**Ridge Regression**

It tries to fit a new line that doesn’t fit the training data well, thereby introducing a small amount of bias. By starting with a slightly worse fit, it can provide a better long-term prediction. Ordinary Least Squares technique tries to fit a model to the data in order to reduce the sum of the squared residuals, but Ridge Regression tries to reduce the sum of the squared residuals along with the term: **lambda \* (slope)^2.** So (slope)^2 adds a penalty to the traditional least square method and lambda determines how severe the penalty is.

Without the small amount of bias the penalty creates the least squares fit has huge variance. In contrast the ridge regression line that has a small amount of bias due to penalty has less variance.

**What effect does the ridge regression penalty has on how the line is fit to the data?**

When the slope of the line is steep, then the prediction for y will be very sensitive to small changes in x (input). It has the opposite effect when the slope of the line is small or not steep: prediction for y barely changes to small changes in x. The ridge regression tries to fit a line with small slope.

**Lambda**

It can vary from 0 to infinity. When lambda is 0 then both ridge and least squares line are the same as they both trying to minimize the sum of the square residuals. The larger we get the lambda the slope gets asymptotically closer to 0. So when lambda gets larger and larger the prediction for y gets lesser and lesser sensitive to changes in x.

**So how do we decide which value to choose for lambda?**

We could use cross validation to choose the value, that results in lower variance, for lambda.

**Benefits of Ridge Regression**

Ridge regression can also be applied to logistic regression. Ridge regression is especially effective in case of small sample size. Least square regression provides poor predictions when small size is small. When applied to logistic regression, it optimizes the sum of the likelihoods than minimizing the sum of the squared residuals as logistic regression is solved using maximum likelihood concept. So what does it do: it shrinks the value of lambda thereby getting out target variable, y less sensitive to the changes with regards to the predictor variables(x).

Ridge regression helps to reduce variance by shrinking the parameters and making our predictions less sensitive to them. In general, the ridge regression penalty contains of all of the parameters except for the y-intercept. So we have to square every individual parameter and sum them altogether and finally multiply by lambda.

**So what do we do when have 10,001 parameters, but only 500 data points?**

We use ridge regression. In general, when small sizes are relatively small then the ridge regression can improve predictions made from new data, by reducing variance, by making the predictions less sensitive to the training data.