<table>
<tr><td>ASSIGNMENT NO</td><td>A1</td></tr>
<tr><td>TITLE</td><td>To find the best fit line for the given data using Linear Regression</td></tr>
<tr><td>PROBLEM STATEMENT/DEFINITION</td><td>The following table shows the results of a recently conducted study on the correlation of the number of hours spent driving with the risk of developing acute backache. Find the equation of the best fit line for this data
<table>
<tr><th>Number of hours spent driving (x)</th><th>Risk score on a scale of 0-100 (y)</th></tr>
<tr><td>10</td><td>95</td></tr>
<tr><td>9</td><td>80</td></tr>
<tr><td>2</td><td>10</td></tr>
<tr><td>15</td><td>50</td></tr>
<tr><td>10</td><td>45</td></tr>
<tr><td>16</td><td>98</td></tr>
<tr><td>11</td><td>38</td></tr>
<tr><td>16</td><td>93</td></tr>
</table></td></tr>
<tr><td>OBJECTIVE</td><td>To understand how linear regression works on the given dataset</td></tr>
<tr><td>OUTCOME</td><td>To find the best scenario for the result to be achieved for a given data set using linear regression</td></tr>
<tr><td>S/W PACKAGES AND HARDWARE APPARATUS USED</td><td>Core 2 DUO/i3/i5/i7 64-bit processor
OS-LINUX 64 bit OS
Editor-gedit/Eclipse
S/w- Jupyter Notebook/ Weka/ Python</td></tr>
<tr><td>REFERENCES</td><td>1. Giuseppe Bonaccorso, “ Machine Learning Algorithms”, Packt Publishing Limited, ISBN-10: 1785889621, ISBN-13: 978-1785889622
2. Josh Patterson, Adam Gibson, “Deep Learning: A Practitioners Approach”, O’REILLY, SPD, ISBN: 978-93-5213-604-9, 2017 Edition 1st.
3. Nikhil Buduma, “Fundamentals of Deep Learning”, O’REILLY publication, Second Edition,</td></tr>
</table>

	2017,ISBN: 1491925612
STEPS	1. Get the points, R^n 2. Frame the equation of line. 3. Define cost function 4. Compute derivatives 5. Calculate co-efficient which gives the minimum squared error.
INSTRUCTIONS FOR WRITING JOURNAL	1. Date 2. Assignment No. 3. Problem Definition 4. Learning Objective 5. Learning Outcome 6. Concepts Related Theory 7. Algorithm 8. Test Cases 9. Conclusion/Analysis

- **Prerequisites:** Basic knowledge about Algorithms and any programming knowledge Java/python

- **Concepts related Theory**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use

it learn for themselves. List of Common Machine Learning Algorithms are:

- Linear Regression.
- Logistic Regression.
- Decision Tree.
- SVM.
- Naive Bayes.
- kNN.
- K-Means.
- Random Forest, and so on.

Let's learn about Linear Regression. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

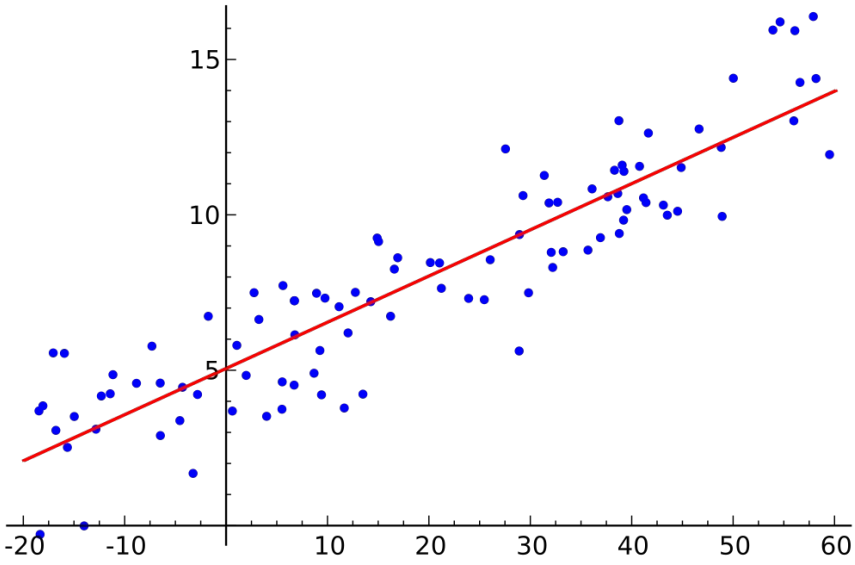

Figure: Linear Regression Graph

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less

commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

Real-time example:

We have a dataset which contains information about relationship between ‘number of hours studied’ and ‘marks obtained’. Many students have been observed and their hours of study and grade are recorded. This will be our training data. Goal is to design a model that can predict marks if given the number of hours studied. Using the training data, a regression line is obtained which will give minimum error. This linear equation is then used for any new data. That is, if we give number of hours studied by a student as an input, our model should predict their mark with minimum error.

Y(pred) = b0 + b1*x

The values b0 and b1 must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.

$$\text{Error} = \sum_{i=1}^{n} (actual\_output - predicted\_output) ** 2$$

Figure : Error Calculation

If we don’t square the error, then positive and negative point will cancel out each other.

For model with one predictor,

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

Figure: Intercept Calculation

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

Figure: Co-efficient Formula

Exploring ‘b1’

If b1 > 0, then x(predictor) and y(target) have a positive relationship. That is increase in x will increase y.
If b1 < 0, then x(predictor) and y(target) have a negative relationship. That is increase in x will decrease y.

Exploring ‘b0’

If the model does not include x=0, then the prediction will become meaningless with only b0. For example, we have a dataset that relates height(x) and weight(y). Taking x=0(that is height as 0), will make equation have only b0 value which is completely meaningless as in real-time height and weight can never be zero. This resulted due to considering the model values beyond its scope.
If the model includes value 0, then ‘b0’ will be the average of all predicted values when x=0. But, setting zero for all the predictor variables is often impossible.
The value of b0 guarantee that residual have mean zero. If there is no ‘b0’ term, then regression will be forced to pass over the origin. Both the regression co-efficient and prediction will be biased.

Co-efficient from Normal equations

Apart from above equation co-efficient of the model can also be calculated from normal equation.

$$\text{Theta} = (X^T X)^{-1} X^T Y$$

Figure: Co-efficient calculation using Normal Equation

Theta contains co-efficient of all predictors including constant term ‘b0’. Normal equation performs computation by taking inverse of input matrix. Complexity of the computation will increase as the number of features increase. It gets very slow when number of features grow

large.

Optimizing using gradient descent

Complexity of the normal equation makes it difficult to use, this is where gradient descent method comes into picture. Partial derivative of the cost function with respect to the parameter can give optimal co-efficient value.

Residual Analysis

Randomness and unpredictability are the two main components of a regression model.

Prediction = Deterministic + Statistic

Deterministic part is covered by the predictor variable in the model. Stochastic part reveals the fact that the expected and observed value is unpredictable. There will always be some information that are missed to cover. This information can be obtained from the residual information.

- **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 minimize.
4. See it as a function of a and b: compute both derivatives, 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 data set is applied and a best fit equation is calculated.