41256

Jan 6, 2022.

# Assignment No. 1.

* Problem Statement :
Find the best fit line for the given data using Linear regression.

* Objective :
To understand how linear regression works on the given data set.

* Outcome :
To find the best Scenario for the results to be achieved for the given dataset using linear regression.

* Theory :

Machine learning is an application of AI that Provides systems the ability to automatically learn & improve from experiance, without being explicitly Programmed. ML focuses on the development of computer Programs that can access data and use it to learn for themselves.

List of common machine learning Algorithms→
i. Linear Regression
ii. Logistic Regression

iii. Decision Tree
iv. SVM
v. Naive Bayes
vi. KNN
vii. K-Means
viii. Random forest, and so on.

In Linear Regression, the relationships are modeled using linear Predector functions, whose unknown model Parameters are Stablished from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explainatory Variables (or Predector) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. like all form of regression analysis, Linear Regression focuses on the conditional Probability distribution of the response given the values of the Predectors, rather than on the joint Probablity of all these Variables, which is the domain of multivariate analysis.

$$b_0 = \bar{y} - b_1 \bar{x}$$

where,

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$b_0 \rightarrow$ Value on hight at $x = 0$

* Algorithm :
1. Got a bunch of Points in $R^2$, $\{(x^i, y^i)\}$
2. want to fit a line $y = ax + b$ that describes the trend.
3. We define a cost function that computes the total Squared error of our predictions w.r.t. observed values $y^i$ $J(a,b) = \sum(ax^i + b - y^i)^2$ that we want to minimise.
4. see it as a function of a and b : compute both derivatives and force them equal to zero and Solve for a and b
5. The coefficients you get give you the minimum Squared error.
6. can do this for specific Points, or in general and find the formulas.
7. More general version in $R^n$.

* Conclusion :

Thus linear regression model on the given Dataset is applied & a best fit eq. is calculated.

$$\boxed{y = 12.585 + 4.588x}$$