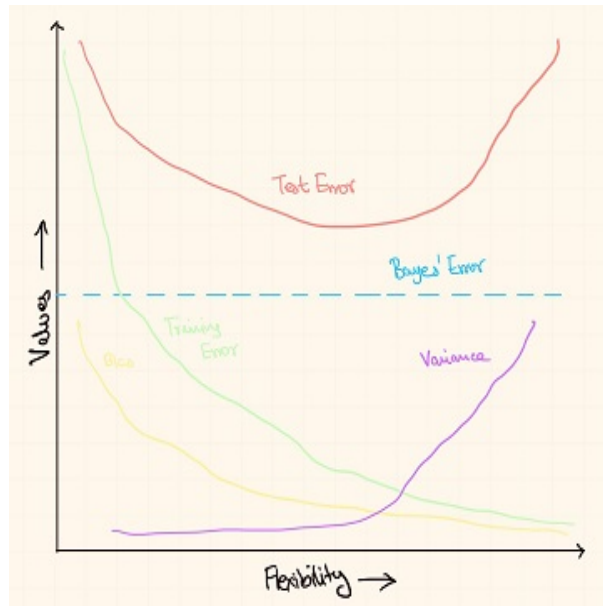


Chapter 2 - Statistical Learning

1.
 - (a) We would expect the flexible algorithm to learn the representation of the data better than the inflexible algorithm. This is because the flexible learning algorithm has lower bias which allows it to model the data more accurately.
 - (b) With smaller data sets flexible models tend to overfit and hence perform worse than inflexible algorithms.
 - (c) We would expect the flexible algorithm to perform better as it has the required higher variance to model the non-linearity in data.
 - (d) The flexible model will perform worse as it will find patterns in the errors present.
2.
 - (a) $n = 500$, $p = 3$, Regression, Inference
 - (b) $n = 20$, $p = 13$, Classification, Prediction
 - (c) $n = 52$, $p = 3$, Regression, Prediction
3.
 - Training Error - It is the error recorded by the statistical model on the training data set. This error generally reduces as model flexibility increases. This reduction in training error is due to the fact that more flexible methods have the ability to model the data more closely (overfitting).
 - Test Error - It is error recorded by the statistical model on the test data set. This error generally has a U-shape. Initially the error reduces as the flexibility increases because the algorithm is able to better fit the data and hence generalizes better on unseen data. However, as the flexibility is increased beyond a point, the error starts to increase again due to the model overfitting.
 - Bias - This is the inherent error occurring due to making an assumption of a real world process. As the number of assumptions we make reduces, flexibility increases, hence bias reduces.
 - Variance - It is the degree to which the model changes when the corresponding data points change. When a model is highly flexible, a small change in the data will lead to a drastic change in the model. This is because the highly flexible model will try to map every data point as closely as possible. Thus with an increase in flexibility, variance increases.

- Bayes' Error - This is also called the irreducible error. It is theoretical lowest possible value of the test error. It is generally equal to the variance of the noise.



4. (a)
 - Detecting scam emails
 - Detecting digits from handwritten digits
 - Differentiating between cancerous and non-cancerous cells
- (b)
 - Predicting house sales
 - Understanding how the data effects the salary of employees
 - Predicting stock prices
- (c)
 - Social Media Analysis
 - Image Segmentation
 - Targeted Advertisements
5. Flexible models are generally used when the datasets are relatively large or when the underlying function is highly non-linear. However, highly flexible models tend to overfit and are harder to interpret.
6. In parametric models, we make an assumption about the underlying function that models the data set. Due to this assumption all we need

to do is find the parameters of our assumed model. This is less computationally intensive as compared to non-parametric methods which try to estimate the underlying function from scratch. However, when the assumption is very off, then the parametric method will perform worse than the non-parametric model.

7. Given that we need to make a prediction for $(0,0,0)$ using KNN, we first need to find the K nearest neighbors to $(0,0,0)$.

$$\begin{aligned}
 \text{(a)} \quad D_{obs1} &= \sqrt{0^2 + 3^2 + 0^2} = 3 \\
 D_{obs2} &= \sqrt{2^2 + 0^2 + 0^2} = 2 \\
 D_{obs3} &= \sqrt{0^2 + 1^2 + 3^2} = \sqrt{10} \\
 D_{obs4} &= \sqrt{0^2 + 1^2 + 2^2} = \sqrt{5} \\
 D_{obs5} &= \sqrt{-1^2 + 0^2 + 1^2} = \sqrt{2} \\
 D_{obs5} &= \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}
 \end{aligned}$$

- (b) With $K=1$, the nearest point is $(-1,0,1)$. Thus $Y_{(0,0,0)}$ is Green.
- (c) With $K=3$, the nearest points are $(-1,0,1)$, $(1,1,1)$, and $(2,0,0)$. Among these points, two are Red and one is Green, thus $Y_{(0,0,0)}$ is Red.
- (d) If the Bayes' decision boundary is highly non-linear we would expect the best value of K to be small. If K is large, the KNN algorithm has to find a large number of points for each observation which reduces the flexibility of the model. Thus to increase flexibility, K has to be kept low.