Intelligent Data Analysis – Fall 2015
Homework #1
Due Date: Sept. 15th, 2015, 11:59PM

The attached excel file contains the data to be used for all the questions listed below. The four columns of the data file are: student id, points obtained in Physics test, points obtained in Maths test, and the letter grade obtained in C++ test. For each problem listed below clearly describe and include with submission: (i) a brief list of steps followed by you to solve the problem, (ii) any matlab/python/other code used by you to get the results, and (iii) the results obtained.

1. (12) Add the two scores for each student and write the sum in a new column. Discretize the sum column into five groups using equal width partitioning and assign one of the five grades (A, B, C, D, and F) to students in the five groups (the highest scores get A and the lowest scores get F). Show the grades assigned to each student in a list sorted according to student id. List the counts of each letter grade awarded.
2. (12) Repeat problem #1 above with the difference that this time use equal frequency partitioning to discretize the sum of points obtained. List the counts of each letter grade awarded. List student ids of those students who would be happier with equal width binning and also of those students' ids who would be happier with equal frequency binning.
3. (12) Compare the grades assigned in problem #1 and #2 above. Make a list of those student ids whose grades changed when the method of binning changed.
4. (12) Convert the Physics and Maths points to their equivalent z-scores in each column. Sum the two z-scores for each student and use equal frequency binning to create five bins. Assign grades to the students and show them in a list sorted by student ids.
5. (12) Compare the grades obtained in #4 and in #2 above. Make a list of students who would be happier with the method in #2 and also a list of those who would be happier with the method in #4.
6. (12) Consider a student who is happier with the method in #4 compared to the method in #2. Briefly explain why his being happier is justified.
7. (12) For Physics and Maths scores individually, perform the following. Use z-scores to assign label "Low" to those students whose z-score is strictly below -0.3, the label "Mid" to those whose z-score is between -0.3 and +0.3 (both values inclusive), and the label "High" to those students whose z-score is strictly greater than 0.3. Show the data table sorted according to student ids.
8. (16) We want to use the labels assigned in #7 above for Physics and Maths scores to predict the letter grade a student would obtain in his/her C++ class. Use the entropy method to determine which course's label (Physics or Maths) is a better predictor of a student's grade in C++ class. Show all your work to arrive at your answer.