Dr. Sudhakar Raju

BUS 6309 – LINEAR & MULTIVARIATE MODELS

SPRING 2016

**QUESTIONS FOR ASSIGNMENT 6**

1. Customer arrivals at a bank are random and independent – the probability of arrival in any one minute period is the same as the probability of arrival in any other one minute period. Suppose the mean number of arrivals in any one minute period is 3.

a.) What is the probability of exactly 3 arrivals in a one minute period?

b.) What is the probability of at least 3 arrivals in a one minute period?

2. Phone calls arrive at the rate of 48 per hour at the reservation desk of Regional Airways.

a.) What is the probability of receiving exactly 3 calls in a 5 minute period?

b.) What is the probability of receiving no calls in a 5 minute period?

3. Use the attached dataset (credit\_default\_data) which pertains to a sample of 10,000 credit card holders. The dependent variable is default/no default and the independent variables are student /not a student, credit card balance and income.

a. Run a Logistic Regression on this data. What proportion of the credit card holders default?

b. What is the predicted probability of default for a student with a credit card balance of $1500 and income of $40,000?

c. What is the predicted probability of default for a non-student with a credit card balance of $1500 and income of $40,000?

d. What is the highest credit card balance in the dataset? What is the predicted probability of default for this individual? Does this result make sense?

e. What is the highest income in the dataset? What is the predicted probability of default for this individual? Does this result make sense?

4. Use the same credit\_default\_data in question 3 above. Instead of a Logit Model, run an LDA (Linear Discriminant Model) on the data. What are the Fisher Classification functions for the different groups?

b. What is the predicted Fisher Score for a student with a credit card balance of $1500 and income of $40,000? Is this person in the default or no default group according to LDA? Compare the LDA result to Logit result. Are they in agreement?

c. What is the predicted LDA scores for the individual with the highest income?

d. What is the predicted LDA scores for a non-student with a balance of $1530 and income of $30,003? What is this individual predicted class?

e. The data above pertains to ID: 264. What is this individual’s actual class? What is his predicted class?

f. Interpret the Confusion Matrix for this dataset.

5. Using the credit\_default\_data apply the k-means algorithm.

a. To apply the k-means algorithm, clean the data. Eliminate the columns with id and default leaving you with purely numeric data.

b. Use the 3 variables (student, balance and income) to create k = 2 clusters.

c. Compare the LDA actual and predicted classification with the k means clusters. How well does cluster analysis perform compared to LDA?