

Assignment no-A1]

* Title :- To find the best fit line for given data using linear regression.

* Problem Statement :- The follo. table shows results of study on correlation of number of hours spent driving with risk of developing acute back-pain. Find equation of best fit line for the data.

Number of hrs. spent driving (x)	Risk score on 0-100 scale (y)
10	95
9	88
2	10
15	50
10	45
16	98
11	38
16	93

* Objective :- i) To understand when to use linear regression, meaning of linear regression.
ii) Understand working of linear regression on dataset.

* Outcome :- i) I studied and implemented linear regression for finding best fit line for given data.

*S/W and H/W Packages:- Google Colab, Python, numpy, matplotlib, pandas, 64 bit OS, Ubuntu 20.04, 8GB RAM, 1TB HDD, mouse keyboard.

*Theory:-

1. Linear regression is a method that allows us to summarize and study relationships between two continuous variables.
2. The line of best fit is a straight line that will go to centre of data points in our scatter plot.
3. Closer the points are to the line, stronger correlation exists between two variables.

$y = mx + b$, is equation of straight line

x : given data point x

y : given data point y .

m : slope

b : intercept.

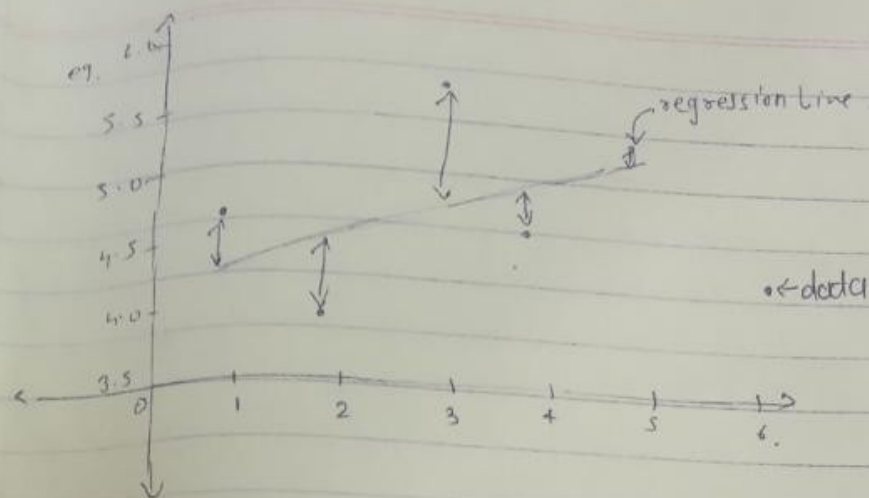
$$m = \frac{\sum y_i x_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

$$b = \frac{\bar{y} \sum x_i^2 - \bar{x} \sum y_i x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

4. Using this formula and python we can find best fit line.

5. Now to calculate how close the points are to the line we use R squared method.

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$



The distance between regression line's y values, and the data's y values is error, then we square that. The line's squared error is either a mean or sum of this, we'll simply sum it.

Conclusion:- I successfully implemented Linear Regression for finding best-fit of line.

Paperwork

1. Here we are given an input feature-output label mapping so this is supervised learning.

SET THEORY

$$S = \{ s, e, X, Y, f_{me}, f_i^f, Mem \mid \emptyset \}$$

s = input set is feature X (number of hours) and initial state for best fit line is $m=0, b=0$

e = output is a value > 0 for m, b

X(hours of driving) = {Feature X}

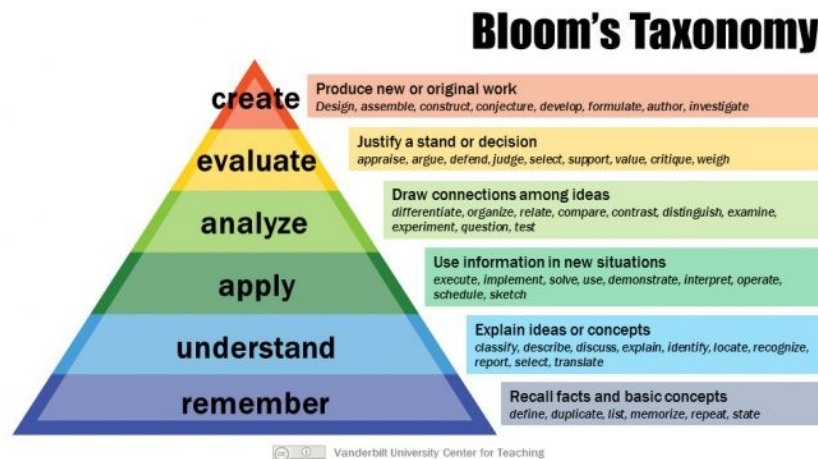
Y(risk score) = Output Label

2. Perception (interpretation of data by machine similar to humans using their senses) – finding of function between input feature X and output label Y
eg. $X = \{1, 2, 3\}$ and $Y = \{1, 4, 9\}$

So human uses his senses and finds out $f(x) = x^2$

3. Cognition : In this we have found the function which can help us map input X to output Y

Blooms Taxonomy -



1) Remember -

1. Linear means a straight line i.e. equation satisfying $y=mx+b$
2. Regression means relationship between two variables (one is dependent Y and other independent X)
3. Linear Regression – method to find relation/mapping function between given input feature X and Output label Y.

2) Understand -

1. In the given data there is an input feature X (hrs) mapping with Output label Y (risk) and hence it is Supervised Learning.

3) Apply -

1. Suppose $X=\{1,2,3\}$ and $Y=\{1,4,9\}$ then using human intelligence we get function as $f(x) = x^2$

So using the Machine we have to find function which maps input feature to output label this can be done by linear regression.

4) Analyze -

1.This involves evaluating of the risk score obtained.Here the dataset has an output label for predicting back pain which can be classified into acute,moderate,severe and dependent on different features on which value of Y depends.

4.In given dataset for input value X we get Output label Y as 95 and 45.So this difference of value appears due to the fact that the Output label is derived from more set of features.

Kaggle link ([Back Pain Dataset](#))

There are 12 features involved for the backpain given in dataset.

WRITTEN BY
Nasir Islam Sujan
Research Assistant |
Machine Learning
Practitioner |
ASP.NET, C#
developer | UI/UX
designer.

Follow

97

Q Search this file...

col. no.	Attribute name	type
Col1	pelvic_incidence	float64
Col2	pelvic_tilt	float64
Col3	lumbar_lordosis_angle	float64
Col4	sacral_slope	float64
Col5	pelvic_radius	float64
Col6	degree_spondylolisthesis	float64
Col7	pelvic_slope	float64
Col8	Direct_tilt	float64
Col9	thoracic_slope	float64
Col10	cervical_tilt	float64
Col11	sacrum_angle	float64
Col12	scoliosis_slope	float64
Class_att	Attribute Class	object
Unnamed	dummy Column	

lower_back_pain_dataset_description.csv hosted with ❤ by GitHub

view raw

Along with this different features on which the back pain depends can be Prolonged seating,constrained posture,low frequency vibration,Poor Posture,Accident or Collision etc.