CS909 Lab 3: Data Mining

January 20, 2014

Student Name:

University ID:

3.1 Use the R hist(), boxplot() and pairs() commands to explore the distributions of each non-categorical attribute in the iris dataset. Comment on what you observe, including outliers in the boxplot and the scatter plots of the attribute values.

Mark:	*None —— Basic —— Good —— Excellent*
Comments:	

3.2 Import the dataset irisMissing.csv into a dataset named irisMissing in your R workspace and use an R command to discover the row numbers of the instances that have missing values.

Mark:	*None —— Basic —— Good —— Excellent*
Comments:	

3.3 Identify an R command that will drop missing values. Apply it to the irisMissing dataset to create a new dataset irisDrop. Briefly describe three other strategies for handling missing values.

Mark:	*None —— Basic —— Good —— Excellent*
Comments:	

3.4 Identify or write your own R functions to implement each of these three strategies.

Mark:	*None* —— *Basic* —— *Good* —— *Excellent*
Comments:	

3.5 Write an R function foo() that takes a dataset and a missing value function as arguments and returns a new dataset with the missing values replaced with values as determined by the missing value function.

Mark:	*None* —— *Basic* —— *Good* —— *Excellent*
Comments:	

3.6 Use the hist() and boxplot() commands to compare results of applying each missing value strategy. Based on this, comment on their relative merits.

Mark:	*None* —— *Basic* —— *Good* —— *Excellent*
Comments:	

I confirm that the work that has been marked is my own and understand that cases of plagiarism will be subject to Departmental and University regulations.

Student Signature **Marker Signature**