DATA MINING

Exercise 1

Name: Chethan Kashyap Bangalore Muralidhara.

Date: 02/11/2023

Completed Exercises: 1,2,3,4,5,6(All)

Question 1:

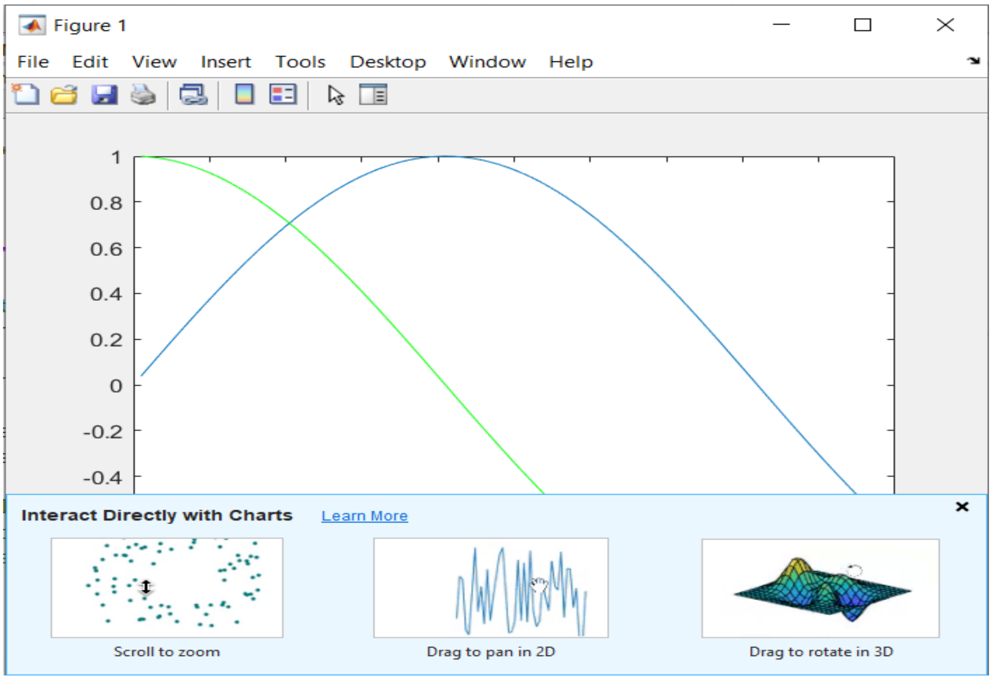
Code file is attached as well.

```
%Numbers from 1 to 100.
t1=1:100

%Question 1
%Numbers from 100 to 1
for i = 100:-1:1
    disp(i);
end

%Numbers from 1 to 100 with interval of 2
for i = 1:2:100
    disp(i);
end
```

Question 2:

```
%Question 2
%Dimensions of "Matrix t". Even a scalar is matrix in Matlab (Matrix laboratory)
l=length(t1)
d=size(t1)

%Transpose of t.
tp=t1'
size(tp)

%100 samples from 50Hz sinusoidal signal sampled at frequency of 8192Hz.
f=50
F=8192
y=sin(2*pi*t1*(f/F))
plot(t1,y)

%Another signal into the same plot.

y2=cos(2*pi*t1*(f/F))

plot(t1,y)
hold on
plot(t1,y2,'g')
```

```
%finding intersection
intersection_indices = find(abs(y - y2) < 0.01);  % You can adjust the tolerance as needed

intersection_points = t1(intersection_indices);

disp(intersection_indices);
disp(intersection_points);

%theoretically(manually) we can't find the intersection of these two waves as both
%sine and cosine waves are infinite series,but
%visually we can see the intersection when x=20 and y=0.720003
```



As we can from the graph, visually the intersection can be found out and it is at x=20 and y=0.720003. Whereas when we try to find the exact intersection manually (theoretically) it is not possible as the sine and waves are infinite.

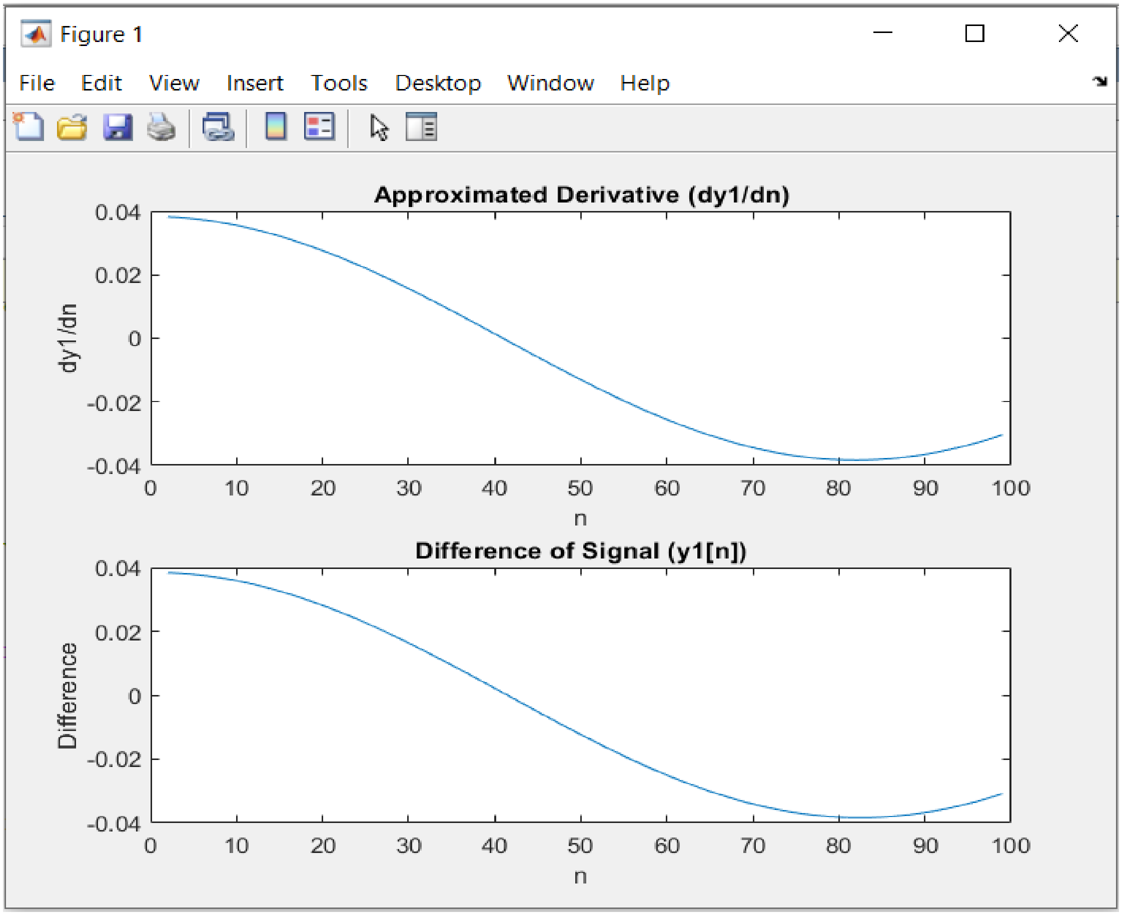
Question 3:

```
%Question 3
% Compute the derivative of y1 with respect to n
n = 2:99;
dy1_dn_approx = (y(n + 1) - y(n - 1)) / 2;

% Compute the difference of y1[n]
diff_y1 = diff(y);

% Plot the derivative and difference vectors using their valid indices
subplot(2, 1, 1);
plot(n, dy1_dn_approx);
title('Approximated Derivative (dy1/dn)');
xlabel('n');
ylabel('dy1/dn');

subplot(2, 1, 2);
plot(n, diff_y1(1:end-1));  % Adjust for the length
title('Difference of Signal (y1[n])');
xlabel('n');
ylabel('Difference');

% Calculate the maximum absolute difference between the two approaches
max_abs_diff = max(abs(dy1_dn_approx - diff_y1(1:end-1)));

fprintf('Maximum absolute difference between the two approaches: %.5f\n', max_abs_diff);
```

Command Window

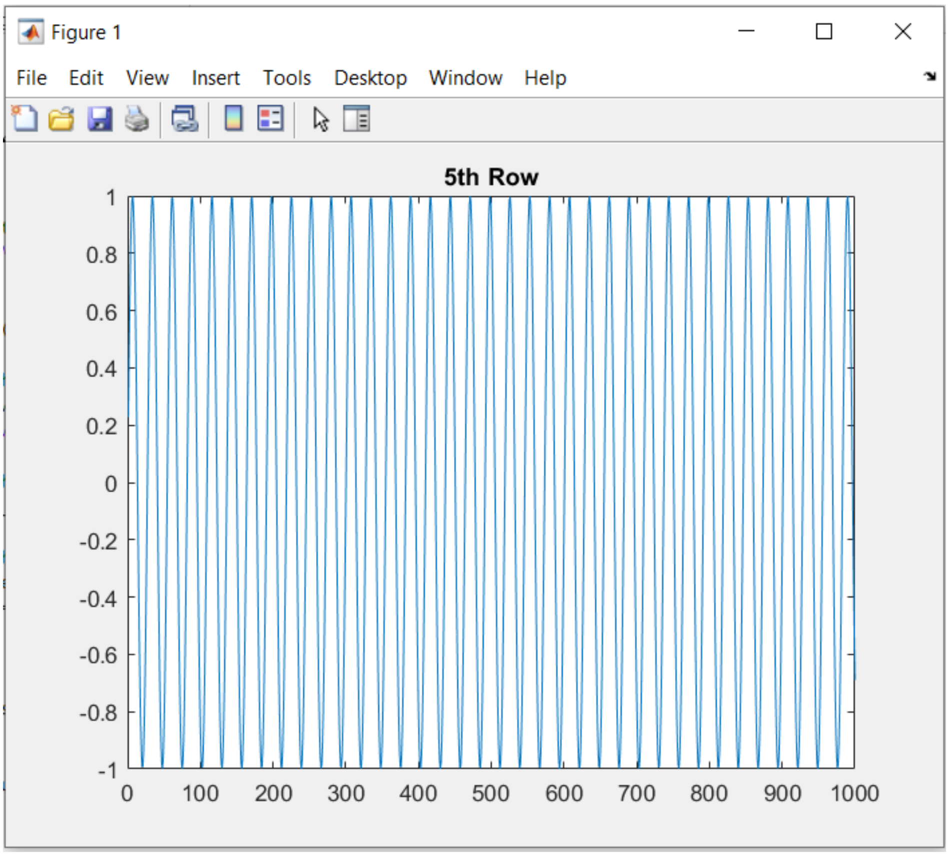
New to MATLAB? See resources for Getting Started.

```
% Compute the derivative of y1 with respect to n
n = 2:99;
dy1_dn_approx = (y(n + 1) - y(n - 1)) / 2;

% Compute the difference of y1[n]
diff_y1 = diff(y);

% Plot the derivative and difference vectors using their valid indices
subplot(2, 1, 1);
plot(n, dy1_dn_approx);
title('Approximated Derivative (dy1/dn)');
xlabel('n');
ylabel('dy1/dn');

subplot(2, 1, 2);
plot(n, diff_y1(1:end-1));  % Adjust for the length
title('Difference of Signal (y1[n])');
xlabel('n');
ylabel('Difference');

% Calculate the maximum absolute difference between the two approaches
max_abs_diff = max(abs(dy1_dn_approx - diff_y1(1:end-1)));

fprintf('Maximum absolute difference between the two approaches: %.5f\n', max_abs_diff);
Maximum absolute difference between the two approaches: 0.00074
```



Question 4:

```
%Question 4
% Load the data from the file "Matrix.txt"
A = load('D:\MS\Data Mining\1\Matrix.txt');

% Extract the 5th row (remember that MATLAB uses 1-based indexing)
fifth_row = A(5, :);

% Plot the 5th row
plot(fifth_row);
title('5th Row');

% Listen to the content of the 5th row
soundsc(fifth_row);

% Calculate the means
mean_rows = mean(A, 2); % Mean of each row
mean_columns = mean(A, 1); % Mean of each column
mean_matrix = mean(A(:)); % Mean of the entire matrix

fprintf('Mean of rows:\n');
disp(mean_rows);

fprintf('Mean of columns:\n');
disp(mean_columns);

fprintf('Mean of the entire matrix: %.4f\n', mean_matrix);
```



```
Mean of rows:
    0.0056
    0.0100
    0.0658
    0.0003
    0.0071

Mean of columns:
  Columns 1 through 12

    0.2120    0.3438    0.3587    0.2832    0.1907    0.1596    0.2288    0.3756    0.5278    0.6049    0.5630    0.4206

  Columns 13 through 24

    0.2486    0.1321    0.1235    0.2146    0.3423    0.4250    0.4094    0.3007    0.1607    0.0733    0.0973    0.2334

  Columns 25 through 36

    0.4233    0.5813    0.6416    0.5939    0.4881    0.4048    0.4096    0.5148    0.6717    0.7977    0.8213    0.7229

  Columns 37 through 48
```

```
  Columns 985 through 996

    0.2070    0.2183    0.1372    0.0356   -0.0070    0.0505    0.1875    0.3336    0.4073    0.3628    0.2160    0.0363

  Columns 997 through 1,000

   -0.0909   -0.1111   -0.0299    0.0912

Mean of the entire matrix: 0.0178
```

Question 5:

```
>> %Question 5
% Calculate the mean amplitude for each row
mean_amplitudes = mean(A, 2);

% Find the row with the lowest mean amplitude
[lowest_amplitude, lowest_row] = min(mean_amplitudes);

fprintf('The lowest sounding signal is in row %d with a mean amplitude of %.4f\n', lowest_row, lowest_amplitude);
The lowest sounding signal is in row 4 with a mean amplitude of 0.0003
```

Question 6:

```
%Question 6
data = readtable('D:\MS\Data Mining\1\inco13par.txt');
% Check the number of cases and variables
[num_cases, num_vars] = size(data);
fprintf('Number of cases: %d\n', num_cases);
fprintf('Number of variables: %d\n', num_vars);

% Initialize a matrix to store missing value counts
missing_values_per_variable = zeros(1, num_vars);

% Convert the data to numeric format and count missing values
for i = 1:num_vars
    variable_data = str2double(data(:, i));
    missing_values_per_variable(i) = sum(isnan(variable_data));
end

fprintf('Number of missing values for each variable:\n');
disp(missing_values_per_variable);
```

```
fprintf('Number of missing values for each variable:\n');
disp(missing_values_per_variable);
Number of cases: 529
Number of variables: 16
Number of missing values for each variable:
     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1     1
```