

# Exercise 13: Molecular Classification of Cancer (part II)

May 24, 2015

## **Reminder from last week:**

One of the challenges in cancer treatment is correct identification of the type of cancer in order to apply the most suitable treatment. We would like to test the feasibility of using gene expression data obtained by DNA microarrays to classify cancer. Our focus is on two types of leukemia: acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).

Our initial leukemia data set consisted of 38 bone marrow samples (27 ALL, 11 AML) obtained from acute leukemia patients at the time of diagnosis:

- **expression** – a 7129x38 matrix containing the log2-scale data from the DNA microarrays.
- **genes** – a 7129x1 cell array with the gene names
- **true\_labels** – a 38x1 matrix indicating the types of leukemia. 1 means ALL, and 2 means AML.

In the MATLAB file **ex13.mat** you will find the same data from exercise 12, together with new unclassified data:

- **expression\_test** – a 7129x34 matrix containing the log2-scale data from DNA microarrays of 34 new cancer patients.

## **1 Brainstorming**

Split into groups of 3-4 students and design a strategy for classifying the 34 new patients into the two leukemia subgroups: ALL or AML.

## **2 Write pseudocode**

As a group, write (on paper) all the steps you need to go through to classify the samples using your algorithm. Remember that in the end of the script, you should have a vector of size 34x1 with the values 1 or 2 depending on the classification.

## **3 Individual work**

Implement your algorithm using MATLAB.