### **Regularization**

**Regularization** is an important concept that is used to avoid overfitting of the data, especially when the trained and test data are varying much.

**Regularization** is implemented by adding a “penalty” term to the best fit derived from the trained data, to achieve a lesser variance with the tested data and also restricts the influence of predictor variables over the output variable by compressing their coefficients.

In regularization, what we do is normally keep the same number of features but reduce the magnitude of the coefficients. We can reduce the magnitude of the coefficients by using different types of regression techniques that use regularization to overcome this problem. So, let us discuss them.

***Warning: The more features in our dataset, the harder our pairplot will be to interpret.***

Sns . boxplox -> box plot

Another important thing to look for while we're exploring our data is multicollinearity.

Multicollinearity means that several variables are essentially measuring the same thing. Not only is there no point in having more than one measure of the same thing in a model, but doing so can actually cause our model results to fluctuate.

***Note: Depending on the situation, it may not be a problem for your model if only slight or moderate collinearity issue occur. However, it is strongly advised to solve the issue if severe collinearity issue exists(e.g. correlation >0.8 between 2 variables)***

***No severe collinearity issues.***

**loss functions**

**Mean Absolute Error** (MAE) is the mean of the absolute value of the errors:(is the easiest to understand, because it's the average error.)

**Mean Squared Error** (MSE) is the mean of the squared errors:(is more popular than MAE, because MSE "punishes" larger errors, which tends to be useful in the real world.)

**Root Mean Squared Error** (RMSE) is the square root of the mean of the squared errors:( is even more popular than MSE, because RMSE is interpretable in the "y" units.)

All of these are **loss functions** because we want to minimize them.

**sklearn** can calculate all of these metrics for us

R^2the percentage of variation in y is explained by all the x variables together. Usually, an R^2 of .70 is considered good.

### **Linear Regression**

### **Ridge Regression** is a technique for analyzing multiple regression data that suffers from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value.

### **Lasso Regression ->** The “LASSO” stands for Least Absolute Shrinkage and Selection Operator. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean.

**The key difference to remember here is that Lasso shrinks the less important feature’s coefficient to zero, thus removing some features altogether. So, this works well for feature selection in case we have a huge number of features.**