

mixOmics: sPLS-DA

Cédric HASSEN-KHODJA

20 juin 2016

Introduction

The Small Round Blue Cell Tumors dataset from Khan et al., (2001) contains information of 63 samples and 2308 genes. The samples are distributed in four classes as follows: 8 Burkitt Lymphoma (BL), 23 Ewing Sarcoma (EWS), 12 neuroblastoma (NB), and 20 rhabdomyosarcoma (RMS).

Now, we will see how to analyze srbc by using sPLS-DA. The aim of this analysis is to select the genes that can help predict the class of the samples.

sPLS-DA Rules

I have one single data set (e.g. microarray data) and I am interested in classifying my samples into known classes:

X = expression data

Y = vector indicating the classes of the samples

I would like to know how informative my data are to rightly classify my samples, as well as predicting the class of new samples: PLS-Discriminant Analysis (PLS-DA)

In addition to the above, I would like to select the variables that help classifying the samples: sparse PLS-DA (sPLS-DA)

Results

Preliminary analysis with PCA

Most of the samples are mixed with each other. Let see what happens now if we include the information about the classes of the samples, and if we are selecting the relevant genes that help classifying the sample.

