**Case Study 1 - Linear regression**

2021F-T1 AISC1003 - Machine Learning 1 01 (M07 Group 1)

Prof. Jagmohan Dutta

Name : Pratik Domadiya

Student Id : **500199494**

*Introduction*

1. Linear regression is machine learning algorithm based on supervised learning method. It targets the prediction values based on independent input variable.
2. Linear Regression is used for finding out relation between variables and forecasting.
3. Hypothesis function for Linear Regression.

**x:** input training data

**y:** labels to data (supervised learning)

The model gets the best regression fit line by finding the best θ1 and θ2 values.  
**θ1:** intercept  
**θ2:** coefficient of x



1. Cost function :

Pred(i) : predicted y value (pred)

Y(i) : true y value (y).



1. In order to find best fit line we need to minimize the values of  θ1 and θ2. So update the values of  θ1 and θ2, to reach the best values that minimize the error between predicted y value and true value of y.

*Assumption of Linear Regression*

1. There are many assumptions are available in Linear Regression, though we are going to discuss most 5 important one.
2. Linear Regression required linear relationship between dependent and independent variable.

* It is checked by scatter plot of the data.
* It is very crucial to check for the outliers as Linear regression is more sensitive towards outliers.

1. Required all variable to be multivariate normal.

* It is checked by ploting histograms or Q-Q plot.
* When data is not normalized then non-linear transformation like log transformation will solve the issue.

1. No or little Multicollinearity .

* Multicollinearity occurs when independent variables are too highly correlated with each other.
* It is tested with 3 central criteria.
* Correlation metrix :
* Compute the metrix of pearson’s bivariate correlation among all the variables and coefficient of the correlation needs to be 0.
* Tolerance :
* If tolerance < 0.1 , there might be multicollinearity
* If tolerance < 0.01, there certainly is
* Variance Inflation factor ( VIF) :
* VIF > 5 , there might be multicollinearity
* VIF > 10, there is certainly is.

1. Linear Regression required little or no correlation in the data.

* It occurs when the residuals are not independent from each other.
* Values of y(x+1) is not independent from y(x). Ex. Stock prices. Current stock price is not independent from previous price.
* It is tested by scatter plot and durbin-watson test.

1. Homo Scendasticity

* Residuals are equal across the regression line.
* It is tested by scatter plot and gold feld quandt test.

*Different Regression Algorithm*

1. Linear Regression

* It includes predict the target variables based on independent input variables.

|  |  |
| --- | --- |
| Pros | Cons |
| * Simple to implement * Less complexity * May lead over fitting but it can be avoided by some dimensionality reduction techniques such as regularization. | * Outlier affect badly to model performance * Over simplify the real world problem by assuming a linear relationship among variables. |

1. Decision Tree Regression

* It captures non linear interaction between features and target variables. It applies to all those data either it is numerical or categorical.

|  |  |
| --- | --- |
| Pros | Cons |
| * Easy to understand and interpret * Work with all the types of the data * Required little data pre processing, No need for one hot encoding or dummy variables etc. | * Tends to over fit * Small change resulted in a big difference, which causes instability. |

1. Support Vector Regression

* Support Vector Regression uses the same idea as support vector machine.
* It uses hyper planes to segregate the data. The main objective is, points are lies within boundary line.

|  |  |
| --- | --- |
| Pros | Cons |
| * Robust to outliers * Excellent generalization capabilities * High prediction accuracy. | * Not suitable for larger dataset * Not perform very well when the data set has more noise. |

1. Lasso Regression

* Lasso stands for “ Least Absolute Selection Shrinkage Operator”. It basically a shrinkage and variable selection method and it helps to determine which of the predictors are the most important.

|  |  |
| --- | --- |
| Pros | Cons |
| * It avoids over fitting | * It will select only one feature from a group of correlated features. * Selected feature can be highly biased. |

1. Random Forest Regression

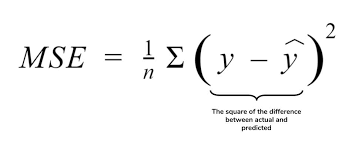
* Combination of the decision trees.
* It is supervised machine learning algorithm used for both classification and Regression.

|  |  |
| --- | --- |
| Pros | Cons |
| * Good at learning complex and non-linear relationship. * Easy to interpret and understandable | * Prone to over fitting. * For larger forest need more memory and that leads to slow down the speed. |

*Metrics for regression*

1. Mean Squared error

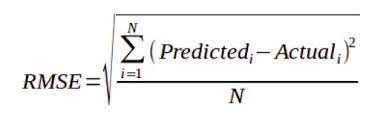
* Minimize the mean squared error between prediction and expected values.



* Larger MSE resulted into large positive error that is effect of punishing model with large values.

1. Root mean squared error

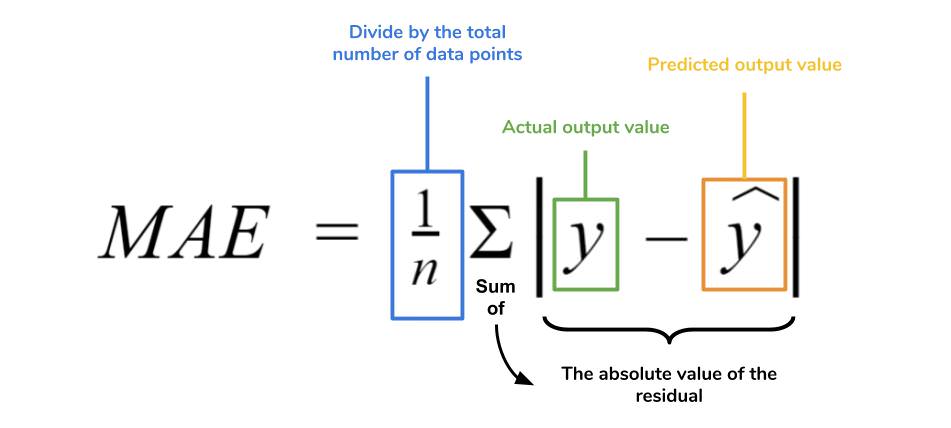
* It is an extension of MSE.



* It reverses the operation of MSE. By means that MSE uses squared operation to remove sign of the each error values and punish large error. The RMSE reverse this operation, though it ensures the result remains positive.

1. Mean absolute error

* MSE & RMSE punishes large error more than smaller one, whilst MAE doesn’t give more or less weight to different types of error and instead the score, it increases linearity with increases in error.

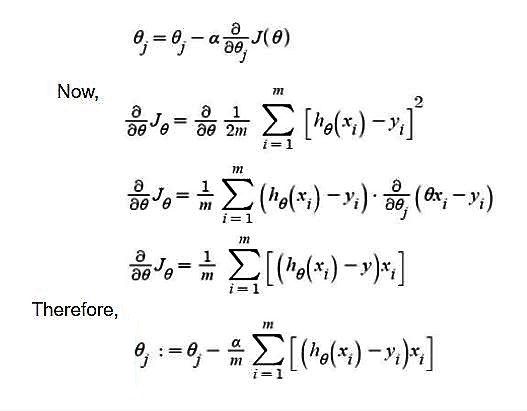


*Gradient Descent*

1. In Linear regression target values predicted based on independent variables. While training, the model will calculate MSE between predicted values and expected values, that is known as cost function. So it should be minimize in order to achieve the best values of θ1 and θ2.
2. Initially Model will select random values of θ1 and θ2, and then update it on the each iteration till it gets minimum value.

* Linear Regression Cost Function:



* Goal : minimize J
* Algorithm :

**θj :** Weights of the hypothesis.

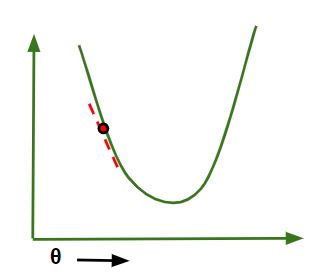
**hθ(xi) :** predicted y value for ith input.

**j :** Feature index number (can be 0, 1, 2, ......, n).

**α :** Learning Rate of Gradient Descent.

Source : <https://media.geeksforgeeks.org/wp-content/uploads/gradiant_descent.jpg>

* Cost Function Graph – Negative Slope



Gradient Descent steps down the cost function in the direction of the steepest descent .

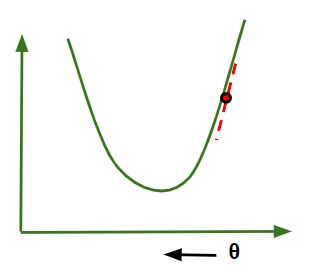
Size of each step is determined by parameter Alpha (α)

θj = θj – (-ve value).

Hence value of θj increases.

Source : https[://media.geeksforgeeks.org/wp-content/uploads/theta-decrease.jpg](https://media.geeksforgeeks.org/wp-content/uploads/theta-decrease.jpg)

* Positive slope



θj = θj – (+ve value).

Hence value of θj decreases

Source : <https://media.geeksforgeeks.org/wp-content/uploads/theta-decrease.jpg>

* Choice of Alpha (α) is very crucial fector.
* If we choose α is to be very high then our model can overshoot the minimum value and fail to converge or diverge.
* If we choose α is to be very low then our model will take small steps to reach minimum value, hence it will take more time.

References

* <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-linear-regression/>
* <https://www.geeksforgeeks.org/gradient-descent-in-linear-regression/>
* <https://www.geeksforgeeks.org/ml-linear-regression/>
* <https://www.analyticsvidhya.com/blog/2021/05/5-regression-algorithms-you-should-know-introductory-guide/>
* <https://machinelearningmastery.com/regression-metrics-for-machine-learning/>