What Are Overfitting and Underfitting?

To train our machine learning model, we give it some data to learn from. The process of plotting a series of data points and drawing the best fit line to understand the relationship between the variables is called Data Fitting. Our model is the best fit when it can find all necessary patterns in our data and avoid the random data points and unnecessary patterns called Noise.



## What Are Overfitting and Underfitting?

To train our machine learning model, we give it some data to learn from. The process of plotting a series of data points and drawing the best fit line to understand the relationship between the variables is called Data Fitting. Our model is the best fit when it can find all necessary patterns in our data and avoid the random data points and unnecessary patterns called Noise.

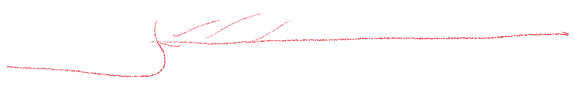
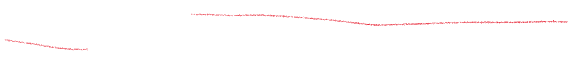
Conversely, in a scenario where the model has not been allowed to look at our data a sufficient number of times, the model won’t be able to find patterns in our test dataset. It will not fit properly to our test dataset and fail to perform on new data too.

A scenario where a machine learning model can neither learn the relationship between variables in the testing data nor predict or classify a new data point is called Underfitting. ----both training set and testing set accuracy is low.

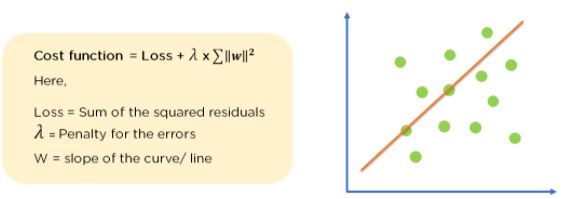
## Ridge Regularization :

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.



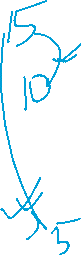
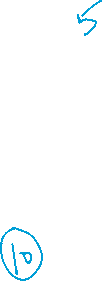
This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present. The function depicted below shows the cost function of ridge regression :







In the cost function, the penalty term is represented by Lambda λ. By changing the values of the penalty function, we are controlling the penalty term. The higher the penalty, it reduces the magnitude of coefficients. It shrinks the parameters. Therefore, it is used to prevent multicollinearity, and it reduces the model complexity by coefficient shrinkage.



Consider the graph illustrated below which represents Linear regression :

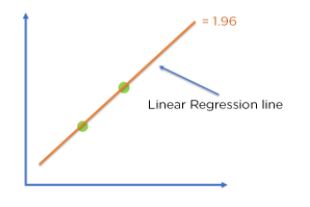
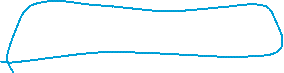


Figure 8: Linear regression model

Cost function = Loss + λ x∑‖w‖^2



For Linear Regression line, let’s consider two points that are on the line,

Loss = 0 (considering the two points on the line)

λ= 1

w = 1.4



Then, Cost function = 0 + 1 x 1.42

            = 1.96



For Ridge Regression, let’s assume,



Loss = 0.32 + 0.22 = 0.13

λ = 1



w = 0.7

Then, Cost function = 0.13 + 1 x 0.72



            = 0.62



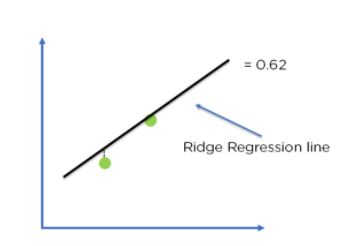


 Figure 9: Ridge regression model

Comparing the two models, with all data points, we can see that the Ridge regression line fits the model more accurately than the linear regression line.

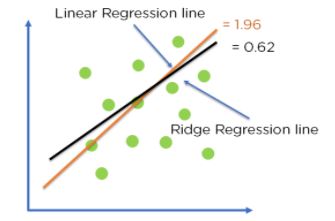


 Figure 10: Optimization of model fit using Ridge Regression

## Lasso Regression

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.



Lasso regression also performs coefficient minimization,  but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients. Consider the cost function for Lasso regression :

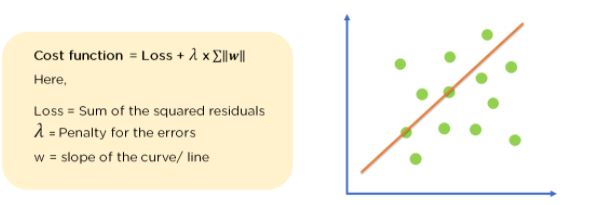


                                                 Figure 11: Cost function for Lasso Regression

We can control the coefficient values by controlling the penalty terms, just like we did in Ridge Regression. Again consider a Linear Regression model :

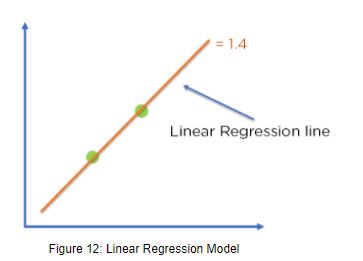


                                                  Figure 12: Linear Regression Model

Cost function = Loss + λ x ∑‖w‖



For Linear Regression line, let’s assume,

Loss = 0 (considering the two points on the line)



λ = 1

w = 1.4

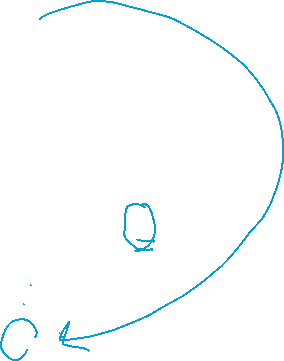


Then, Cost function = 0 + 1 x 1.4

            = 1.4



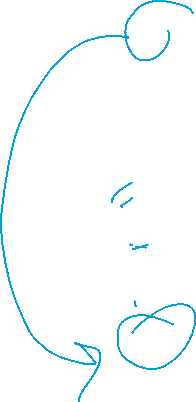
For Ridge Regression, let’s assume,



Loss = 0.32 + 0.12 = 0.1



λ = 1



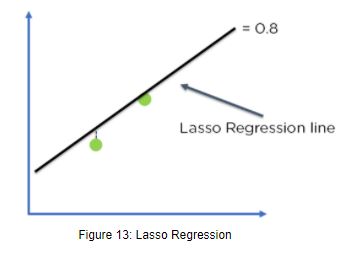
w = 0.7

Then, Cost function = 0.1 + 1 x 0.7



            = 0.8





Comparing the two models, with all data points, we can see that the Lasso regression line fits the model more accurately than the linear regression line.



**Lasso** is especially useful for feature selection on large data sets because it can effectively reduce noise from irrelevant variables. Additionally, Lasso can be used to identify key drivers of returns by shrinking coefficients toward zero. This helps eliminate redundant variables from consideration while still accounting for nonlinear relationships between features.



Ridge Regression is ideal for predicting outcomes when there are too many input variables, and multicollinearity is present. This technique can help reduce the effects of this by shrinking coefficients toward zero while still preserving their sign, allowing you to identify which inputs are most important for predicting an outcome.



Ridge Regression is also useful for predicting outcomes when there are outliers in the data, as it helps mitigate potential overfitting from these outliers by penalizing them less than other inputs.

