

# M4L5 Exercises

CQF January 2024

## Exercise 3

Generate a training dataset containing 30 observations with two predictors centered around -0.5 and 0.5 with a cluster standard deviation of 0.4 and one qualitative response variable. Define classes that takes 'Red' when response variable is positive and 'Blue' otherwise. Use this generated dataset to make a prediction for  $y$  when  $X_1 = X_2 = 0.25$  using K-nearest neighbours.

- a) Compute the Euclidean distance between each observations and the test points.
- b) What is the class prediction with  $K = 1$ ?
- c) What is the class prediction with  $K = 5$ ?
- d) Plot the classification points with decision boundary for  $K = 5$ .

## Exercise 4

For this exercise, use the [admission dataset](#). The dataset contains three predictor variables: gre, gpa and rank and one binary response variable called admit.

- a) List all tunable hyperparameters.
- b) Select the best model by searching over a range of hyperparameters based on cross validation score using an Exhaustive Search.

\* \* \*