This tutorial contains navigation buttons that enable you to move throughout the tutorial.

Please use the navigation buttons and not the page up/page down or arrow keys to navigate through the tutorials.

This is the 'Next' button. It takes you to the next frame or stop point.

This is the 'Previous' button. It takes you to the previous frame or stop point.

This is the 'Go to frame' button. It takes you to a specified frame.

This is the 'Go to URL' button. It takes you to a website link.

Press the 'Next' button below to start this tutorial.
This tutorial will cover how to use the Dice/Classify Image Function from the MVA menu.

This function uses a set of control spectra to classify the data in an image matrix using PCA.

For this a PCA model is created using the control spectra. Then the image matrix is diced into n x n "pixels" and each "pixel" is then projected into the PCA model and classified based on its distance from the mean of a group from the control spectra. If the distance between the projected "pixel" and all control groups is greater than 3 times the standard deviation of the given group it is considered unclassified.

This function uses the scatter value defined in Med Biol Eng Comput (2008) 46:985–995

It also uses the knnsearch function written by Yi Cao at Cranfield University on 25 March 2008 found on the Matlab file exchange.
<table>
<thead>
<tr>
<th>Name of Image Matrix</th>
<th>Name of Variable Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>yimage</td>
<td>exactmass</td>
</tr>
</tbody>
</table>

To use this function the user must first load the image matrix that is to be used.
This data set consists of a spot of a chemical on a substrate. As seen in this image the spot can be easily seen and makes a good controls set for demonstrating the Dice/Classify function.
The data, samplenames and variables in the data matrix for the controls, and the image to be classified must be loaded in the current Matlab workspace in order to use the Dice/Classify Image function.

The image and control data matrices must be created with the same set of peaks.

The image and control data should contain unmodified peak areas (i.e. no normalization or scaling).
As controls for this data set, we have spectra taken from the chemical (On the spot) and from the substrate (Off the spot).
Open the Dice/Classify Image panel by choosing 'Dice/Classify Image' from the MVA menu.
Data Selection Panel

Name of Image Matrix: yimage
Name of Variable Matrix: exactmass

Name of Control Matrix:

Sample Names for Control Matrix:

Choose a normalization method and then a scaling method. If you change the normalization method you need to re-select the chosen scaling method.

Choose Normalization Method:
Choose one...

Choose Scaling Method:
Choose one...

Run PCA

PCA Summary:
PC# | %Var | %VarTotal

# of PCs to keep in model =

Save Model

"Pixel" size of squares nxn, n=

Dice and Model

This is the Dice/Classify Image panel.

This upper part of the panel is used to build the original PCA model based on the control spectra.
Data Selection Panel

Name of Image Matrix: yimage
Name of Variable Matrix: exactmass

Name of Control Matrix

Sample names for Control Matrix

Choose a normalization method and then a scaling method. If you change the normalization method, you need to re-select the chosen scaling method.

Choose Normalization Method
Choose one...

Choose Scaling Method
Choose one...

Run PCA

PCA Summary
PC#  %Var  %Vartotal

# of PCs to keep in model =

Save Model

"Pixel" size of squares nxn, n=

Once the model is created, this part of the panel allows the user to dice and classify the image selected above.
To use the function, first enter the names of the control data and samplenames that are in the current Matlab workspace.
Choose a normalization method. This will be applied to the control data and the image matrix.

Here we will choose to normalize to the 'Sum of Selected Peaks'.

Choose one...

Choose one...

None

Sum of Selected Peaks

Run PCA
Choose a scaling method. This will be applied to the control data and the image matrix.

Here we will choose 'Square root scaling' (takes the square root of the data and then mean centers.)
Press the 'Run PCA' button to create the model based on the control spectra.
This brings up a new box on the panel that allows the user to browse the PCA scores and loadings for the model in order to see which PCs contain useful information and which contain noise.
Choose a PC to plot and press the 'Plot Scores' button to see the scores for the given PC.
PC1 clearly separates the control spectra.
PC2 looks like it is showing scatter in the ‘off’ samples.
To start let's look at how the classification works with a 1 PC model. Enter the number of PCs to keep and press the 'Save Model' button.
### Data Selection Panel

**Name of Image Matrix**
yimage

**Name of Variable Matrix**
exactmass

**Name of Control Matrix**
data

**Sample Names For Control Matrix**
samplenames

Choose a normalization method and then a scaling method. If you change the normalization method you need to re-select the chosen scaling method.

**Choose Normalization Method**
Sum of Selected Peaks

**Choose Scaling Method**
Square root scaling

**Run PCA**

**PCA Summary**

<table>
<thead>
<tr>
<th>PC#</th>
<th>%Var</th>
<th>%Vartotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.9</td>
<td>94.9</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>1.2</td>
<td>99.2</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>99.7</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>99.8</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>99.9</td>
</tr>
<tr>
<td>7</td>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

**# of PCs to keep in model**
1

**Save Model**

**“Pixel” size of squares nxn, n=**

Acceptable pixel sizes include:

1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192

**Dice and Model**

---

Once the model is saved the PCA scores and loadings plot box closes. This is so the classification image can be shown.
For this function a "Pixel" is defined as an n x n set of original pixels that will be summed together and projected into the model.

If a value of 1 is used every pixel in the image is projected into the control model. If a value of 2 is used then the image matrix is subdivided into groups of pixels 2 pixels wide and 2 pixels tall (which are summed together and then projected into the control model). After being classified each original pixel from within the 2 x 2 area is assigned the same classification value.

Using larger "Pixel" sizes speeds up the classification process but degrades the resulting image resolution.

A list of acceptable 'Pixel' sizes is shown here. The pixels must be square so this shows what values can be used for a 'Pixel'.

Enter the desired value and press the 'Dice and Model' button.
Data Selection Panel

Name of Image Matrix: yimage
Name of Variable Matrix: exactmass

Name of Control Matrix: data

Sample names for Control Matrix: samplenames

Choose a normalization method and then a scaling method. If you change the normalization method you need to re-select the chosen scaling method.

Choose Normalization Method:
- Sum of Selected Peaks

Choose Scaling Method:
- Square root scaling

Run PCA

PCA Summary

<table>
<thead>
<tr>
<th>PC#</th>
<th>%Var</th>
<th>%Vartotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.9</td>
<td>94.9</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>1.2</td>
<td>99.2</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>99.7</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>99.8</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>99.9</td>
</tr>
<tr>
<td>7</td>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

# of PCs to keep in model = 1

Save Model

"Pixel" size of squares n x n = 2

Acceptable pixel sizes include:
1 2 4 8 16 32 64 128 256 512
1024 2048 4096 8192

Dice and Model

A progress bar appears and shows the time remaining for the classification.

6% Progress
Classifying "pixels"...
Estimated time remaining: 00:00:00
When the function finishes the calculations the 'classification image' is shown.
In this image light grey corresponds to "Pixels" that were classified with the 'On' control.

Dark grey corresponds to "Pixels" that classified with the 'Off' control.

White correspond with "Pixels" that were not classified with any of the controls. This means the distance of the "Pixel" to the model groups was greater than the matching tolerance (3 standard deviations from the mean of the group).
As can be seen the classification is good, but let's see what happens if we change the number of PCs in the PCA model.
Let’s see what a 2 PC model does to the classification.

Enter 2 in the “Pixel” box and press the “Dice and Model” button.
The classification is redone and the new classification image is shown.

As can be seen the classification has improved.

Of course one would have to decide which model is more realistic. Saving more PCs in a model can improve the classification, but can result in overfitting the data and creating a model that is too specific to the current data set.
The classification image can be saved to a file by pressing the ‘Save Image To File’ button.
Select the location to save the file, give it a name and press the 'Save' button.
The image is saved in the directory chosen.
Press the 'Close Panel' button to close the panel.
Data Selection Panel

Name of Image Matrix: yimage
Name of Variable Matrix: exactmass

That ends this tutorial. Press the button on the left to go back to the previous step. Press the button on the right to start the tutorial over.