Assignment 7

1. How sci-kit learn module performs feature scaling?
   Normalization *and* **Standardization** are two techniques commonly used during Data Preprocessing to adjust the features to a common scale.
   From **Data preprocessing** package of **sci-kit learn StandardScaler** and **MinMax scaler** classes are imported to perform data scaling. The training data is fitted first using **fit ()** method to find its mean and standard deviation. Then the entire data is transformed using **transform ()** method to be normally distributed with mean and standard deviation

2. If you are performing feature scaling, what you will prefer Standardisation or Normalization?
   Normalization is good to use when the distribution of data does not follow a Gaussian distribution.
   Standardization, on the other hand, can be helpful in cases where the data follows a Gaussian distribution.

3. What is the difference between the residual sum of squares and regularization?
   The **residual sum of squares** measures the variation in the error between the observed data and modelled values.
   Regularization is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. Regularization is performed by adding an penalty term to the residual sum of squares. A regularized model finds a line that reduces sum of squared residuals added with a penalty.

4. Differentiate between L1, L2, and Elastic Net regularization on the basis of the alpha parameter in sci-kit learn?
   Lasso Regression also known as L1 Regularization, is similar to Ridge regression. It also fits a line with little bias but which has less variance than linear regression. This is done by adding penalty to shrink the slope of the line until we get the best fit. So, when the penalty of the model that regularizes is the Sum of absolute slopes then it is called **Lasso regression.**

Sum of squared residuals + λ*(Sum of |slopes|)

- While the Ridge regression reduce the slope very **close to 0**, the Lasso regression fully reduce it to 0.
- It removes the dependency of the output variable on some of the features.
- It is useful when our data contains some unimportant features.

Ridge Regression also known as L2 Regularization. This model finds a line that doesn't fit the training data well but reduce the error with testing data. It means we introduce a small bias to the line fitting the data. But in return, we get drop in the variance.
This new line called the Ridge regression line, provides better predictions than the linear model line though this line doesn't best fit the training data.
A regularized model finds a line that reduces sum of squared residuals added with a penalty. When the penalty is **Sum of squared slopes** then it is called **Ridge regression**.

Sum of squared residuals + λ*Sum of (slopes) ^2

- While the Ridge regression reduce the slope very close to 0.
- Ridge regression works better when the output variable depends on all the features (it means, when all the features are important).

Elastic net is a combination of Ridge and Lasso regularization. It tries to find a line that reduces both **λ1** (lasso) and **λ2** (ridge).

- When there are very large number of features and we don't know which features are useful and which are not.

Sum of squared residuals + λ1*(Sum of |slopes|) + λ2*Sum of (slopes) ^2

5. Suppose over fitting is happening, what kind of regularization you will prefer to perform?
   Regularization of a model deals with the over-fitting problem. There are many techniques for regularization. Choosing the best regularization method depends on the usefulness of the features. If all the features are important to output then the **Ridge regression** is chosen. If only some features are important to the output then **Lasso Regression** is used. When we have very large number of features and we don't know about their usefulness then we choose **Elastic Net Regression**.