Linear Model Selection & Regularization

Before moving to more flexible, non-linear models in the next chapters, we gonna discuss ways and methods to improve these linear models which often are competitive with non-linear methods.

The methods discuss in this chapter are primarily alternative fitting procedures to least squares but they can be **generalized** to be alternatives for any fitting method

Least squares Core Problem

It can have very **high variance** if the training data is slightly different the resulting model can be very different, which makes the **least squares** very sensitive to the Training Data

These alternatives can yield better :

- **Prediction Accuracy** : if the true relationship between the Response and predictors is approximately linear the **least squares** estimates will have low **bias**. if $n \gg p$ and also results in low **variance**, However if n is not much larger than p there will be a a lot of variability in the least squares fit which results in overfitting and picking up noise instead of learning the data patterns, and if p is larger than n the least squares method no longer have unique solutions(Since the X^TX is no longer invertible, the number of unknowns p is larger then the amount of equations n) which results in *zero* training error, By constraining or shrinking the estimated coefficients we can reduce the variance with an increase in the bias Bias-Variance Trade-Off
- **Model Interpretability** : It is often the case many of the variables and features used in Multiple Linear Regression model are in fact no associated with the response and can be **irrelevant** that leads to unnecessary complexity in the model, some approaches that automatically exclude the irrelevant variables

In This chapter we discuss the following methods :

- Subset Selection
- Shrinkage
- Dimension Reduction