

Package ‘ber’

February 19, 2015

Type Package

Title Batch Effects Removal

Version 4.0

Date 2013-12-03

Author Marco Giordan

Maintainer Marco Giordan <marco.giordan@fmach.it>

Depends MASS

Description The functions in this package remove batch effects from microarray normalized data. The expression levels of the genes are represented in a matrix where rows correspond to independent samples and columns to genes (variables). The batches are represented by categorical variables (objects of class factor). When further covariates of interest are available they can be used to remove efficiently the batch effects and adjust the data.

License GPL-2

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2013-03-12 10:01:16

R topics documented:

| | |
|---------------------------|---|
| ber | 2 |
| ber_bg | 3 |
| combat_np | 4 |
| combat_p | 5 |
| mean_centering | 6 |
| standardization | 7 |

| | |
|--------------|----------|
| Index | 8 |
|--------------|----------|

ber *Batch Effects Removal*

Description

Batch effects are removed using a two-stage regression approach.

Usage

```
ber(Y, b, covariates = NULL)
```

Arguments

| | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |
| covariates | An object of class <code>data.frame</code> where each column corresponds to a quantitative variable (of class <code>numeric</code>) or a qualitative variable (of class <code>factor</code>). |

Details

In this implementation NA values are not allowed.

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

References

M. Giordan. February 2013. A Two-Stage Procedure for the Removal of Batch Effects in Microarray Studies. *Statistics in Biosciences*.

See Also

[ber_bg](#), [combat_np](#), [combat_p](#), [mean_centering](#), [standardization](#)

Examples

```
Y<-matrix(rnorm(6000),nrow=12)
class<-gl(2,6,labels=c("Control","Treat"))
class<-data.frame(class)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-ber(Y,batch,class)
```

ber_bg *Batch Effects Removal using Bagging*

Description

Batch effects are removed using a two-stage regression approach and bagging.

Usage

```
ber_bg(Y, b, covariates = NULL, partial=TRUE, nSim=150)
```

Arguments

| | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |
| covariates | An object of class data.frame where each column corresponds to a quantitative variable (of class numeric) or a qualitative variable (of class factor). |
| partial | A logical value indicating if partial bagging or full bagging have to be performed. See reference below. |
| nSim | Number of bootstrap samples. |

Details

In this implementation NA values are not allowed.

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

References

M. Giordan. February 2013. A Two-Stage Procedure for the Removal of Batch Effects in Microarray Studies. *Statistics in Biosciences*.

See Also

[ber](#), [combat_np](#), [combat_p](#), [mean_centering](#), [standardization](#)

Examples

```

Y<-matrix(rnorm(6000),nrow=12)
class<-gl(2,6,labels=c("Control","Treat"))
class<-data.frame(class)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-ber_bg(Y,batch,class)

```

| | |
|-----------|------------------------------|
| combat_np | <i>non-parametric combat</i> |
|-----------|------------------------------|

Description

Batch effects are removed using a non-parametric empirical Bayes approach, as described in Johnson et al. 2007.

Usage

```
combat_np(Y, b, covariates = NULL)
```

Arguments

| | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |
| covariates | An object of class data.frame where each column corresponds to a quantitative variable (of class numeric) or a qualitative variable (of class factor). |

Details

In this implementation pre-processing of the data is obtained through Moore-Penrose pseudo inverse. The final matrix of adjusted data can be different from that obtained through the original ComBat algorithm (see sva package) due to small differences in the optimization procedure. NA values are not allowed.

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

References

Johnson et al. 2007 <http://www.ncbi.nlm.nih.gov/pubmed/16632515>

See Also

[ber](#), [ber_bg](#), [combat_p](#), [mean_centering](#), [standardization](#)

Examples

```
Y<-matrix(rnorm(1200),nrow=12)
class<-gl(2,6,labels=c("Control","Treat"))
class<-data.frame(class)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-combat_np(Y,batch,class)
```

| | |
|----------|--------------------------|
| combat_p | <i>parametric combat</i> |
|----------|--------------------------|

Description

Batch effects are removed using a parametric empirical Bayes approach, as described in Johnson et al. 2007.

Usage

```
combat_p(Y, b, covariates = NULL, prior.plots=T)
```

Arguments

| | |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |
| covariates | An object of class data.frame where each column corresponds to a quantitative variable (of class numeric) or a qualitative variable (of class factor). |
| prior.plots | A logical value; if true prior plots to compare kernel density estimate and parametric estimate are produced. |

Details

In this implementation pre-processing of the data is obtained through Moore-Penrose pseudo inverse. The final matrix of adjusted data can be different from that obtained through the original ComBat algorithm (see sva package) due to small differences in the optimization procedure. NA values are not allowed.

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

References

Johnson et al. 2007 <http://www.ncbi.nlm.nih.gov/pubmed/16632515>

See Also

[ber](#), [ber_bg](#), [combat_np](#), [mean_centering](#), [standardization](#)

Examples

```
Y<-matrix(rnorm(1200),nrow=12)
class<-gl(2,6,labels=c("Control","Treat"))
class<-data.frame(class)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-combat_p(Y,batch,class)
```

mean_centering

mean_centering

Description

Batch effects are removed using the means of the batches.

Usage

```
mean_centering(Y, b)
```

Arguments

- | | |
|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

See Also

[ber](#), [ber_bg](#), [combat_np](#), [combat_p](#), [standardization](#)

Examples

```
Y<-matrix(rnorm(6000),nrow=12)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-mean_centering(Y,batch)
```

| | |
|-----------------|------------------------|
| standardization | <i>standardization</i> |
|-----------------|------------------------|

Description

Batch effects are removed using the means and the standard deviations of the batches.

Usage

```
standardization(Y, b)
```

Arguments

| | |
|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Y | A matrix with n rows and g columns, where n is the number of objects and g is the number of variables. In the case of gene expression data, columns correspond to genes (probe sets) and rows to samples. |
| b | A vector of class factor with the element in position i ($i = 1, \dots, n$) representing the batch from which observation i belongs to. |

Value

A matrix of adjusted data with n rows and g columns.

Author(s)

Marco Giordan

See Also

[ber](#), [ber_bg](#), [combat_np](#), [combat_p](#), [mean_centering](#)

Examples

```
Y<-matrix(rnorm(6000),nrow=12)
batch<-rep(gl(2,3,labels=c("Batch1","Batch2")),2)
YEadj<-standardization(Y,batch)
```

Index

ber, 2, 3, 5–7

ber_bg, 2, 3, 5–7

combat_np, 2, 3, 4, 6, 7

combat_p, 2, 3, 5, 5, 7

mean_centering, 2, 3, 5, 6, 6, 7

standardization, 2, 3, 5–7, 7