Assignment 2

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

Problem 3

3.1.

- *K-NN*: has no assumptions about the data distribution.
- *LDA*: assumes that the data is normally distributed and the covariance of the predictors is the same across all classes.
- *QDA*: also assumes the data is normal distribution however the covariance matrix between the predictors can have different values between different responses "classes".
- *LR*: has the assumption that the observations are independent. And the correlation between the predictors is not high.

If the shape of the decision boundary is:

- *Linear*: Both LDA and LR can be used.
- *Quadratic*: QDA can be used.
- *Highly non-linear*: K-NN.

3.2.

LDA vs LR LR is preferred since

- does not have an assumption that data is drawn from the gaussian distribution
- Logistic regression also works when the predictors are categorical