Credit Card Clients

The [UCI "Default of Credit Card Clients" data set](#) contains various fields describing 30,000 credit card customers in Taiwan in 2005 (Yeh & Lien, [doi://10.1016/j.eswa.2007.12.020](#)).

1. Split the data set 50/50 into a training set and test set.

2. Logistic Regression

   (a) Create several logistic regression models to predict `default.payment.next.month` (at least 10).

   (b) Calculate the confusion matrices for each model based on the test set.

   (c) Present a summary of the five models that gave the best overall accuracy on the test set. Which variables did you include, and which ones mattered the most?

3. Discriminant Analysis

   (a) Predict `default.payment.next.month` using linear discriminant analysis. Calculate the confusion matrix on the test set.

   (b) Predict `default.payment.next.month` using quadratic discriminant analysis. Calculate the confusion matrix on the test set.

4. KNN Classifiers

   (a) Create four different KNN classifiers to predict `default.payment.next.month` with four different values of k. You will have to choose how to measure distance in a vector space that includes indicator variables and payment amounts ($NT 100k).

   (b) Calculate the confusion matrices for each classifier based on the test set.

5. Which of these classifiers would you recommend? You may make your recommendation based on any accuracy metric (not ncessarily the best overall accuracy).