EKT 720 Assignment 5

Question 1

1. The objective of k-nearest classification is to classify an observation based on the neighbourhood it is closest to using the training set as means to estimation. The estimation can be based on averaging the results or using some distance formula to determine the closest neighbourhood to the test point.
2. A
3. The classification of the two variable model is based on averaging the sum of the yis to estimate the classification of the xi points. The resulting estimate is classified based on a ‘majority vote’ in the neighbourhood.

The classification of the multi-class response model is similar to that of the two variable except that the estimation is based on the distance of the test point from the neighbourhoods. The smaller the distance, the more likely the point is in that neighbourhood.

1. We can use the data to evaluate the predictions of the model by using the bootstrapping technique to split the data such that we have data to test on. Using the predicted results, we can then apply it on the data not used in prediction to see if the predictions come close to those we already have. If so, the predicted model is useful for the classification.
2. The k-nearest model is well suited for categorical (qualitative) data.