Assignment 1: CS 215

Akash Trehan-150050031, Bhavya Bahl-150050110

August 11, 2016

$$\mu = Mean, \ \nu = Median, \ \sigma = Standard \ Deviation$$

Ans 1.

$$|\mu - \nu| = \left| \frac{\sum\limits_{i=1}^{n} x_i}{n} - \nu \right|$$

$$= \left| \frac{\sum\limits_{i=1}^{n} x_i - n\nu}{n} \right|$$

$$= \left| \frac{\sum\limits_{i=1}^{n} (x_i - \nu)}{n} \right|$$

$$\leq \frac{\sum\limits_{i=1}^{n} |x_i - \nu|}{n} \ (Extended \ Triangle \ Inequality)$$

$$\leq \frac{\sum\limits_{i=1}^{n} |x_i - \mu|}{n} \ (\because Absolute \ deviation \ is \ minimum \ for \ median)$$

$$\leq \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \mu)^2}{n}} \ (R.M.S \geq A.M)$$

$$\leq \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \mu)^2}{n - 1}} = \sigma$$

$$Thus \ |\mu - \nu| \leq \sigma$$

Hence proved.

Ans 2.

$$z_i = ax_i + b, a \neq 0$$

$$w_i = cy_i + d, c \neq 0$$

$$So\ \mu_z = a\mu_x + b\ and\ \mu_w = c\mu_y + d$$

$$r(z, w) = \frac{\sum\limits_{i=1}^{n}(z_i - \mu_z)(w_i - \mu_w)}{\sqrt{\sum\limits_{i=1}^{n}(z_i - \mu_z)^2 \sum\limits_{i=1}^{n}(w_i - \mu_w)^2}}$$

$$Now\ z_i - \mu_z = ax_i + b - (a\mu_x + b) = a(x_i - \mu_x)\ \forall i,\ 1 \leq i \leq n$$

$$Similarly\ w_i - \mu_w = c(y_i - \mu_y)\ \forall i,\ 1 \leq i \leq n$$

$$Substituting\ values\ for\ (z_i - \mu_z)\ and\ (w_i - \mu_w), we\ have$$

$$r(z, w) = \frac{\sum\limits_{i=1}^{n}(a(x_i - \mu_x))(c(y_i - \mu_y))}{\sqrt{\sum\limits_{i=1}^{n}(a(x_i - \mu_x))^2 \sum\limits_{i=1}^{n}(c(y_i - \mu_y))^2}}$$

$$= \frac{ac}{|ac|} \frac{\sum\limits_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \mu_x)^2 \sum\limits_{i=1}^{n}(y_i - \mu_y)^2}}$$

$$= \frac{ac}{|ac|} r(x, y) = \pm r(x, y)\ (Since\ a \neq 0\ and\ c \neq 0)$$

Hence proved.

Now, $\frac{ac}{|ac|} = 1$ when a and c have the same sign and $\frac{ac}{|ac|} = -1$ when a and c have opposite signs.

Thus r(z,w)=r(x,y) when a and c have same sign and r(z,w)=-r(x,y) when a and c have opposite signs.

Ans 3.

$$To\ prove:\ |x_i - \mu| \leq \sigma\sqrt{n-1}\ \forall i$$

By definition of standard deviation

$$\sigma = \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{n-1}}$$

$$\Rightarrow\ \sigma^2 = \frac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{n-1}$$

$$\Rightarrow \ \sigma^2 \geq \frac{(x_i - \mu)^2}{n - 1}, \forall i, 1 \leq i \leq n \ (\because \ all \ terms \ are \ positive)$$

$$\Rightarrow \ (n - 1)\sigma^2 \geq (x_i - \mu)^2 \ (\because \ n - 1 > 0)$$

Both sides are positive, so taking square root on both sides we have,

$$\sigma\sqrt{n - 1} \geq |x_i - \mu|$$

Hence proved.

Instructions for running code

1. For Ques 4

   - For first part of the question with random values added between 5 and 10 run file task4_1.m in MATLAB.
   - For first part of the question with random values added between 100 and 120 run file task4_2.m in MATLAB.
   - The relative mean squared error will be displayed on the console.

2. For Ques 5

   - Add the files UpdateMedian.m, UpdateMean.m and UpdateStd.m to your Matlab working directory.
   - Then call the function UpdateMean(OldMean, NewDataValue, A, N) with the Old mean, New value added, array A and the number of items originally in A respectively. This function returns the new mean of the data set.
   - Call the function UpdateMedian(oldMedian, NewDataValue, A, N) with the Old median, New value added, array A and the number of items originally in A respectively. The function returns the new median of the data set.
   - Call the function UpdateStd(OldMean, OldStd, NewMean, NewDataValue, A, N) with the Old mean, Old standard deviation, New mean, New value added, array A and the number of items originally in A respectively. The function returns the new Standard deviation of the data set.

Ans 4.

Moving **Median** filtering produced better relative mean squared error than Moving **Mean** filtering. The reason is that **Median is robust to outliers**. This means that the points far away from the mean do not have much effect on the median. This is because the middle element of the set either remains the same or changes to an element closeby. Mean on the other hand depends on the sum of all elements and outliers can have a huge effect on this sum. The graphs on the next page give a really good demonstration of this.
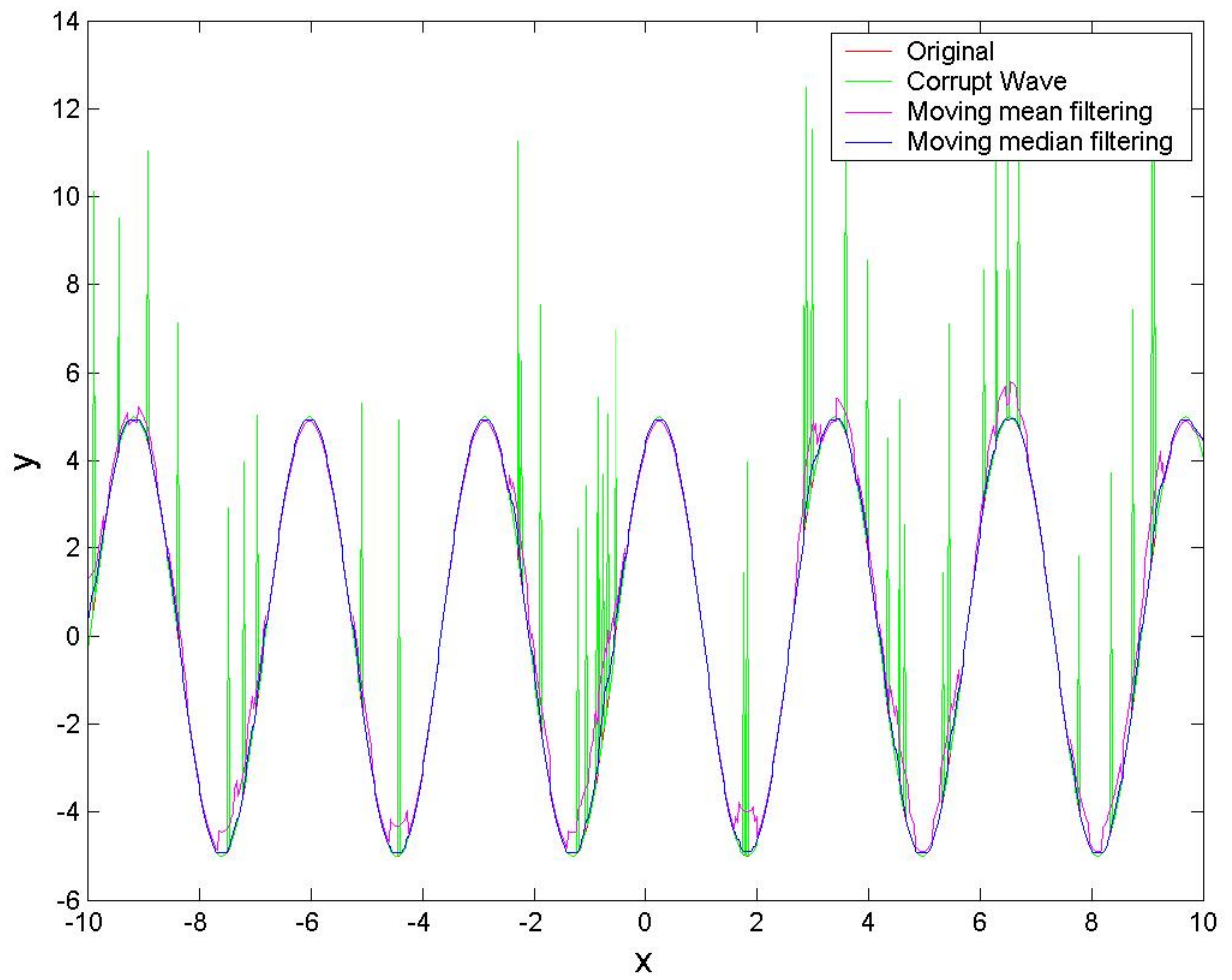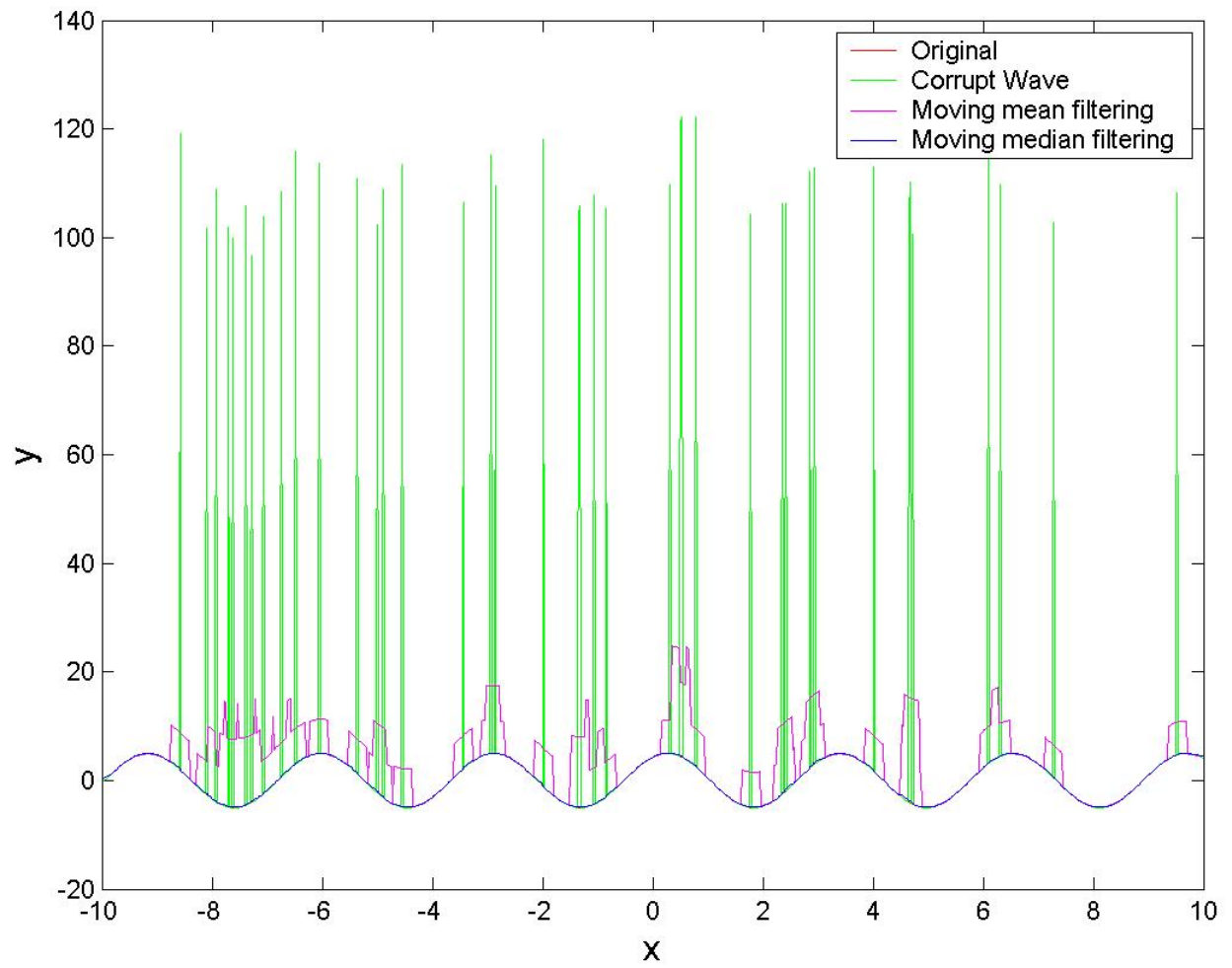
Figure 1: Task 4(1)

Figure 2: Task 4(2)

Ans 5.

Formula for Updating Mean

$$newMean = \frac{OldMean * n + NewDataValue}{n+1}$$

Derivation

$$OldMean = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\Rightarrow \sum_{i=1}^{n} x_i = OldMean * n$$

Now adding new value,

$$\sum_{i=1}^{n+1} x_i = OldMean * n + NewDataValue$$

$$\Rightarrow newMean = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{OldMean * N + NewDataValue}{n+1}$$

Formula for Updating Standard Deviation

$$newStd = \sqrt{\frac{(n-1)oldStd^2 + NewDataValue^2 + n\mu_{old}^2 - (n+1)\mu_{new}^2}{n}}$$

Derivation

$$oldStd = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_{old})^2}{n-1}}$$

$$newStd = \sqrt{\frac{\sum_{i=1}^{n+1} (x_i - \mu_{new})^2}{n}}$$

$$oldStd = \sqrt{\frac{\sum_{i=1}^{n} (x_i^2 + \mu_{old}^2 - 2x_i\mu_{old})}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} x_i^2 + n\mu_{old}^2 - 2\mu_{old}\sum_{i=1}^{n} x_i}{n-1}}$$

$$= \sqrt{\frac{\sum\limits_{i=1}^{n} x_i^2 + n\mu_{old}^2 - 2n\mu_{old}^2}{n-1}}$$

$$= \sqrt{\frac{\sum\limits_{i=1}^{n} x_i^2 - n\mu_{old}^2}{n-1}}$$

$$Similarly\ newStd = \sqrt{\frac{\sum\limits_{i=1}^{n+1} x_i^2 - (n+1)\mu_{new}^2}{n}}$$

$$= \sqrt{\frac{\sum\limits_{i=1}^{n} x_i^2 + NewDataValue^2 - (n+1)\mu_{new}^2}{n}}$$

$$Now\ using\ oldStd\ we\ see \sum_{i=1}^{n} x_i^2 = (n-1)oldStd^2 + n\mu_{old}^2$$

$$newStd = \sqrt{\frac{(n-1)oldStd^2 + NewDataValue^2 + n\mu_{old}^2 - (n+1)\mu_{new}^2}{n}}$$