

Econometrics 720

Assignment 5

Consider the data file `yx.sas7bdat`

1. Give a full theoretical discussion on k-nearest neighbor classification methods.

Use the following headings:

- a. Objective
- b. The underlying statistical model(s)
- c. Two and multi-class response variables
- d. Evaluation of prediction results
- e. Practical use

2. Use the attached data to

- a. Plot the data
- b. Write a SAS *iml* program to fit a k-nearest neighbor classification model to the data. Base your model on a random selection of 80% of the data. The value of k should be a parameter that can be adjusted as and when needed.
- c. Make a graphical representation that will indicate the classification areas clearly.
- d. Evaluate the model by using the test data (20% of the data that was not used in model estimation). Give a full discussion on the prediction accuracy of the model using the modeling data versus the test data.