**ISyE 6404 - Take-Home Exam #3**

**This is an individual-student exam. Students can discuss about software usage, but not solution for exams. Your report needs to include description of problem/data, source of data, R-codes and study-results (summary table and figures) and -comments.**

1. **Wavelets (50%):**

Locate *a one-dimensional data set* that has sharp-changes like those presented in the recent lectures for applying the following wavelet procedures. It is best that the data size is larger than 512, and is in 2-factorial, e.g., 2*10* = 1024. Note that you need to locate proper R-package/codes to perform the tasks below.

1. Select two families of wavelets for completing the two tasks ii) and iii) below, and make comparisons for the results impacted from distinct wavelet families.
2. Show a multi-resolution plot of mother and father discrete wavelet-coefficients (DWTs), and make comments about their values.
3. Apply **two** thresholding/shrinkage methods to reduce number of non-zero DWT coefficients. Apply the IDWT to *reconstruct* the original data signal by using the thresholded DWTs. Comment on the quality of the reconstructions from plotting the original and reconstructed signals (by overlaying them in one figure). Compare the two methods about their data-reduction ratios and the MSEs.
4. Artificially alter the data in one “local-region” and one large-size “global region” for creating 3 distinct “fault-class” data sets. See lecture presentation about details of this task. Apply the best thresholding method (and the best wavelet-family) to model the data from all FOUR classes (one from the original data and the other 3 fault-class data). Use the multi-resolution plots of thresholded-DWTs to see how they are different in these FOUR classes of data signals.
5. Discuss the possibility and steps for developing a rigorous decision-making procedure to detect faulty-signals against the original data, and distinguish classes of faulty signals with reduced-size data presented by thresholded DWTs.
6. **Categorical Data Analysis (25%):**

Chapter 9 of the textbook on categorical data analysis includes 6 sections with various problems/data and methods. Locate **three sets** of problems/data matching three distinct methods taught in lectures. Apply proper statistical software (preferred to be in R-codes) to analyze the data. Provide in-depth comments about the findings in your statistical analyses.

1. **Nonparametric Regression (25%):**

Locate *a data set* suitable for **both** kernel and spline regressions. The data should include at least 3 x-variables. It is okay to focus on **additive-models** discussed in lectures.

1. Go through one-variable-at-a-time kernel- and spline-regression fits to the data for all 3 x-variables. Compare the fitting results using the leave-one-out cross-validation procedure.
2. Select 2 sets of x-locations, e.g., (1st set: x1 = 3, x2 = 5, x3 = 2, where 3, 5, 2 are values within x-data range). These 2 sets of x-locations should be from a location close to the center of x-data-range and another location closer to the edge. Make predictions of Y at these x-data. Compare the predictions from Kernel and spline methods, and also comment on the impact from x-data-locations.
3. Construct the 90% Bias-Corrected Bootstrap CIs for the predictions at the selected 2-sets of x-locations. Show details of the bias-correction process. Compare the CIs at two x-locations, and also from two nonparametric regression methods.