**Data Mining and Machine Learning Lab 10 – Quiz!**

**Instructions:** Save this worksheet as xxxxxxxx.doc where <xxxxxxxx> is your UCD student number. Write your answers in this worksheet and save it to your own computer so you don’t lose your answers. Then upload worksheet to moodle as soon as you have completed it.

**Name:**

**UCD Student Number:**

**Using a Logistic Regression Model to predict admission into graduate school**

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, affect admission into graduate school. The response variable, admit/don’t admit, is binary.

This dataset has a binary response (outcome, dependent) variable called admit. There are three predictor variables: gre, gpa and rank. We will treat the variables gre and gpa as continuous. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.

See <https://stats.idre.ucla.edu/r/dae/logit-regression/> for more details.

**Visualize the data**

In Lab 9 you created boxplots and a correlation matrix showing the features in your dataset.

**1.** What did the boxplots tell you about the features?

What did the correlation matrix tell you about:

**2. Predictive** features?

**3. Redundant** features?

**4. Irrelevant** features?

**5. Interacting** features?

**6.** What else should you do to check the quality of your data?

### **Training And Test Sets**

In Lab 9 you divided the data set into two parts: a training set and a test set.

**7.** What percentage of students in the training set are admitted?

**8.** What percentage students in the test set are admitted?

**9.** What issues does this raise?

**Logistic Regression Model and Model Evaluation**

You created a model of your training data using GRE, GPA and Rank as descriptive features.

**10.** What results would you get if you use a difference set of features?

|  |  |
| --- | --- |
|  |  |
| gre+gpa+rank |  |
| gre+gpa |  |
| gpa+rank |  |
| gre |  |
| gpa |  |
| rank |  |

**11.** What feature(s) should you chose for your final model? Why?

**12.** Does this result surprise you? Why?