Assignment 1

Title : To find the best fit line for diven data using linear regression.

Problem statement:

The following table shows the result of recently conducted study on correlation of number of hrs spent driving with risk of developing acute backache. Find the equation of best fit line for this data

No. of hrs spent driving (x)	Risk Score (y)
10	95
9	80
2	10
15	50
10	45
16	98
11	38
16	93

Objective : To understand linear regression

Outcome : To find best scenario for result to be achieved for given dataset using linear regression.

S/W & H/W Package:

i5 / i7 64 bit processor

Os - Windows / Linux

Jupyter Notebook

Scanned by TapScanner

Concept related theory:

Linear Regression:

It is a linear approach to modelling the relationship between scalar response and one or more explanatory variables.

The case of one explanatory variable is linear regression.

The relationships are modeled using linear predictor functions whose unknown model parameters are estimated from data. Such models are called linear models.

Linear regression focuses on conditional probability distribution of response given the values of predictors, rather than joint probability distribution of all of these variables, which is domain of Multivariate analysis.

A regression line is obtained which will give minimum error.

$$y = mx + c$$

The values of m and c must be chosen so that they minimize the error.

$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

If $m > 0$, x & y have positive relationship
If $m < 0$, x & y have negative relationship

Scanned by TapScanner

$$c = \bar{y} - m \times \bar{x}$$

Mean square error (MSE):

$$Mse = \frac{1}{n} \sum_{i=1}^{n} (\text{actual o/p} - \text{predicted o/p})^2$$

Root mean square error (RMSE)

$$rmse = \sqrt{mse}$$

Results:

for the given ~~input~~ dataset,

we get

$$m = 4.58$$
$$c = 12.58$$

∴ Equation of line is-

$$y = 4.58 * x + 12.58$$

RMSE = 22.875

MSE = 518.0047

Scanned by TapScanner

Conclusion:

Thus, linear regression model on given data set is applied and best fit equation is calculated.

Scanned by TapScanner