Page:
Date: / /

LP1

Assignment C3

Data Analysis

Title :- Bigmart Sales Analysis Date of completion :- 20/11/20

Problem Statement :- For data comprising of transaction records of sales store. The data has 8523 rows of 12 variables. predict the sales of a store.

Learning objectives :- 1) Learn regression algorithm
2) Summarize properties of dataset
3) Learn to split the dataset.

Learning Outcomes :- Students will be able to develop a predictive model for sales of an item at Bigmart.

Learning Software / Hardware requirement :- OS (Linux), python libraries.

Theory :-

Linear Regression :-

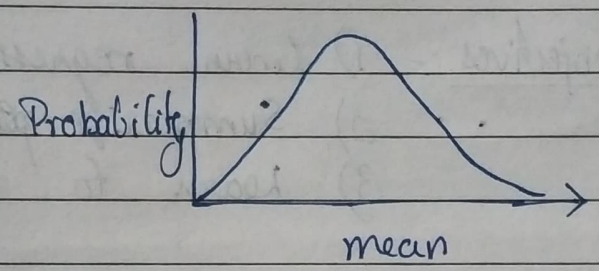
1) It is a linear approach to model the relationship between a scalar response & one or more explanatory variables. The case of one explanatory variable is called simple linear regression.

$$y = B_0 + B_1 * x$$

input x, output y
The line is a plane or hyperplane.

2) Gaussian Distribution:-

1) It is a symmetric distribution where most of the observations cluster around the central ~~peak~~ peak of the probability for values further away the mean taper off equally in both direction



Dataset:-
2013 Sales data for 1559 product across 10 stores in different cities.
Certain attributes of each product & store have been defined. The aim is to build a predictive model.

Attributes :-

Item_Identifier	Outlet_Identifier
Item_weight	Outlet_Establishment_Ye
Item_fat_Content	Outlet_Size
Item_Visibility	Outlet_Location_Type
Item_Type	Outlet_Type
Item_MRP	Item_Outlet_Sales.

Page:
Date: / /

Test Cases:-

Case	Expected o/p	Actual o/p	Remark
Mean K Neighbour for Regression	MSE:- 2720662 Root MSE:- 1649 MAE:-1239 R2:- 0.029	MSE:-2720662 Root MSE:- 1649 MAE:-1239 R2:- 0.029	Passed
Decision Tree	RMSE:- 1546 MAE:-1076 R2:-0.146	RMSE:-1546 MAE:-1076 R2:-0.146	Passed
Linear regression	RMSE:- 1065 MAE:- 752 R2:- 0.59	RMSE:-1065 MAE:-752 R2:-0.59	Passed

MAE → mean absolute error R2 → Coefficient of determination.

Conclusion:-

Thus I have completed bigmart Sales Analysis. ~~And concluded that~~