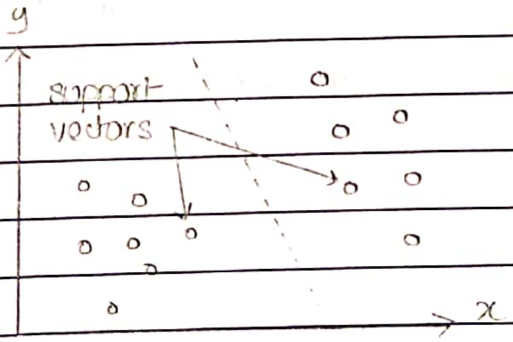
No.	security	salary	Debt	reputation	risk
1.	none	\$0 to 30K	high	bad	high
2.	none	30 to 60K	high	unknown	high
3.	none	30 to 60K	low	unknown	moderate
4.	none	0 to 30K	low	unknown	high
5.	none	over 60K	low	unknown	low
6.	adequate	over 60K	low	unknown	low
7.	none	0 to 30K	low	bad	high
8.	adequate	over 60K	low	bad	moderate
9.	none	over 60K	low	good	low
10.	adequate	over 60K	high	good	low
11.	none	0 to 30K	high	good	high
12.	none	30 to 60K	high	good	moderate
13.	none	over 60K	high	good	low
14.	none	30 to 60K	high	bad	high

8. Define Support vector Machine, hyperplane, margin and support vectors with suitable diagram.

Support Vector Machines: In machine learning, SVMs are supperuised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

1. - Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary classification.

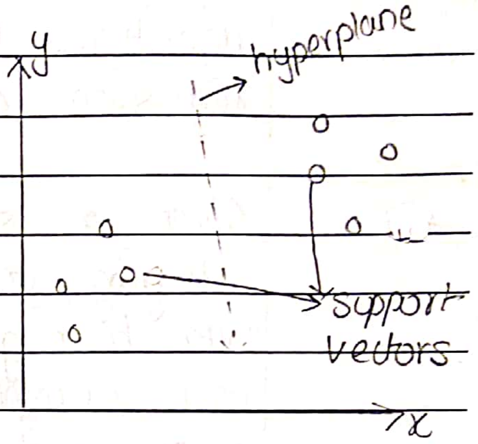
- A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap.



Hyperplane :

- The optimum hyperplane is the linear classifier with the maximum margin for a given finite set of learning pattern.

Margin :



- A margin is a separation of line to the closest class points.
- A good margin is one where there is this separation is larger for both the classes. It allows the points to be in their respective classes without crossing to other class.

Support vectors :

- The vectors that define the hyperplane are the support vectors.
- The extreme points in the data sets that define the hyperplane.