Authors: Wojciech Ptaś (7042843), Mhd Jawad Al Rahwanji (703890)

Problem 3

1.

K-NN

We do not make any assumptions w.r.t. the data distribution.

LDA

For univariate LDA: - every class is Gaussian-distributed - all classes have the same variance

For multivariate LDA: - each class is a multivariate Gaussian - all the clesses share the same covariance matrix

QDA

- every class is Gaussian-distributed
- classes do not share the same covariance

Logistic Regression

- independent variables should not be correlated with each other
- assumes linear decision boundary

Decision boundary

If the decision boundary is highly non-linear and the sample is big enough, then the kNN algorithm will yield the best results.

If the decision boundary is quadratic, the QDA will perform best and all the linear methods will perform poorly.

If the decision boundary is linear, then the LDA will perform best, but only if the data distribution is Gaussian. Otherwise, the Logistic Regression is going to be the best.

2.

Reasons why LR is preferred over LDA:

- LDA makes much stronger assumptions w.r.t the data distribution
- LDA is much more sensitive to outliers