**Assignment 1**

**Due Date: Tuesday*, January 19th, 2015, by midnight***

**Total number of points: 25**

This assignment covers the topics from Lectures #1 and #2.

**Problem 1 (5 points):** Differentiate between the following terms:

1. a. classification and clustering
2. b. classification and prediction
3. c. data warehouse and database
4. d. data mining and OLAP
5. e. machine learning and statistics

**Problem 2 (5 points):**

Discuss whether or not each of the following activities is a data mining task.

(a) Monitoring the heart rate of a patient for abnormalities.

(b) Computing the total sales of a company.

(c) Sorting a student database based on student identification numbers.

(d) Predicting the outcomes of tossing a (fair) pair of dice.

(e) Monitoring seismic waves for earthquake activities.

**Problem 3**: (15 points) Fisher's iris data (download the IRIS dataset from <http://archive.ics.uci.edu/ml/datasets/Iris>) consists of measurements on the sepal length, sepal width, petal length, and petal width of 150 iris specimens. There are 50 specimens from each of three species.

Use SPSS to answer the following questions:

a. Visualize and interpret the relationship between the two sepal variables, sepal length and sepal width. Provide the scatterplot that you created to visualize the data along with your interpretation. When you plot the data, you may want to use different colors/signs for representing the data points belonging to the different three class species. Do you think that a classification algorithm will be successful in classifying the data with respect to these two variables? Justify your answer.

b. Repeat part a. for the petal variables.

c. Draw the histograms of the four variables and interpret the distributions of each one of the four variables.

d. Determine if there are any outliers in the data with respect to the sepal length.

e. Repeat d. for the petal length.

**Submission Instructions**

1. Answer the problems and write your answers in a Word document.
2. Submit your file online at the website at <http://d2l.depaul.edu> and check your submission
3. Keep a copy of all your submissions!
4. If you have questions about the homework, email me BEFORE the deadline.
5. The assignment will lose 25%, if submitted after the due date.
6. Assignments submitted one week after the due date will not be accepted.