Assignment 1: CS 215

Due: 10th August before 11:55 pm

Remember the honor code while submitting this (and every other) assignment. All members of the group should work on all parts of the assignment. We will adopt a zero-tolerance policy against any violation.

Submission instructions:

1. You should type out all the answers to the written problems in Word (with the equation editor) or using Latex, or write it neatly on paper and scan it. In either case, prepare a pdf file.

2. Put the pdf file and the code for the programming parts all in one zip file. The pdf should contain the names and ID numbers of all students in the group within the header. The pdf file should also contain instructions for running your code. Name the zip file as follows: A1-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A1-IdNumber.zip).

3. Upload the file on moodle BEFORE 11:55 pm on the due date (i.e. 10th August). We will nevertheless allow and not penalize any submission until 2:00 am on the following day (i.e. 11th August). No assignments will be accepted thereafter.

4. Note that only one student per group should upload their work on moodle.

5. Please preserve a copy of all your work until the end of the semester.

Questions:

1. Given n values $\{x_i\}_{i=1}^n$ having mean μ, median ν and standard deviation σ, prove that $|\mu - \nu| \leq \sigma$. Assume n is even. [10 points]

2. Consider four sets of n values each: $\{x_i\}_{i=1}^n, \{y_i\}_{i=1}^n, \{z_i\}_{i=1}^n, \{w_i\}_{i=1}^n$. Consider for all $i, 1 \leq i \leq n$, we have $z_i = ax_i + b$ and $w_i = cy_i + d$ where $a \neq 0, c \neq 0$. Here a, b, c, d are constants. Then prove that $r(z, w) = \pm r(x, y)$ where r is the correlation coefficient. Comment on when the we would have $r(z, w) = r(x, y)$ and when we would have $r(x, y) = -r(z, w)$. [10 points]

3. Given n distinct values $\{x_i\}_{i=1}^n$ with mean μ and standard deviation σ, prove that for all i, we have $|x_i - \mu| \leq \sigma\sqrt{n-1}$. [10 points]

   In the following problems, you can use the mean, median and standard deviation functions from MATLAB.

4. Generate a sine wave in MATLAB of the form $y = 5\sin(2x + \pi/3)$ where x ranges from -10 to 10 in steps of 0.02. Now randomly select 40 values in the array y (using MATLAB function 'randperm') and corrupt them by adding random values from 5 to 10 using the MATLAB function 'rand'. This will generate a corrupted sine wave which we will denote as z. Now your job is to filter z using the following steps.

   - Create a new array y_{median} to store the filtered sine wave.
   - For a value at index i in z, consider a neighborhood $N(i)$ consisting of $z(i)$, 8 values to its right and 8 values to its left. For indices near the left or right end of the array, you may not have 8 neighbors in one of the directions. In such a case, the neighborhood will contain fewer values.

- Set $y_{median}(i)$ to the median of all the values in $N(i)$. Repeat this for every i.

This process is called as 'moving median filtering', and will produce a filtered signal in the end. Repeat the entire procedure described here using the arithmetic mean instead of the median. This is called as 'moving average filtering'. Plot the original (i.e. clean) sine wave y, the corrupted sine wave z and the filtered sine wave using mean and median on the same figure in different colors. Introduce a legend on the plot (find out how to do this in MATLAB). Include an image of the plot in your report. Now compute and print the relative mean squared error between each result and the original clean sine wave. The relative mean squared error between y and its estimate $\hat{y}$ is defined as $\dfrac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i y_i^2}$.

Now repeat all the steps above when the random values to corrupt the sine wave lay in a range from 100 to 120, and include the plot of the sine waves in your report, and write down the relative mean square error values.

Which method (median or arithmetic mean) produced better relative mean squared error? Why? Explain in your report. [6+3+3+3=15 points]

5. Suppose that you have computed the mean, median and standard deviation of a set of n numbers stored in array A where n is very large. Now, you decide to add another number to A. Write a MATLAB function to update the previously computed mean, another MATLAB function to update the previously computed median, and yet another MATLAB function to update the previously computed standard deviation. Note that you are <u>not</u> allowed to simply recompute the mean and standard deviation by looping through all the data. You may need to derive a formula for this. Include the formula and its derivation in your report. Your MATLAB functions should be of the form function newMean = UpdateMean (OldMean, NewDataValue, A, N), function newMedian = UpdateMedian (oldMedian, A, N) and function newStd = UpdateStd (OldMean, OldStd, NewMean, NewDataValue, A, N). [5+5+5 = 15 points]