Usage Tips

Junyan Song

2017-12-07

Introduction

This is an example of using HMMdmdv package in R. HMMdmdv implements Hidden Markov Model for large-scale multiple testing of differential mean and variance in the framework of a proposed hierarchical mixture model. This vignette aims to demonstrate the usage of HMMdmdv through some example codes and example data.

Usage of HMMdmdv

(1) initialization

```
install.packages("HMMdmdv_0.1.1.tar.gz", repos = NULL, type="source")
```

(2) loading example data

The example dataset has 1000 rows for 1000 test sites and 100 columns for 100 samples. In each individual test, we are comparing the first 50 samples with the second 50 samples.

```
library(HMMdmdv)
# Loading dataset
data(example_data)
# Check dataset
example_data[1:10, 1:8]
```

```
##              [,1]         [,2]       [,3]        [,4]        [,5]
##  [1,]   2.6242907  -4.99641619  -3.062018  -2.5659846 -6.30053963
##  [2,] -12.3220464 -40.42143479  25.411854 -28.5906662  2.02547908
##  [3,]   8.7694668  -7.04233809  -4.673113 -13.1961235 -7.99669568
##  [4,]   0.7413598   7.65175867   3.205561   4.1768603 -4.14000223
##  [5,]  -4.8978253   0.04991304   7.680206   0.5627407 -1.61020660
##  [6,]  -2.2359698   9.48256606  26.846845  14.5789586 10.78583876
##  [7,]   0.1591787  -4.26908042 -10.676433   6.5384808  9.87601784
##  [8,]  -3.7506621  -2.43609395   8.550669   2.5052252 -0.06575597
##  [9,]   3.5021217  -9.13166883  10.334842   1.8310631  1.55223636
## [10,]   4.7141413   3.55702020   1.968636   7.6128951 -7.51867982
##              [,6]       [,7]       [,8]
##  [1,]   0.26759842  10.155866  2.6142248
##  [2,] -30.06000343 -15.966786  1.4682712
##  [3,]   0.40475621  -4.136894  0.7610810
##  [4,]   9.15686516  10.286435 -1.3841398
##  [5,]  -2.35525596   7.211297  8.0231026
##  [6,]  26.45608382  10.809190 18.0868189
##  [7,]  10.72438530   2.839356  5.5018218
##  [8,]  -0.05796556  -1.122700 -5.7768993
##  [9,]   3.35646679   4.928918 -0.7005532
## [10,]   6.33632074  -5.462055  1.2078019
```

```
dim(example_data)
```

```
## [1] 1000  100
```

```
# 1000 tests, 50 samples in each group
n<-1000
n1 <- 50
n2 <- 50
```

(3) Initial (heuristic) parameter estimate

```
# remove case 3
data_df <- remove_case3(dat = example_data, mean_thresholdPV = 0.1, var_thresholdPV =
0.05, n1 = n1, n2 = n2)
# parameter estimate
init_para_est <- init_est(dat_df = data_df, mean_thresholdPV = 0.05, var_thresholdPV
= 0.1, n1 = n1, n2 = n2, n = n)
```

```
## [1] "initial parameter estimate:"
## [1]  0.6480000  0.1300000  0.1080000  0.1140000  4.7036645 27.6750509
## [7]  2.3092714  0.8935815
```

(4) Independent Gaussian Mixture parameter estimate

```
# upper limit of EM iterations
niter<-1000
# parameter estimate
indep_para_est <- runEM(dat_df = data_df, n1 = n1, n2 = n2, init_para_est= init_para_
est, niter = niter)
```

```
## [1] "independent EM parameter estimate:"
## [1]  0.67557563  0.14673058  0.10800000  0.06969379  3.92439481 24.52639798
## [7]  2.25630233  0.98702806
```

(5) Calculate emission densities

```
emissions <- emission_probs(indep_para_est[5], indep_para_est[6], indep_para_est[7],
indep_para_est[8], example_data, n1 = n1, n2 = n2)
head(emissions, 10)
```

```
##                                  emission_numeric
##  [1,] 7.797759e-133 1.122620e-132    1.800511e-132 1.916198e-132
##  [2,] 1.326501e-174 2.018706e-175    5.770411e-164 6.035901e-165
##  [3,] 1.129686e-150 4.341097e-151    5.774301e-151 1.615323e-151
##  [4,] 1.927956e-117 4.974494e-118    4.334189e-118 1.304342e-118
##  [5,] 9.004879e-132 2.643047e-132    4.421538e-132 9.514479e-133
##  [6,] 6.585679e-171 1.307460e-171    9.775128e-173 1.428814e-172
##  [7,] 1.657701e-142 6.077785e-143    5.566504e-143 2.039573e-143
##  [8,] 3.715715e-146 1.176543e-146    1.861085e-146 4.281675e-147
##  [9,] 9.213114e-131 2.658738e-131    9.824302e-131 2.130905e-131
## [10,] 8.582819e-128 1.869420e-127    3.145988e-128 7.789714e-128
```

(6) HMM parameter estimates

```
HMM_resList <- runHMM(emissions)
```

```
## [1] "HMM initial state probability estimate:"
## [1] 0 0 1 0
## [1] "HMM transition probability estimate:"
##           [,1]       [,2]         [,3]       [,4]
## [1,] 0.7049952 0.12262464 1.058833e-01 0.06649688
## [2,] 0.6586202 0.19083008 2.441677e-74 0.15054972
## [3,] 0.5969715 0.07636621 1.085234e-01 0.21813892
## [4,] 0.6643522 0.16963576 5.424930e-08 0.16601195
## [1] "HMM iterations:"
## [1] 915
```

(7) Posterior inference of hidden states

```
result <- posterior_inference(HMM_resList[[1]], HMM_resList[[2]], train = FALSE)
```

```
## pos_state
##   1   2   3   4
## 783 103  56  58
```

```
head(result, 10)
```

```
##  [1] 3 3 1 1 1 1 1 1 1 1
```

More Information

Details for arguments and functions can be found by typing e.g. `help(package="HMMdmdv")`, `?runEM`.