Package 'robustDA'

September 29, 2014

Type Package

Title Robust Mixture Discriminant Analysis

Version 1.0

Date 2014-09-26

Author Charles Bouveyron & Stephane Girard

Maintainer Charles Bouveyron <charles.bouveyron@parisdescartes.fr>

Depends MASS, mclust, Rsolnp

Description Robust mixture discriminant analysis (RMDA, Bouveyron & Girard, 2009) allows to build a robust supervised classifier from learning data with label noise. The idea of the proposed method is to confront an unsupervised modeling of the data with the supervised information carried by the labels of the learning data in order to detect inconsistencies. The method is able afterward to build a robust classifier taking into account the detected inconsistencies into the labels.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-09-29 07:27:55

R topics documented:

robustDA-package *Robust Mixture Discriminant Analysis*

Description

Robust mixture discriminant analysis (RMDA, Bouveyron & Girard, 2009) allows to build a robust supervised classifier from learning data with label noise. The idea of the proposed method is to confront an unsupervised modeling of the data with the supervised information carried by the labels of the learning data in order to detect inconsistencies. The method is able afterward to build a robust classifier taking into account the detected inconsistencies into the labels.

Details

Package:	robustDA
Type:	Package
Version:	1.0
Date:	2014-09-26
License:	GPL-2

Author(s)

Charles Bouveyron & Stéphane Girard

Maintainer: Charles Bouveyron <charles.bouveyron@parisdescartes.fr>

References

C. Bouveyron and S. Girard, Robust supervised classification with mixture models: Learning from data with uncertain labels, Pattern Recognition, vol. 42 (11), pp. 2649-2658, 2009.

Examples

```
set.seed(12345)

## Simulated data
N = 600
n = N/4
S1 = S2 = S3 = S4 = 2*diag(2)
m1 = 1.5*c(-4,0)
m4 = 1.5*c(0,-4)
m3 = 1.5*c(0,4)
m2 = 1.5*c(4,0)
Z.data = rbind(mvrnorm(n,m1,S1),mvrnorm(n,m2,S2),
  mvrnorm(n,m3,S3),mvrnorm(n,m4,S4))
Z.cls = c(rep(1,n),rep(1,n),rep(2,n),rep(2,n))
```

```
# Split in training and test sets
ind = sample(1:N,N)
X.data = Z.data[ind[1:(3*N/4)],]
X.cls = Z.cls[ind[1:(3*N/4)]]
Y.data = Z.data[ind[(3*N/4+1):N],]
Y.cls = Z.cls[ind[(3*N/4+1):N]]

## Adding noise label
cls = X.cls
nois = rbinom(length(cls),1,0.3)
lbl = cls
lbl[cls==1 & nois] = 2
lbl[cls==2 & nois] = 1

# Plot
par(mfrow=c(2,2))
plot(X.data,col=X.cls,pch=(18:19)[X.cls],
  main='Learning set with actual labels',xlab='',ylab='')
plot(X.data,col=lbl,pch=(18:19)[lbl],
  main='Learning set with noisy labels',xlab='',ylab='')


## Classification with LDA
c.lda = lda(X.data,lbl)
res.lda <- predict(c.lda,Y.data)$class

## Classification with MDA
c.mda = MclustDA(X.data,lbl,G=2)
res.mda = predict(c.mda,Y.data)$cl
plot(Y.data,col=res.mda,pch=(18:19)[res.mda],
     main='Classification of test set with MDA',xlab='',ylab='')

## Classification with RMDA
c.rmda <- rmda(X.data,lbl,K=4,model='VEV')
res.rmda <- predict(c.rmda,Y.data)
plot(Y.data,col=res.rmda$cls,pch=(18:19)[res.rmda$cls],
     main='Classification of test set with RMDA',xlab='',ylab='')

## Classification results
cat("* Correct classification rates on test data:\n")
cat("\tLDA:\t",sum(res.lda == Y.cls) / length(Y.cls),"\n")
cat("\tMDA:\t",sum(res.mda == Y.cls) / length(Y.cls),"\n")
cat("\tRMDA:\t",sum(res.rmda$cls == Y.cls) / length(Y.cls),"\n")
```

`predict.rmda` *Prediction method for 'rmda' class objects.*

Description

The prediction method for 'rmda' class objects which allows to predict the labels for test observations.

Usage

```
## S3 method for class 'rmda'
predict(object, X, ...)
```

Arguments

object	a supervised classifier genarated by the rmda function (a 'rmda' object).
X	the test data.
...	additional options for internal functions.

Value

A list with:

- cls: the predicted class labels,

- P: the posterior probabilities that observations belong to the classes.

Author(s)

Charles Bouveyron & Stéphane Girard

References

C. Bouveyron and S. Girard, Robust supervised classification with mixture models: Learning from data with uncertain labels, Pattern Recognition, vol. 42 (11), pp. 2649-2658, 2009.

Examples

```
set.seed(12345)

## Simulated data
N = 600
n = N/4
S1 = S2 = S3 = S4 = 2*diag(2)
m1 = 1.5*c(-4,0)
m4 = 1.5*c(0,-4)
m3 = 1.5*c(0,4)
m2 = 1.5*c(4,0)
Z.data = rbind(mvrnorm(n,m1,S1),mvrnorm(n,m2,S2),
  mvrnorm(n,m3,S3),mvrnorm(n,m4,S4))
Z.cls = c(rep(1,n),rep(1,n),rep(2,n),rep(2,n))

# Split in training and test sets
ind = sample(1:N,N)
X.data = Z.data[ind[1:(3*N/4)],]
```

```
X.cls = Z.cls[ind[1:(3*N/4)]]
Y.data = Z.data[ind[(3*N/4+1):N],]
Y.cls = Z.cls[ind[(3*N/4+1):N]]

## Adding noise label
cls = X.cls
nois = rbinom(length(cls),1,0.3)
lbl = cls
lbl[cls==1 & nois] = 2
lbl[cls==2 & nois] = 1

# Plot
par(mfrow=c(2,2))
plot(X.data,col=X.cls,pch=(18:19)[X.cls],
  main='Learning set with actual labels',xlab='',ylab='')
plot(X.data,col=lbl,pch=(18:19)[lbl],
  main='Learning set with noisy labels',xlab='',ylab='')


## Classification with LDA
c.lda = lda(X.data,lbl)
res.lda <- predict(c.lda,Y.data)$class

## Classification with MDA
c.mda = MclustDA(X.data,lbl,G=2)
res.mda = predict(c.mda,Y.data)$cl
plot(Y.data,col=res.mda,pch=(18:19)[res.mda],
     main='Classification of test set with MDA',xlab='',ylab='')

## Classification with RMDA
c.rmda <- rmda(X.data,lbl,K=4,model='VEV')
res.rmda <- predict(c.rmda,Y.data)
plot(Y.data,col=res.rmda$cls,pch=(18:19)[res.rmda$cls],
     main='Classification of test set with RMDA',xlab='',ylab='')

## Classification results
cat("* Correct classification rates on test data:\n")
cat("\tLDA:\t",sum(res.lda == Y.cls) / length(Y.cls),"\n")
cat("\tMDA:\t",sum(res.mda == Y.cls) / length(Y.cls),"\n")
cat("\tRMDA:\t",sum(res.rmda$cls == Y.cls) / length(Y.cls),"\n")
```

---

rmda *Robust Mixture Discriminant Analysis*

Description

Robust mixture discriminant analysis (RMDA, Bouveyron & Girard, 2009) allows to build a robust supervised classifier from learning data with label noise. The idea of the proposed method is to confront an unsupervised modeling of the data with the supervised information carried by the labels of the learning data in order to detect inconsistencies. The method is able afterward to build a robust classifier taking into account the detected inconsistencies into the labels.

Usage

```
rmda(X, cls, K = 4, model = "VEV")
```

Arguments

X	the data.
cls	the known labels.
K	the total number of groups which are assumed to exist within the known classes.
model	the model name (according to the Mclust nomenclature).

Value

An object of class "rmda" with: - K: the total number of groups, - prms: all model parameters as returned by Mclust, - R: the posterior probabilities that the fitted groups belong to the known classes.

Author(s)

Charles Bouveyron & Stéphane Girard

References

C. Bouveyron and S. Girard, Robust supervised classification with mixture models: Learning from data with uncertain labels, Pattern Recognition, vol. 42 (11), pp. 2649-2658, 2009.

Examples

```
set.seed(12345)

## Simulated data
N = 600
n = N/4
S1 = S2 = S3 = S4 = 2*diag(2)
m1 = 1.5*c(-4,0)
m4 = 1.5*c(0,-4)
m3 = 1.5*c(0,4)
m2 = 1.5*c(4,0)
Z.data = rbind(mvrnorm(n,m1,S1),mvrnorm(n,m2,S2),
  mvrnorm(n,m3,S3),mvrnorm(n,m4,S4))
Z.cls = c(rep(1,n),rep(1,n),rep(2,n),rep(2,n))

# Split in training and test sets
ind = sample(1:N,N)
X.data = Z.data[ind[1:(3*N/4)],]
X.cls = Z.cls[ind[1:(3*N/4)]]
Y.data = Z.data[ind[(3*N/4+1):N],]
Y.cls = Z.cls[ind[(3*N/4+1):N]]

## Adding noise label
cls = X.cls
```

```
nois = rbinom(length(cls),1,0.3)
lbl = cls
lbl[cls==1 & nois] = 2
lbl[cls==2 & nois] = 1

# Plot
par(mfrow=c(2,2))
plot(X.data,col=X.cls,pch=(18:19)[X.cls],
  main='Learning set with actual labels',xlab='',ylab='')
plot(X.data,col=lbl,pch=(18:19)[lbl],
  main='Learning set with noisy labels',xlab='',ylab='')


## Classification with LDA
c.lda = lda(X.data,lbl)
res.lda <- predict(c.lda,Y.data)$class

## Classification with MDA
c.mda = MclustDA(X.data,lbl,G=2)
res.mda = predict(c.mda,Y.data)$cl
plot(Y.data,col=res.mda,pch=(18:19)[res.mda],
     main='Classification of test set with MDA',xlab='',ylab='')

## Classification with RMDA
c.rmda <- rmda(X.data,lbl,K=4,model='VEV')
res.rmda <- predict(c.rmda,Y.data)
plot(Y.data,col=res.rmda$cls,pch=(18:19)[res.rmda$cls],
     main='Classification of test set with RMDA',xlab='',ylab='')

## Classification results
cat("* Correct classification rates on test data:\n")
cat("\tLDA:\t",sum(res.lda == Y.cls) / length(Y.cls),"\n")
cat("\tMDA:\t",sum(res.mda == Y.cls) / length(Y.cls),"\n")
cat("\tRMDA:\t",sum(res.rmda$cls == Y.cls) / length(Y.cls),"\n")
```

Index