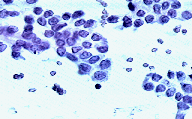
**Machine Learning Exercise**

In this exercise, we will use the **Breast cancer dataset** to build a classification model that differentiate benign from malignant samples.

The dataset source: <http://pages.cs.wisc.edu/~olvi/uwmp/cancer.html>



Files for the exercise

1. wdbc.names – dataset description.
2. wdbc-data.data – dataset consist of 569 cell samples.

Each sample includes the following information:

* The **ID** **number** of the patient
* **30** **pre-calculated features**
* A **label** set by an expert to Malignant (M) or Benign (B).

Steps

1. Import the data from **wdbc-data.data** as a **table** called "data" using the Import Data Tool:

* To open the Import Data Tool - right click on the file name and choose "Import Data…"
* On theIMPORT tab, choose "Table" as the imported data type.

**Q:** What is the size of the table? What do the rows represent? What do the columns represent?

1. Generate a **function** that can automatically import the data, and save it with the name "importData.m".
2. Split the table "data" into two tables by the **holdout** method using the **cvpartition** function:

"training\_set" - Consists of 90% of the samples

"testing\_set" - Consists of 10% of the samples

1. Open the **Classification Learner App** and start a new session with the data you imported, and a validation method **of** **your choice**.

Notice: Should the ID number column be a part of the training set?

1. Run a few classifiers and compare their **Accuracy**.

**Q:** What is the number of examples used to test each classifier?

(Hint: the answer depends on the validation method you chose on step 4)

1. Choose the best classifier, and create a Confusion Matrix to represent the results.

**Q:** What is the **False Negative** **Rate** for the M class (True M class classified by the classifier as B Class)?

1. Create a ROC curve for the M class of the classifier you chose.

**Q:** What is the **False Positive** **Rate** for the M class (True B class classified by the classifier as M Class)?

1. Export the classifier you chose to the workspace as a variable with the name "myClassifier", using the "**Export** **Model**" button.

**Q:** What is the variable type of the classifier?

1. Generate a function that recreates the classifier you trained in the Classification Learner app.
2. Use the classifier you exported to the workspace in order to **predict** the class of each sample in "**testing**\_**set**" (M or B). Compare the predicted classes to the true classes.

(Hint: display the content of the field "HowToPredict" of the classifier struct)

**Q:** How many **misclassifications** did you get using the classifier for "testing\_set"?