MATLAB MDI on simple data

Stephen Coleman

11/02/2020

Data

The data generated is derived from two multivariate distirbutions defined by their means:

$$\mu_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mu_2 = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

Both distributions use $\mathbf{I}$ for the covariance matrix.

MDI

MATLAB MDI is run on two datasets for 50,000 iterations, thinning by a factor of 50. One can see that the labels indicating which distribution the data originates from aligns (almost) perfectly with the clustering captured in the PSM.

```r
# Read in the tibble containing the MDI analysis
my_tib <- readRDS("./compare_tibble.rds")

# The original data
my_data <- my_tib$expression_data[[1]]

# Briefly summarise
summary(my_data)
```

```
##      Gene_1              Gene_2              Gene_3
##  Min.   :-3.888921   Min.   :-2.91436   Min.   :-3.21470
##  1st Qu.:-1.052855   1st Qu.:-1.18651   1st Qu.:-0.97872
##  Median : 0.005483   Median :-0.09020   Median : 0.01928
##  Mean   :-0.007423   Mean   :-0.03847   Mean   : 0.08024
##  3rd Qu.: 0.985904   3rd Qu.: 0.97446   3rd Qu.: 1.08699
##  Max.   : 3.024842   Max.   : 4.81028   Max.   : 2.97134
```

```r
# The PSM
my_psm <- my_tib$similarity_matrix[[1]]

# The labels for which subpopulation each person actually belongs to
labels <- as.factor(c(
  rep(1, 0.5 * nrow(my_data)),
  rep(2, 0.5 * nrow(my_data))
))

# These labels in an annotation dataframe for the pheatmap
```
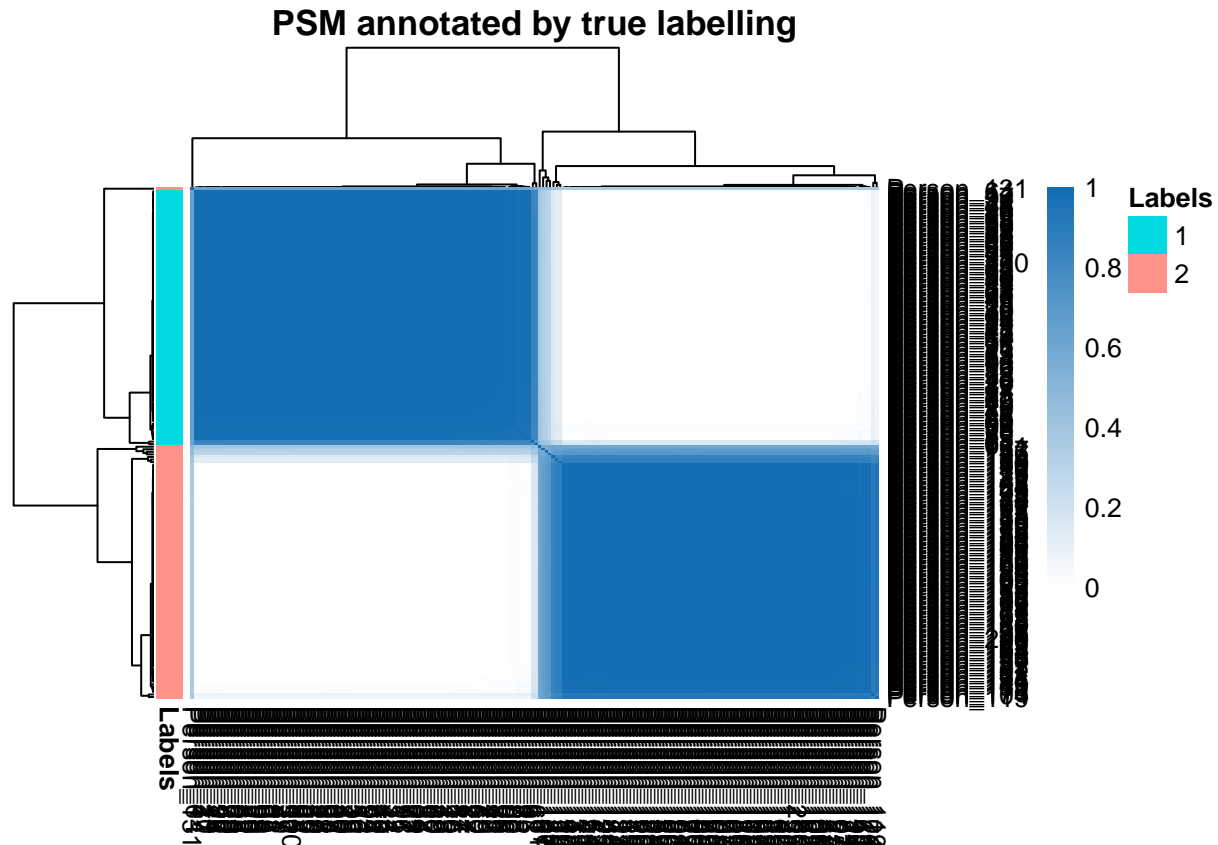
```
annotation_row <- data.frame(Labels = labels) %>%
  set_rownames(row.names(my_data))

# Visualise the PSM annotated by true labelling
pheatmap(my_psm,
  annotation_row = annotation_row,
  color = col_pal,
  breaks = my_breaks,
  main = "PSM annotated by true labelling"
)
```


Comparing this PSM directly with the expression data:

```
# Extract the order from the pheatmap
row_order <- hclust(dist(my_psm))$order

# Find the column order for the expression data
col_order <- hclust(dist(t(my_data)))$order

# Re order the matrices to have a common row order
my_psm <- my_psm[row_order, row_order]
my_data <- my_data[row_order, col_order]

colnames(my_psm) <- NULL
colnames(my_data) <- NULL

# Save the heatmaps of these to the same grid
```

```
compareHeatmapPSMandData(my_psm, my_data,
  save_name = NULL,
  main = "Comparison of PSM and data",
  col_pal_sim = col_pal,
  col_pal_expr = col_pal_expr,
  expr_breaks = expr_breaks,
  sim_breaks = my_breaks
)
```

Comparison of PSM and data