(3)

Mean:-

arithmetic average of a dataset

- adding no. in dataset divided by obser

$$\underset{\text{mean}}{\bar{x}} = \frac{\Sigma x \rightarrow \text{observation}}{N \rightarrow \text{no. of observation}}$$

(2) Mode :-

Most frequently occuring observation or value.

$$\text{mode} = l + \left[\frac{f_m - f_1}{2f_m - f_1 - f_2}\right] \times h$$

f_m = freq. possereb by modal clasr

f_1 - clasr before

f_2 : Uasr after Modal clarr

h = Width

3) Median

Middle number in a dataset When no listed either assending and desending Order.

for odd obs

$$\text{Media} = \left(\frac{n+1}{2}\right)^{th} \text{observation}$$

for even

$$\text{Median} = \frac{(n/2)^{th}\text{obs} + (n/2+1)^{th}\text{obs}}{2}$$

smoothing :→

Process that is used to remove noise from the dataset using some Algorithms

It helps in predicting the Patterns

Aggregation :→

is nothing but data collection is the method of storing and presenting data in a summary format

Generalization !-

It converts low level data attributes to high-level attribute using concept hierarchy.

eg. Age in Numerical form(20, 2 converted into categorical value (Young, old)

Normalization :-

Data Normalization involves converting all data variables into a given range.

Min- Max normalization

Transfer original data linearly.

Z-score normalization:-

In Z Score Normalization the value of an attribute (A), are normalized based on the Mean of A and Its S.D.

an
and data
ly & Commu
for two

hing binary

has a
r not
using
ikelihood
(MLE)

Standard deviation: sqre of variance σ^2

$$s = \sqrt{\frac{\Sigma(\bar{x}-x)^2}{n-1}}$$

x – each value
$\bar{x}$ - Mean
n = no. of value

④ Linear Regression:-
Machine learning algorithm based on supervised learning
It targets prediction valuer on the basis of independent variables

Multivariate Regression:-
it concerns the study of two or more predictor variable

$$y = a + bx + cx_2$$

Mean square Error:-
represents the error of the estimator or predictive model created based on the given set of observation in the sampl

$$MSE = \frac{1}{n}\Sigma(y-\bar{y})^2$$

sq of diff betn
actual & predicted

⑤ logistic regression

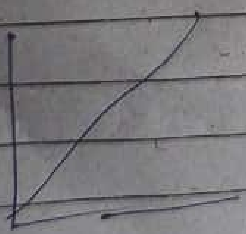
Logistic Regression :-

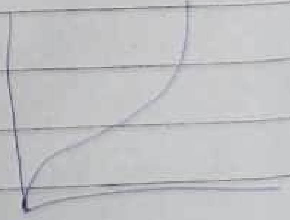
classification techniques are an essential part of Machine learning and data Mining application.

- is one of the most simply & common used machine learning algorithm for two class classification

statistical method for predicting binary class

linear	logistic
continuous o/p	constant o/p
eg. House price	eg patient has a cancer or not
estimated using Ordinary Least Square (OLS]	estimated using Maximum Likelihood Estimation (MLE)





Page No.
Date

Accuracy:- calculated as the number of correctly instance divided by total number of instance.

$$acc = \frac{Tp + TN}{Tp + Fp + TN + FN} = \frac{Tp + TN}{Pos + Neg}$$

errorRate:- calculated as the no. of incorrectly classified instance divided by the total Number of instance

$$error = 1 - acc.$$

precision:-

calculated as the no. of correctly classified +ve instance divided by the total no. of instance Which are predicted positive. it is also called confidence

$$precision = \frac{Tp}{Tp + Fp}$$

Recall:-

$$= \frac{Tp}{Tp + fN}$$

Q. Types of Logistic Regression

Binary Logistic Regression:- The target variable has only two possible outcomes such as spam or Not spam, concer or No cancer.

Multinomial Logistic Regression:- The target variable has three or more nominal categories such as predicting the type of wine.

Ordinal logistic Regression: the target Variable has three or More ordinal categories such as restaurant or product rating from 1to5

Confusion Matrix

it contains information about actual and predicted classification done by classification system.

	Predictel	
actual	TP	FN
	FP	TN

Confusion Matrix

Page No:
Date

(6) Naive. Bayes:- used for classification of categorical data

prior probabilities:-
Calculate for some event based on no other information

$$\text{Bayes Rule} :- P\left(\frac{A}{B}\right) = \frac{P(B/A)\,P(A)}{P(B)}$$

(7) Text Analysis
is a deriving & extending high quality information from text

Pre-processing Method in Text analysis
i. Tokenization.
The process of breaking text paragraph into small chunk or sentence