R for cytometry - Lasso

Nello Blaser

Department of Mathematics, University of Bergen

June 1st 2018

Download data from cytobank

```r
# load CRAN package
require(CytobankAPI)
# Authenticate via authentication token
cyto_session <- authenticate(site="cellmass",
                             auth_token="xxx")
```

```r
# specify experiment id
experiment_id <- 7399
# specify data directory
data_dir <- paste0("data_", experiment_id)
```

```r
dir.create(data_dir, recursive = TRUE)
# list fcs files
files <- fcs_files.list(cyto_session,
                        experiment_id = experiment_id)
```

Download data from cytobank

```
# download fcs files
downloaded_zip <- fcs_files.download_zip(
  UserSession = cyto_session,
  experiment_id = experiment_id,
  fcs_files = files[, "id"],
  directory = data_dir,
  timeout = 60*nrow(files))
unzip(downloaded_zip, exdir = data_dir)
```

Read fcs files

```
require(flowCore)
# read fcs file
filenames <- paste0(data_dir, "/",
                    list.files(data_dir,
                               pattern = "\\.fcs"))
fset <- read.flowSet(filenames)
# decide on lineage markers
#pData(fset[[1]]@parameters)
lineage_markers <- c(
  "CD3(110:114)Dd", "CD45(In115)Dd",
  "CD4(Nd145)Dd", "CD20(Sm147)Dd",
  "CD33(Nd148)Dd", "CD123(Eu151)Dd",
  "CD14(Gd160)Dd", "CD7(Yb176)Dd")
```

FlowSOM

```
require(FlowSOM)
# Set SOM parameters
flowSOM_metaClusters <- 10
flowSOM_xdim <- 7
flowSOM_ydim <- 7
flowSOM_seed <- 20180601 # for reproducible results

# run self-organizing maps
fsom <- FlowSOM(fset,
                colsToUse = lineage_markers,
                xdim=flowSOM_xdim,
                ydim=flowSOM_xdim,
                nClus = flowSOM_metaClusters,
                seed = flowSOM_seed)
```

Summarize clusters

```
# specify functional markers
functional_markers <-
  c("pNFkB(Nd142)Dd", "pp38(Nd144)Dd",
    "pStat5(Nd150)Dd", "pAkt(Sm152)Dd",
    "pStat1(Eu153)Dd", "pSHP2(Sm154)Dd",
    "pZap70(Gd156)Dd", "pStat3(Gd158)Dd",
    "pSlp76(Dy164)Dd", "pBtk(Er166)Dd",
    "pErk(Er168)Dd", "pS6(Yb172)Dd")
# get clusters
clusters <- fsom$FlowSOM$map$mapping[, 1]
meta_clusters <- as.numeric(fsom$metaclustering[clusters])
```

Summarize clusters

```
# combine functional markers and cluster data
dd <- cbind.data.frame(
  meta_clusters = meta_clusters,
  pat = rep(fset@phenoData@data$name,
            times = fsApply(fset, length,
                            use.exprs = TRUE)),
  fsApply(fset, identity, use.exprs = TRUE)[
    , functional_markers])
```

Summarize clusters

```
require(dplyr)
require(tidyr)
# get medians
median_summary <- dd %>%
  group_by(pat, meta_clusters) %>%
  summarize_all(median) %>%
  ungroup %>%
  gather(variable, value, -pat, -meta_clusters) %>%
  unite(temp, variable, meta_clusters) %>%
  spread(temp, value)
```

Lasso regression

```
require(glmnet)
# get medians
X <- as.matrix(select(median_summary, -pat))
# get outcome
y <- grepl("BCR-XL", median_summary$pat)
# lasso regression
cvfit <- cv.glmnet(X, y, family = "binomial")
# get coeficients
cvcoefs <- as.matrix(coef(cvfit, s="lambda.min"))
cv_pos_coefs <- cvcoefs[abs(cvcoefs)>0, ]
cv_pos_coefs
```

```
##          (Intercept) pNFkB(Nd142)Dd_10  pSHP2(Sm154)Dd_2
##           -2.3127587          0.1943530       -25.1324799
##  pSHP2(Sm154)Dd_4 pStat1(Eu153)Dd_6 pStat5(Nd150)Dd_1
##           -0.0545207         -5.0576088         15.1871620
```