DS-630 Week 3 Assignment

1. Why would you want to use
    a. Ridge Regression instead of plain Linear Regression(i.e without any regularization)?
    ➔ we use linear regression when we want to include all parameters in the model. However, in this case, the model leads to overfitting, we should use ridge regression instead of plain regression to engineer the model for the best trade between alfa and test set score.

    b. Lasso instead of Ridge Regression?
    ➔ Lasso regression is a good choice when we have a large number of features and need to reduce the number of features in the model to make it interpretable and simplified. In Lasso coefficient can go to Zero, whereas in ridge regression, they become small, but they never become zero.

    c. Elastic Net instead of Lasso?
    ➔ Lasso regression can sometimes introduce a small bias into the model where the prediction is reliant on a certain variable. In this situation, Elastic performs better than Lasso regression.
2. Suppose you are using Polynomial Regression. You plot the learning curves and you notice that there is a large gap between the training error and the validation error. What is happening? What are three ways to solve this?

➔ A large gap between the training error and the validation error happened because of the overfitting model. This gap exists because the training set error is lower than the testing or validation set error. Improve the overfitting of a model by providing more data to the training set. By reducing the number of features in the dataset. Another way is to reduce the complexity of the model. Or we can add either ridge regression or lasso regression to our model.