KAMMAVARI SANGHAM (R) - 1952

K S GROUP OF INSTITUTIONS

K.S. INSTITUTE OF TECHNOLOGY

Accredited by NAAC

(Approved by AICTE & Affiliated to VTU)

#14, Raghuvanahalli, Kanakapura Road, Bengaluru - 560109

BLUE BOOK

Name of the student: ANUSHA PHANIRAN

Sem / Section: 6th sem **Branch:** AIML

USN: 1 K S 2 0 A I 0 0 4

SUBJECT: MACHINE LEARNING **SUBJECT CODE :** 18AI61

MAXIMUM MARKS

Test	I	II	III	Average Marks Obtained
Date	17.04.2023	5.06.2023	06.07.2023	Test 26
Marks Obtained	30	29	17	Assignment 10
Signature of Student	Anusha	Anusha	Anusha	Activity
Initials of Faculty				Total 36
		Final Assessment		36

NAME OF FACULTY : Ana Matheews

SIGNATURE OF FACULTY

SIGNATURE OF H.O.D.

K. S. INSTITUTE OF TECHNOLOGY

VISION

"To impart quality technical education with ethical values, employable skills and research to achieve excellence"

MISSION

- ★ To attract and retain highly qualified, experienced & committed faculty.
- ★ To create relevant infrastructure
- ★ Network with industry & premier institutions to encourage emergence of new ideas by providing research & development facilities to strive for academic excellence
- ★ To inculcate the professional & ethical values among young students with employable skills & knowledge acquired to transform the society

K.S. INSTITUTE OF TECHNOLOGY

First Internal Test

Q.No	Marks		Q.No	Marks	CO	CO	Total
1 (a)	6		2 (a)		CO1	CO1 - 18	18
1 (b)	6	OR	2 (b)		CO1		
1 (c)	6		2 (c)		CO1	CO2 - 12	12
3 (a)	6		4 (a)		CO2		
3 (b)	6	OR	4 (b)		CO2		
						Grand Total	30

Second Internal Test

Q.No	Marks		Q.No	Marks	CO	CO	Total
1 (a)	6		2 (a)		CO3	CO3 - 17	17
1 (b)	6	OR	2 (b)		CO3		
1 (c)	5		2 (c)		CO3	CO2 - 6	12
3 (a)			4 (a)	6	CO2	CO4 - 6	
3 (b)		OR	4 (b)	6	CO4		
						Grand Total	29

Third Internal Test

Q.No	Marks		Q.No	Marks	CO	CO	Total
1 (a)	4		2 (a)		CO5	CO5 - 9	9
1 (b)	3	OR	2 (b)		CO5		
1 (c)	2		2 (c)		CO5	CO4 - 8	8
3 (a)	6		4 (a)		CO4		
3 (b)	2	OR	4 (b)		CO4		
						Grand Total	17

Assignment	A1	A2	Total

Activity				Total

SIGNATURE OF FACULTY.

IA - 1

PART-A

1a. In simple words, machine learning is a program that learn from data given to it.

A more engineered definition is, a computer program performs a task T which is measured by performance measure P when performance of task T as measured by P increases with experience E.

Applications of Machine learning:

- Learning to recognize spoken words.

A program called SPHINX can recognize specific speech pattern by analyzing sounds and phrases detected by speech signals. Similar operations can be applied to other use cases.

- Learning to drive unmanned vehicle.

A program called ALVINN was design to drive a car controlled by a computer. It could drive a car at 70 kmph for 90 kms on a public highway.

- To recognize new astronomical structures

In the Palomar Sky Survey observatory, decision tree learning algorithms is applied on large databases of images containing data on astronomical structures. The algorithm finds the general regularities in data to classify new astronomical structures.

- Learning to play backgammon

The best game program for backgammon is

TD-GAMMON, it has learnt its game playing strategies by playing about a million practice games against itself. It can play on level of ~~B~~ expert human players.

1b.

Size	Colour	Shape	Label
Small	Red	Circle	Yes
Big	Red	Circle	No
Small	Red	Triangle	No
Big	Blue	Circle	No
Small	Blue	Circle	Yes.

$h_0 = \langle \phi, \phi, \phi, \phi, \phi \rangle$

$x_1 = \langle$Small, Red, Circle$\rangle$, $h_1 = \langle$Small, Red, Circle$\rangle$

$c(x_2) = 0$ $\therefore$ x_2 is not considered

$c(x_3) = 0$ $\therefore$ x_3 is not considered

$c(x_4) = 0$ $\therefore$ x_4 is not considered

$x_4 = \langle$Small, Blue, Circle$\rangle$, $h_2 = \langle$Small, ?, Circle$\rangle$

$\therefore$ Maximally specific hypothesis is

$\langle$Small, ?, Circle$\rangle$.

1c. Supervised learning: Data for model fitting is labelled. So, the model only has to learn the underlying patterns in each category to be able to generalise new data.

Ex: Spam and ham emails are labelled correctly and the model is trained with labelled data. The model learns the pattern behind spam and ham

emails and then tries to filter a new emails into spam or ham.

Unsupervised learning: The data is not labelled, the model has to first classify the data into different categories ~~and then learn~~. Different categories are established and when new data is fed, the model applies its learning to classify it into appropriate category.

Ex: Spam and ham email data is unlabelled. Model finds frequently used word patterns in subject, sender email ID, body of email in spam mails and classifies them as span.

Some unsupervised learning models are k-means clustering, heirarchical clustering algorithm etc.

Reinforcement learning: Here, the model is given positive rewards for accurate prediction and negative rewards for wrong prediction. The model learns to predict in a manner where it receieves positive rewards.

Some examples of supervised learning are: linear regression, logistic regression, k-nearest neighbors etc.

Classification algorithms predict ~~whether said data~~ in which class the said data belongs to.

Regression algorithms predict a numerical value (Ex: house price) by considering various numerical parameters.

Logistic regression is actually a classification algorithm which also mentions the probability (Ex: 80% probability of image being that of an orange).

PART - B

3a.

i) Data Cleaning.

Often times we find missing or inconsistent val in data. At times like these, we need to decide whether the feature is an important predicto or not. If the feature does not contribute much to the learning process and has lots of missing values, the whole column can be dropped. But if it is important, either the missing values can be filled with mean, median or mode of column data or rows with missing values can be dropped

Ex: In housing price prediction dataset, the total_bedroom column has some null values. It is an important feature while considering house price therefore it cannot be dropped. 20433 rows have values and 207 are null. In a dataset of 20640 rows, 207 rows are about 1% of the dataset and can be dropp or filled with median value of total_bedroom

ii) Handling text and categorical data.

Machine learning models require numerical dat The predictions will may not be accurate in the presen of text or categorical data since those attributes will not be considered. But they can be converted to numerical data by One-hot encoding or Ordin encoding.

In One-hot encoding, suppose the data is:

Red, Blue, Green

It is encoded as Red = [1, 0, 0]

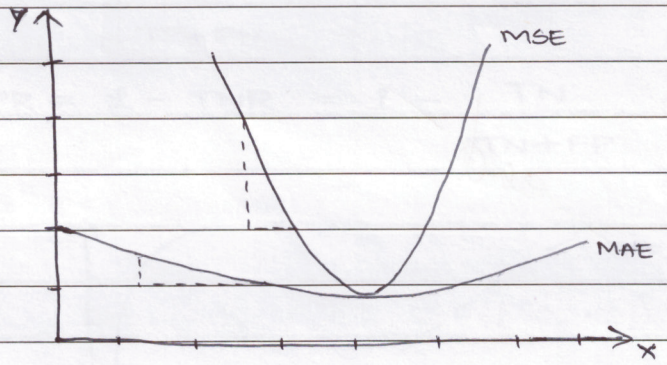
Blue = [0, 1, 0]

Green = [0, 0, 1]

But in Ordinal encoding, it will be encoded as Red = 0, Blue = 1 and Green = 2.

The encoded data is considered for fitting a model.

3b.



MAE (Mean absolute error)

$$MAE = \frac{1}{n} \sum_{i=0}^{n} |x_i - \hat{x}_i|$$

where, $n \rightarrow$ size of test set

$x_i \rightarrow$ predicted value of ith row

$\hat{x}_i \rightarrow$ real value of ith row.

MAE calculates the average of errors between predicted and real values.

This measure is used when we don't want to give much importance to outlier. MAE is an almost linear function and does not provide much weightage to outliers.

RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(x_i - \hat{x}_i)^2}$$

where, $n \rightarrow$ size ~~of test set~~

$x_i \rightarrow$ ~~actual~~ predicted value of ith row

$\hat{x}_i \rightarrow$ actual value of ith row

RMSE calculates the root of average of square ~~of~~ difference between actual and predicted value. This measures is used when outliers need to be considered towards performance of model.

end

IA - 2

PART - B

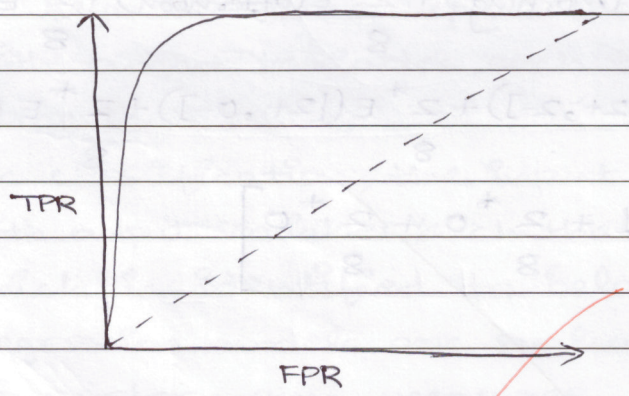
4.

a) ROC (Receiver Operating Characteristics) and AUC (Area under curve) are performance measures for classification models.
ROC is a plot of True Positive Rate vs False Positive Rate i.e Sensitivity vs 1 - Specificity.

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = 1 - TNR = 1 - \frac{TN}{TN+FP} = \frac{FP}{FP+TN}$$

b)



The farther the curve is from the diagonal (more towards top left corner) the better the model is. AUC is a value between 0 to 1. When AUC = 1, it means that the model can perfectly classify input data. When AUC = 0.5, the model is as good as a random guesser. Any value between 0.5 to 1 suggests a good model. AUC is derived from ROC.

Precision-recall curve is used when there are a lot of ~~false~~ positives and ROC, AUC is used when there are a lot of ~~false~~ negatives.

4b. $E(S) = E([6+, 2-]) = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 0.81$

$G(S, \text{Gender}) = 0.81 - \left[\frac{5}{8} * E(\text{female}) + \frac{3}{8} * E(\text{male})\right]$

$= 0.81 - \left[\frac{5}{8} * E([4+, 1-]) + \frac{3}{8} * E([2+, 1-])\right]$

$= 0.81 - \left[\frac{5}{8} * 0.721 + \frac{3}{8} * 0.918\right]$

$= 0.015$

$G(S, \text{slot}) = 0.81 - \left[\frac{4}{8} * E(\text{morning}) + \frac{2}{8} * E(\text{afternoon}) + \frac{2}{8} * E(\text{evening})\right.$

$= 0.81 - \left[\frac{4}{8} * E([2+, 2-]) + \frac{2}{8} * E([2+, 0-]) + \frac{2}{8} * E([2+, 0\right.$

$= 0.81 - \left[\frac{4}{8} * 1 + \frac{2}{8} * 0 + \frac{2}{8} * 0\right]$

$= 0.31$

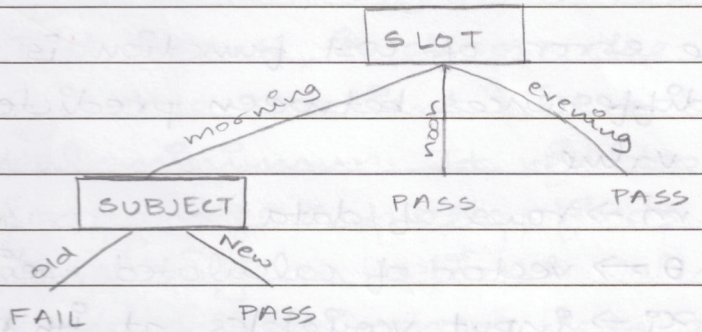
$G(S, \text{subject}) = 0.81 - \left[\frac{3}{8} * E(\text{new}) + \frac{5}{8} * E(\text{old})\right]$

$= 0.81 - \left[\frac{3}{8} * E([3+, 0-]) + \frac{5}{8} * E([3+, 2-])\right]$

$= 0.81 - \left[\frac{3}{8} * 0 + \frac{5}{8} * 0.97\right]$

$= 0.203$

∴ The root node has to be Slot since it has the highest information gain.



PART B – A

1a.

- In Regression, the input variables have a relationship with output variables such that output is a continuous numerical value.
  In classification, the input variables have a relationship with output variables such that output is a discrete value which is identified to belong to a certain class.
- Regression models are evaluated using RMSE, MAE as performance measures.
  Classification models are evaluated using ROC-AUC, precision-recall curve as performance measures.

Vector form of Linear Regression model

$$\hat{y} = \theta^T . x$$

where θ is the vector of calculated weights + bias term
x is the vector of input data
$\hat{y}$ is the predicted output.

MSE cost function for Linear Regression model.

$$MSE(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\theta . x_i - y_i)^2$$

Mean square error of cost function is the mean of squared differences between predicted value and actual value.

Here, m → rows of data

θ → vector of calculated weights

x_i → input variables at i_{th} row

$\theta . x_i$ → predicted output at i_{th} row

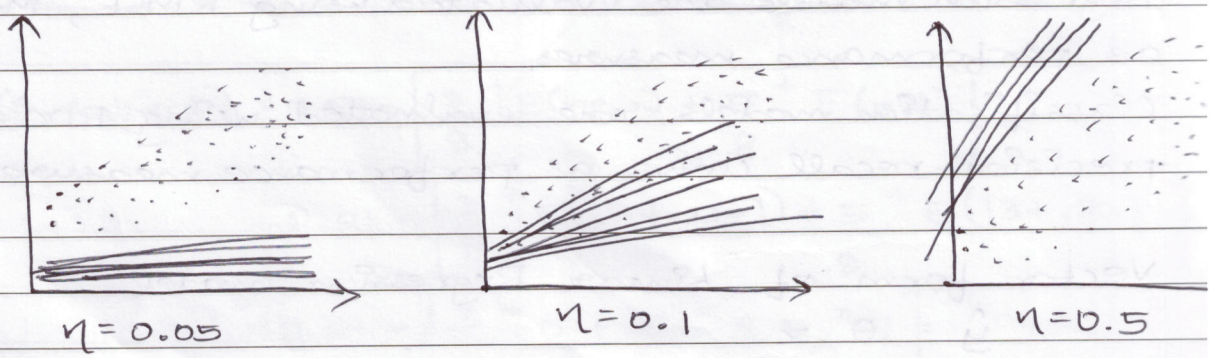
y_i → actual output at i_{th} row.

1b. Gradient descent is an iterative algorithm used to find minimum cost function.

If the learning rate is too low, the algorithm will converge at minimum but take too long.

If the learning rate is too high, the algorithm might completely miss the global minimum and end up on a different slope of the function.

∴ learning rate has to be optimal



1c. • Stochastic Gradient descent randomly picks one instance of training data to calculate next gradien This way it is time efficient and not computation heavy. But because, it choose the next instance randomly, the gradient will not directly find the global minimum. It will bounce around in the region and eventually finds a good value not necessarily an optimal value. SGD is a great solution for datasets where data arrives sequentially

• Mini-Batch Gradient descent randomly picks a subset of training data to calculate next gradien This gives a better result than SGD.

5

sd

IA-3

PART-B

3a. i) Bagging and pasting

In bagging and pasting, the predictors are trained with random subsets of data. The difference lies in how the random subset is dealt with.

In bagging the chosen subset for a predictor is put back and can be chosen again for a different predictor.

In pasting, the subsets used to train predictors are not repeated. Each predictor get a different subset.

Hence, bagging produces more diverse predictors and is less prone to overfitting.

ii) Hard voting and soft voting classifiers.

In hard voting, each predictor predicts a label and the aggregate of all predictions is the final prediction.

In soft-voting, a probability is assigned to each class by a predictor. The probabilities of all classes given by all predictors are aggregated and the class with highest probability is the final prediction