Comparison of Classification Methods

Analytical Comparison (Mathematical)

Performing an **analytical** comparison between all the classification methods covered in the **Classification** chapter 4.

Consider K class and we assign the observation x to the class that maximize $\Pr(Y=k|X=x)$, and class K as a baseline results in :

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right)$$

Linear Discriminant Analysis :

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right)=\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)$$

$$\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)=\log\left(\frac{\pi_k}{\pi_K}\right)-\frac{1}{2}(\mu_k+\mu_K)^T\Sigma^{-1}(\mu_k-\mu_K)+x^T\Sigma^{-1}(\mu_k-\mu_K)$$

$$\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)=ak+\sum_{j=1}^{p}b_{k,j}x_j$$

With :

- $a_k=\log\left(\frac{\pi_k}{\pi_K}\right)-\frac{1}{2}(\mu_k+\mu_K)^T\Sigma^{-1}(\mu_k-\mu_K)$ constant part
- $b_{kj}=\Sigma^{-1}(\mu_k-\mu_K)$ is the $j\text{th}$ component

Quadratic Discriminant Analysis :

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right)=a_k+\sum_{j=1}^{n}b_{kj}x_j+\sum_{j=1}^{p}\sum_{l=1}^{p}c_{kjl}x_jx_l$$

Naive Bayes :

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right)=\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)$$

$$\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)=\log\left(\frac{\pi_k}{\pi_K}\right)+\sum_{j=1}^{p}\log[\frac{\pi_k f(x)_k}{\pi_K f(x)_K}]$$

$$=a_k+\sum_{j=1}^{p}g_{kj}(x_j)$$

With :

- $a_k=\log\left(\frac{\pi_k}{\pi_K}\right)$
- $g_{kj}(xl=\log\left(\frac{\pi_k f(x)_k}{\pi_K f(x)_K}\right)$ this takes form of a *generalized additive model* which will be the topic of chapter 7

Logistic Regression :

$$\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right)=\beta_{k_0}+\sum_{j=1}^{p}\beta_{kj}x_j$$

- The coefficients in **Logistic Regression** are maximized with **Likelihood function**
- Where the **LDA** makes the assumptions that the X follows a Gaussian distribution

K-Nearest Neighbor :

Since **KNN** is a non-parametric with no assumptions made which make it completely different for the other **parametric** classifier we covered in this chapter.

- **KNN** dominate the linear decision boundaries such **Logistic Regression** and **LDA** when the boundary is highly non-linear , given large enough n and relatively small p

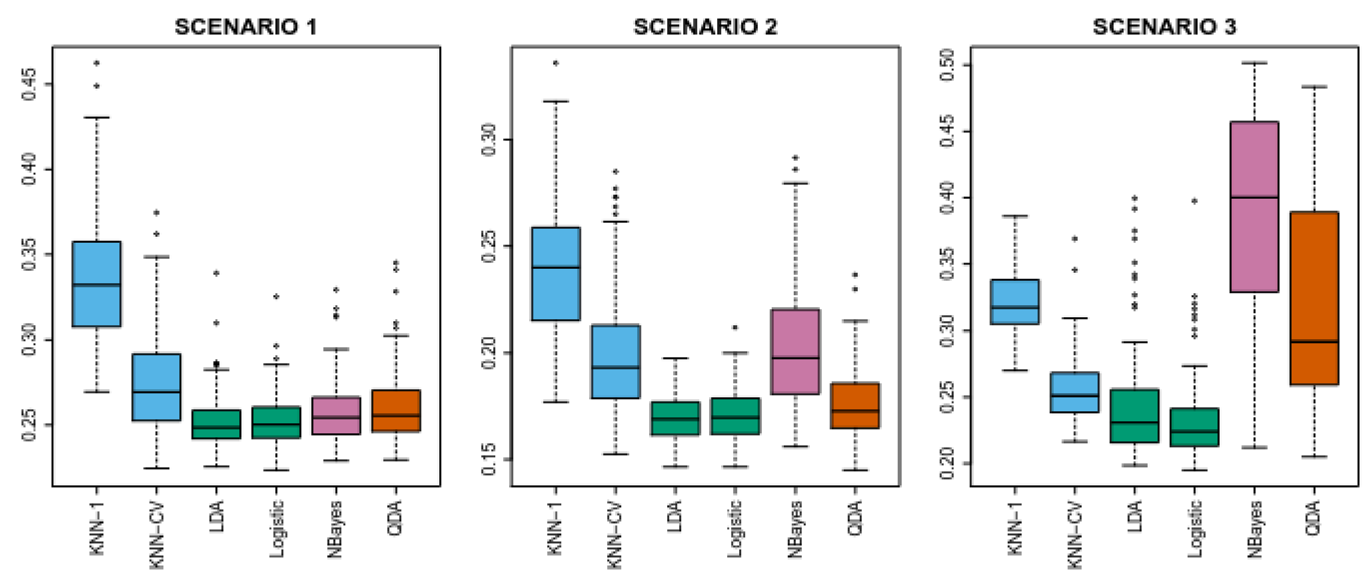
- **KNN** requires very large number of observations since its a non-parametric approach
- **KNN** doesn't tell us any inference or interpretation of the coefficients and the response

Insights

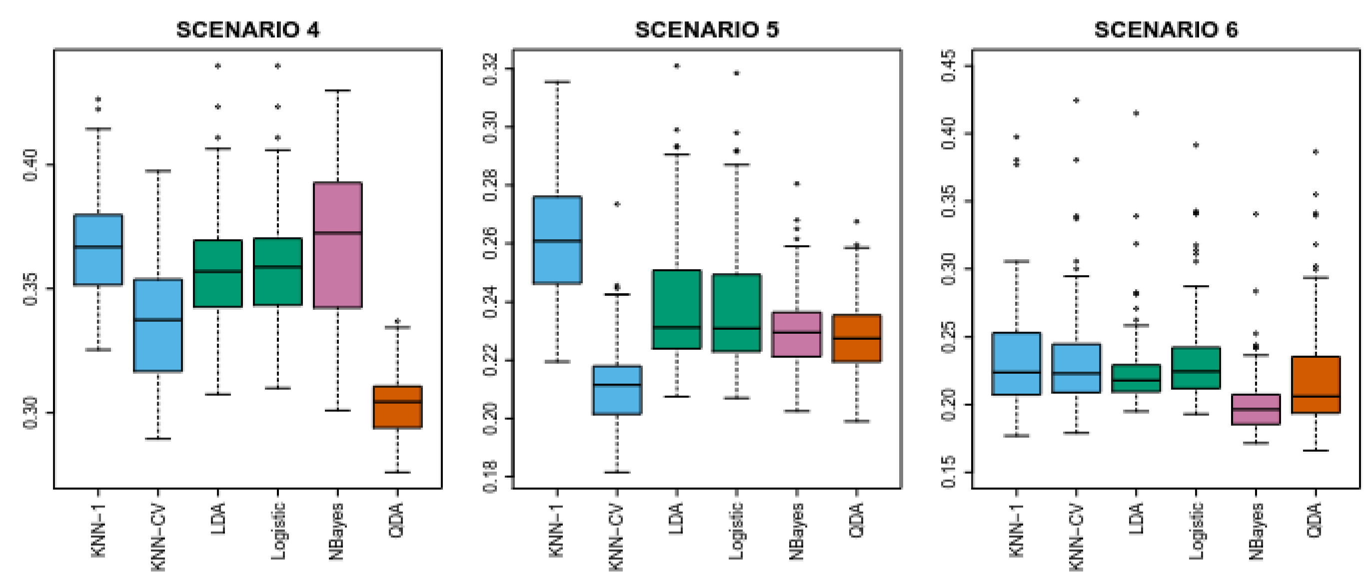
- **LDA** is a special case of **QDA** without the class covariance matrix term $\sum_{j=1}^{p}\sum_{l=1}^{p} c_{kjl}x_j x_l$, **LDA** is just a restricted version of **QDA** where the covariance matrix is shared across all classes k
- Any classifier with **Linear decision boundary** is a special case of **Naive Bayes** with $g_{kj}(xj) = b_{kj}x_j$, Which means that **LDA** is a special case of **Naive Bayes** since it follows the generalized additive models form
- And **Gaussian Bayes Naive** is a special case of **LDA** where the **covariance matrix** forced to be **diagonal** which eliminate the co-variances terms and leave only the variances

An Empirical Comparison

Instead of a mathematical comparison **Empirical** focuses on practical performance of the following **classification** methods : **Logistic regression, LDA, QDA, Naive Bayes, KNN**



- Boxplots for the test error rate for all the classification methods, for a linear decision boundary
- **SCENARIO 1** with uncorrelated variables Both logistic regression and LDA performs quite well due to their linear nature
- **SCENARIO 2** with negative correlation between the variables, we notice a dip in **Naive Bayes** error rate since it violet the assumption of independence
- **SCENARIO 3** with high negative correlation between the features, and still linear , again **Naive Bayes** perform th worse due to the assumptions and same for **QDA** since its non linear
- Logistic Regression is the favored approach since it doesn't depend on any assumption while performing very well on linear problems



- Generated data from a normal distribution with correlated features and non-linear
- **SCENARIO 4** QDA outperform since all the assumption it made were met the Gaussian and the non linearity
- **SCENARIO 5** Highly non linear Which in results the **KNN-CV** preform the best since its the best suited for highly flexible settings
- **SCENARIO 6** With different covariance matrices for each class with a very small observation size $n = 6$ the **Naive Bayes** preforms the best since the assumption of independence and the very low correlation since the observation size is small

- Naive Bayes ,QDA, KNN are the favored in non linear problems and the assumptions they make

Conclusion

Each of the classification methods and approaches are useful and useble and yield good results it all depends on where to use them and when to use them, that's why understanding the nature of the problem is the most important when it comes to picking the Statistical learning methods