**Regularization**

*Regularization comes into play when we have overfitting model. At times, our model is overfitting which means it is trying to accommodate itself too much to train data, hence there are lots of chances that it will not perform well with future data. In other terms, it will accommodate all the data’s which is of less significance, causing too much variance with test data. To avoid this kind of problem regularization comes into picture.*

*Avoiding overfitting will improve the model performance to greater extend. Regularization technique prevents the data from overfitting by adding extra information to the model. It reduces the dependency of one variable with another and thereby generalizing the model with more accuracy.*

*Regularization works by adding penalty to the residual sum of squares. In this we are reducing the variance at the cost of introducing some bias. This is almost always beneficial for the predictive performance of the model.*

**Types of Regularization**

*There are 3 different types of regularization techniques available*

1. *Ridge*
2. *Lasso*
3. *Elastic Net*

***Ridge regression***

*In Ridge regression, the cost function is altered by adding the penalty term, which is multiplies by the square of slopes. Setting the λ size to zero means its same as least squares. λ can take value between 0 to +∞, while the larger its value the stronger it gets penalized. As the λ becomes larger, the variance decreases and the bias increases.*

*Now comes the question, how much bias we can introduce the model? This we can find by doing cross validation. By performing validation with different λ values, we can find out which model or value fits in the test data well.*

*Ridge regression can be used when we know about the significance of data and high collinearity exists between independent variables. In Ridge regression as we generalize the mode more and more the coefficients become very close to zero but never to zero.*

***Lasso regression***

*Lasso is similar to the Ridge Regression except that the penalty term includes the absolute weights instead of a square of weights.* *Penalty has the eﬀect of forcing some of the coeﬃcient estimates to be exactly equal to zero which means there is a complete removal of some of the features from model evaluation when the tuning parameter λ is suﬃciently large.*

*Lasso tends to do well if there are a small number of significant parameters, and the others are close to zero. We can go for lasso when we are not known of the significance column, while ridge in contrary can be used when we know the significance column.*

***Elastic Net***

*In practice, we don't know the true parameter values, so we are unsure which one to use either ridge or lasso. We can either perform a cross validation and finds out which suits best or combine both ridge and lasso, there comes the elastic net. It is the combination of both ridge and lasso.*