
Example of proteomics in cancer biology with Matlab

Table of Contents

Summary	1
Loading pre-processed dataset	1
Initializing variables	2
Visualizing a set of the samples	3
Ranking features	4

Summary

This report represents a represents an example Matlab proteomic data analysis. The data set analyze in this report can be found [here](#) which is the FDA-NCI Clinical Proteomics Program Databank. The samples downloaded from the FDA-NIC Proteomics Programa Databank corresponds to SELDI Mass-Spec profiles of overian cancer samples: **Cancer Group** vs **Normal Group**. The study related to this dataset was published, in 2004, the Endocrine Related Cancer journal.

Loading pre-processed dataset

After preprocessing the dataset (find the code [here](#), the dataset was loaded.

```
load OvarianCancerQAQCdataset
whos
```

<i>Name</i>	<i>Size Attributes</i>	<i>Bytes</i>	<i>Class</i>
<i>Cidx</i>	<i>216x1</i>	<i>216</i>	<i>logical</i>
<i>Cvec</i>	<i>121x1</i>	<i>968</i>	<i>double</i>
<i>MZ</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>N</i>	<i>1x1</i>	<i>8</i>	<i>double</i>
<i>Nidx</i>	<i>216x1</i>	<i>216</i>	<i>logical</i>
<i>NumberCancerDatasets</i>	<i>1x1</i>	<i>8</i>	<i>double</i>
<i>NumberNormalDatasets</i>	<i>1x1</i>	<i>8</i>	<i>double</i>
<i>Nvec</i>	<i>95x1</i>	<i>760</i>	<i>double</i>
<i>Y</i>	<i>15000x216</i>	<i>25920000</i>	<i>double</i>

<i>ans</i>	<i>1x10</i>	<i>80</i>	<i>double</i>
<i>feat</i>	<i>100x1</i>	<i>800</i>	<i>double</i>
<i>files</i>	<i>1x216</i>	<i>32400</i>	<i>cell</i>
<i>filesCancer</i>	<i>121x1</i>	<i>113519</i>	<i>struct</i>
<i>filesNormal</i>	<i>95x1</i>	<i>89209</i>	<i>struct</i>
<i>grp</i>	<i>216x1</i>	<i>26784</i>	<i>cell</i>
<i>hC</i>	<i>10x1</i>	<i>0</i>	
<i>matlab.graphics.chart.primitive.Line</i>			
<i>hN</i>	<i>10x1</i>	<i>0</i>	
<i>matlab.graphics.chart.primitive.Line</i>			
<i>max_C</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>max_N</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>mean_C</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>mean_N</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>min_C</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>min_N</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>repository</i>	<i>1x85</i>	<i>170</i>	<i>char</i>
<i>repositoryC</i>	<i>1x92</i>	<i>184</i>	<i>char</i>
<i>repositoryN</i>	<i>1x92</i>	<i>184</i>	<i>char</i>
<i>sig_Masses</i>	<i>100x1</i>	<i>800</i>	<i>double</i>
<i>stat</i>	<i>15000x1</i>	<i>120000</i>	<i>double</i>
<i>xAxisLabel</i>	<i>1x17</i>	<i>34</i>	<i>char</i>
<i>yAxisLabel</i>	<i>1x13</i>	<i>26</i>	<i>char</i>

Initializing variables

A set of vector variables, which will be used in the downstream workflow, are initialized.

```

N = numel(grp);           % vector of number of samples
Cidx = strcmp('Cancer',grp); % logical index vector for
    Cancer samples
Nidx = strcmp('Normal',grp); % logical index vector for
    Normal samples

```

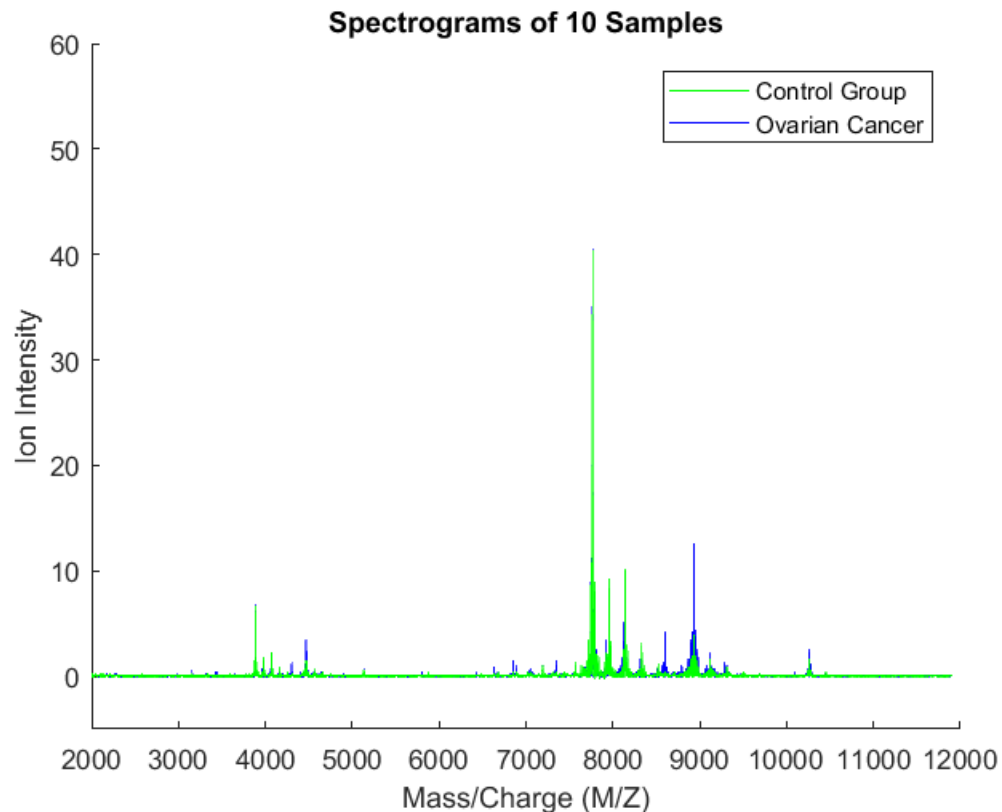
```
Cvec = find(Cidx); % index vector for Cancer
samples
Nvec = find(Nidx); % index vector for Normal
samples
xAxisLabel = 'Mass/Charge (M/Z)'; % x-axis label for plots
yAxisLabel = 'Ion Intensity'; % y-axis label for plots
```

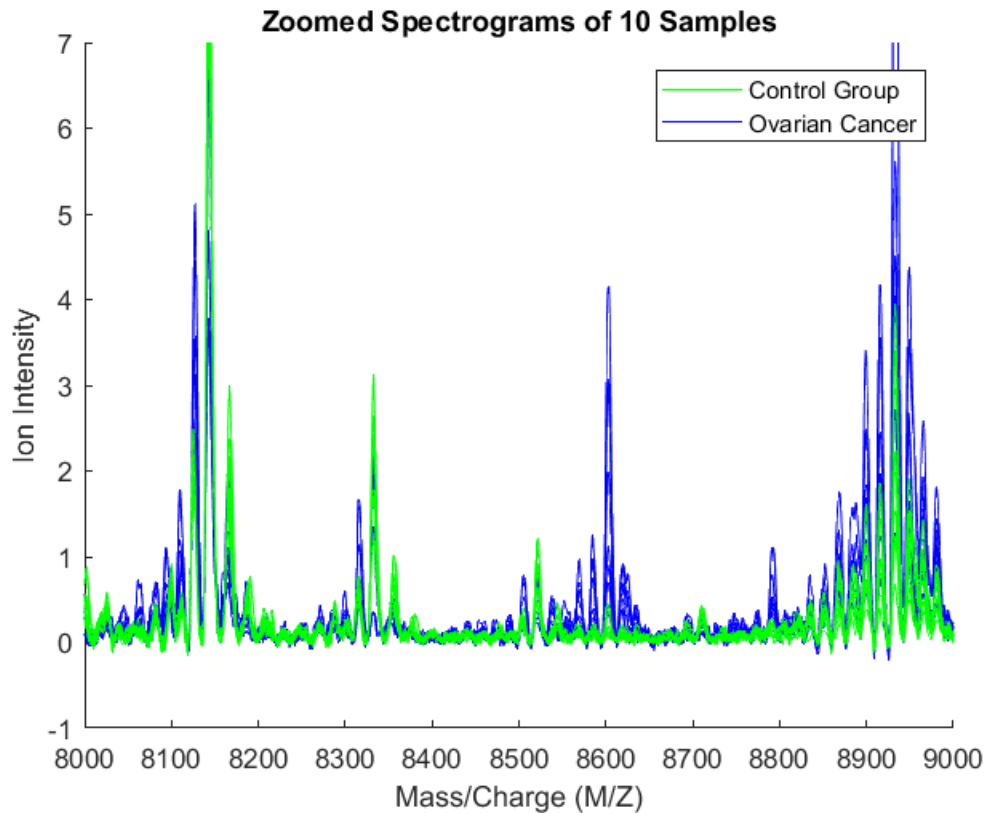
Visualizing a set of the samples

The spectrogram of 10 samples

```
figure; hold on;
hC = plot(MZ,Y(:,Cvec(1:10)),'b');
hN = plot(MZ,Y(:,Nvec(1:10)),'g');
xlabel(xAxisLabel); ylabel(yAxisLabel);
axis([2000 12000 -5 60])
legend([hN(1),hC(1)],{'Control Group','Ovarian Cancer'})
title('Spectrograms of 10 Samples')
```

```
figure; hold on;
hC = plot(MZ,Y(:,Cvec(1:10)),'b');
hN = plot(MZ,Y(:,Nvec(1:10)),'g');
xlabel(xAxisLabel); ylabel(yAxisLabel);
axis([8000 9000 -1 7])
legend([hN(1),hC(1)],{'Control Group','Ovarian Cancer'})
title('Zoomed Spectrograms of 10 Samples')
```





Ranking features

```
% Significant masses using a two-way t-statistic
[feat,stat] = rankfeatures(Y,grp, 'CRITERION', 'ttest', 'NUMBER',100);
sig_Masses = MZ(feat);
sig_Masses(1:10)' %display the first 10 significant masses

% Set variables for two-way t-statistic plotting
mean_N = mean(Y(:,Nidx),2); % group average for control samples
max_N = max(Y(:,Nidx),[],2); % top envelopes of the control samples
min_N = min(Y(:,Nidx),[],2); % bottom envelopes of the control samples
mean_C = mean(Y(:,Cidx),2); % group average for cancer samples
max_C = max(Y(:,Cidx),[],2); % top envelopes of the control samples
min_C = min(Y(:,Cidx),[],2); % bottom envelopes of the control samples

% Plotting the two-way t-statistic
figure;

yyaxis left
plot(MZ, [mean_N mean_C]);
ylim([-1,20])
xlim([8000,9000])
title('Significant M/Z Values')
xlabel(xAxisLabel);
ylabel(yAxisLabel);
```

```
yyaxis right
plot(MZ,stat);
ylim([-1,22])
ylabel('Test Statistic');
```

```
legend({'Control Group Avg.', 'Cancer Group Avg.', 'Test Statistics'})
```

```
ans =
```

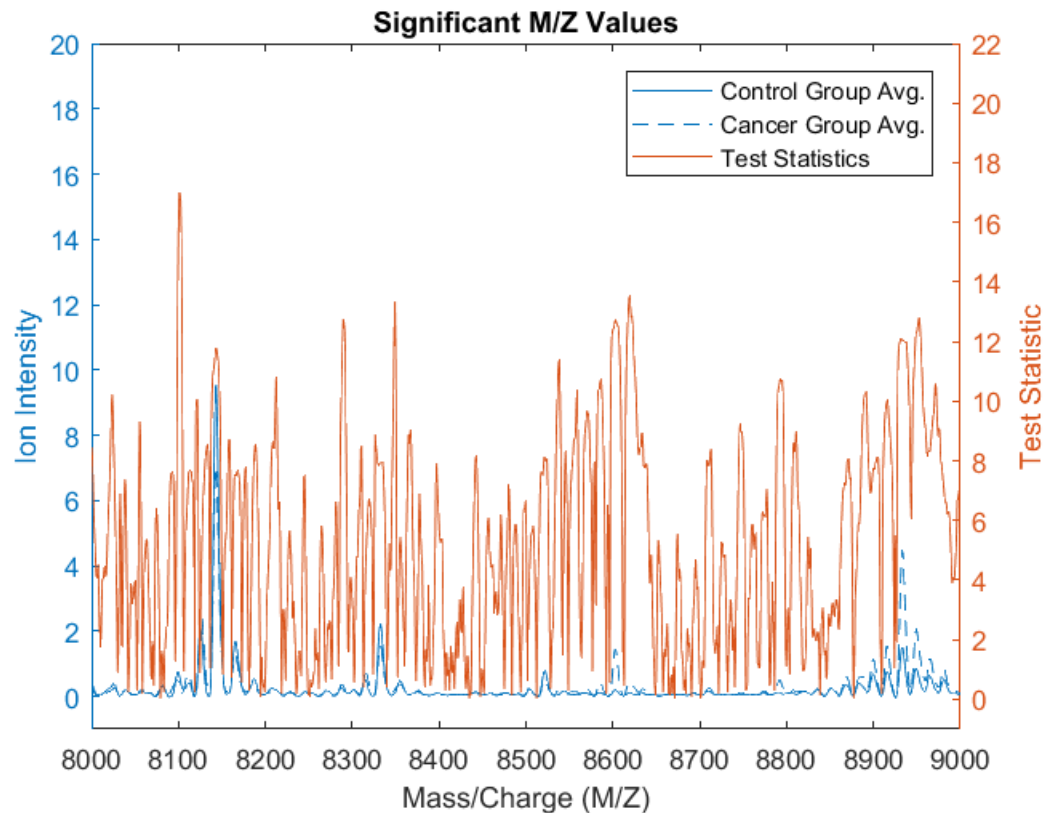
```
1.0e+03 *
```

```
Columns 1 through 7
```

```
8.1009    8.1016    8.1024    8.1001    8.1032    7.7366    7.7359
```

```
Columns 8 through 10
```

```
7.7374    7.7253    7.7245
```



Published with MATLAB® R2017a