Package 'FLLat'

October 12, 2022

Version 1.2-1

Date 2017-05-23

Title Fused Lasso Latent Feature Model

Imports graphics, stats, gplots

Description Fits the Fused Lasso Latent Feature model, which is used for modeling multi-sample aCGH data to identify regions of copy number variation (CNV). Produces a set of features that describe the patterns of CNV and a set of weights that describe the composition of each sample. Also provides functions for choosing the optimal tuning parameters and the appropriate number of features, and for estimating the false discovery rate.

License GPL (>= 2)

LazyData yes

NeedsCompilation yes

Author Gen Nowak [aut, cre],
Trevor Hastie [aut],
Jonathan R. Pollack [aut],
Robert Tibshirani [aut],
Nicholas Johnson [aut]

Maintainer Gen Nowak <gen.nowak@gmail.com>

Repository CRAN

Date/Publication 2017-05-23 03:42:21 UTC

R topics documented:

FLLat   *Fused Lasso Latent Feature Model*

Description

Fits the Fused Lasso Latent Feature (FLLat) model for given values of J (the number of features), and λ_1 and λ_2 (the two fused lasso tuning parameters).

Usage

```
FLLat(Y, J=min(15,floor(ncol(Y)/2)), B="pc", lam1, lam2, thresh=10^(-4),
      maxiter=100, maxiter.B=1, maxiter.T=1)
```

Arguments

Y	A matrix of data from an aCGH experiment (usually in the form of log intensity ratios) or some other type of copy number data. Rows correspond to the probes and columns correspond to the samples.
J	The number of features in the FLLat model. The default is the smaller of either 15 or the number of samples divided by 2.
B	The initial values for the features. Can be one of "pc" (the first J principal components of Y), "rand" (a random selection of J columns of Y), or a user specified matrix of initial values, where rows correspond to the probes and columns correspond to the features. The default is "pc".
lam1	The tuning parameter λ_1 in the fused lasso penalty that controls the level of sparsity in the features.
lam2	The tuning parameter λ_2 in the fused lasso penalty that controls the level of smoothness in the features.
thresh	The threshold for determining when the solutions have converged. The default is 10^{-4}.
maxiter	The maximum number of iterations for the outer loop of the algorithm. The default is 100.
maxiter.B	The maximum number of iterations for the inner loop of the algorithm for estimating the features B. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.
maxiter.T	The maximum number of iterations for the inner loop of the algorithm for estimating the weights Θ. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.

Details

This function fits the Fused Lasso Latent Feature model to multi-sample aCGH data, as described in Nowak and others (2011), for given values of J, λ_1 and λ_2. Given aCGH data consisting of S samples and L probes, the model is given by:

$$Y = B\Theta,$$

where Y is an L-by-S matrix denoting the aCGH data (with samples in columns), B is an L-by-J matrix denoting the features (with features in columns), and Θ is a J-by-S matrix denoting the weights. Each feature describes a pattern of copy number variation and the weights describe the composition of each sample. Specifically, each sample (column of Y) is modeled as a weighted sum of the features (columns of B), with the weights given by the corresponding column of Θ.

The model is fitted by minimizing a penalized version of the residual sum of squares (RSS):

$$RSS + \sum_{j=1}^{J} PEN_j$$

where the penalty is given by:

$$PEN_j = \lambda_1 \sum_{l=1}^{L} |\beta_{lj}| + \lambda_2 \sum_{l=2}^{L} |\beta_{lj} - \beta_{l-1,j}|.$$

Here β_{lj} denotes the (l,j)th element of B. We also constrain the L_2 norm of each row of Θ to be less than or equal to 1.

For more details, please see Nowak and others (2011) and the package vignette.

Value

An object of class FLLat with components:

Beta	The estimated features $\hat{B}$.
Theta	The estimated weights $\hat{\Theta}$.
niter	The number of iterations taken by the algorithm (outer loop).
rss	The residual sum of squares of the fitted model.
bic	The BIC for the fitted model. See FLLat.BIC for more details.
lam1	The value of λ_1 used in the model.
lam2	The value of λ_2 used in the model.

There is a plot method and a predict method for FLLat objects.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

plot.FLLat, predict.FLLat, FLLat.BIC, FLLat.PVE, FLLat.FDR

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Run FLLat for J = 5, lam1 = 1 and lam2 = 9.
result <- FLLat(simaCGH,J=5,lam1=1,lam2=9)

## Plot the estimated features.
plot(result)

## Plot a heatmap of the estimated weights.
plot(result,type="weights")
```

FLLat.BIC   *Optimal Tuning Parameters for the Fused Lasso Latent Feature Model*

Description

Returns the optimal values of the fused lasso tuning parameters for the Fused Lasso Latent Feature
(FLLat) model by minimizing the BIC. Also returns the fitted FLLat model for the optimal values
of the tuning parameters.

Usage

```
FLLat.BIC(Y, J=min(15,floor(ncol(Y)/2)), B="pc", thresh=10^(-4), maxiter=100,
          maxiter.B=1, maxiter.T=1)
```

Arguments

Y	A matrix of data from an aCGH experiment (usually in the form of log intensity ratios) or some other type of copy number data. Rows correspond to the probes and columns correspond to the samples.
J	The number of features in the FLLat model. The default is the smaller of either 15 or the number of samples divided by 2.
B	The initial values for the features. Can be one of "pc" (the first J principal components of Y), "rand" (a random selection of J columns of Y), or a user specified matrix of initial values, where rows correspond to the probes and columns correspond to the features. The default is "pc".
thresh	The threshold for determining when the solutions have converged. The default is 10^{-4}.
maxiter	The maximum number of iterations for the outer loop of the FLLat algorithm. The default is 100.
maxiter.B	The maximum number of iterations for the inner loop of the FLLat algorithm for estimating the features B. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.
maxiter.T	The maximum number of iterations for the inner loop of the FLLat algorithm for estimating the weights Θ. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.

Details

This function returns the optimal values of the fused lasso tuning parameters, λ_1 and λ_2, for the FLLat model. The optimal values are chosen by first re-parameterizing λ_1 and λ_2 in terms of λ_0 and a proportion α such that $\lambda_1 = \alpha\lambda_0$ and $\lambda_2 = (1 - \alpha)\lambda_0$. The values of α are fixed to be $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ and for each value of α we consider a range of λ_0 values. The optimal values of λ_0 and α (and consequently λ_1 and λ_2) are chosen by minimizing the following BIC-type criterion over this two dimensional grid:

$$(SL) \times \log\left(\frac{RSS}{SL}\right) + k_{\alpha,\lambda_0} \log(SL),$$

where S is the number of samples, L is the number probes, RSS denotes the residual sum of squares and k_{α,λ_0} denotes the sum over all the features of the number of unique non-zero elements in each estimated feature.

Note that for extremely large data sets, this function may take some time to run.

For more details, please see Nowak and others (2011) and the package vignette.

Value

A list with components:

lam0	The optimal value of λ_0.
alpha	The optimal value of α.
lam1	The optimal value of λ_1.
lam2	The optimal value of λ_2.
opt.FLLat	The fitted FLLat model for the optimal values of the tuning parameters.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

FLLat

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Run FLLat.BIC to choose optimal tuning parameters for J = 5 features.
result.bic <- FLLat.BIC(simaCGH,J=5)
```

```
## Plot the features for the optimal FLLat model.
plot(result.bic$opt.FLLat)

## Plot a heatmap of the weights for the optimal FLLat model.
plot(result.bic$opt.FLLat,type="weights")
```

FLLat.FDR   *False Discovery Rate for the Fused Lasso Latent Feature Model*

Description

Estimates the false discovery rate (FDR) over a range of threshold values for a fitted Fused Lasso Latent Feature (FLLat) model. Also plots the FDRs against the threshold values.

Usage

```
FLLat.FDR(Y, Y.FLLat, n.thresh=50, fdr.control=0.05, pi0=1, n.perms=20)

## S3 method for class 'FDR'
plot(x, xlab="Threshold", ylab="FDR", ...)
```

Arguments

Y	A matrix of data from an aCGH experiment (usually in the form of log intensity ratios) or some other type of copy number data. Rows correspond to the probes and columns correspond to the samples.
Y.FLLat	A FLLat model fitted to Y. That is, an object of class FLLat, as returned by [FLLat](#).
n.thresh	The number of threshold values at which to estimate the FDR. The default is 50.
fdr.control	A value at which to control the FDR. The function will return the smallest threshold value which controls the FDR at the specified value. The default is 0.05.
pi0	The proportion of true null hypotheses. For probe location l in sample s, the null hypothesis $H_0(l,s)$ states that there is no copy number variation at that location. The default is 1.
n.perms	The number of permutations of the aCGH data used in estimating the FDRs. The default is 20.
x	An object of class FDR, as returned by [FLLat.FDR](#).
xlab	The title for the x-axis of the FDR plot.
ylab	The title for the y-axis of the FDR plot.
...	Further graphical parameters.

Details

Identifying regions of copy number variation (CNV) in aCGH data can be viewed in a multiple-testing framework. For each probe location l within sample s, we are essentially testing the hypothesis $H_0(l, s)$ that there is no CNV at that location. The decision to reject each hypothesis can be based on the fitted values $\hat{Y} = \hat{B}\hat{\Theta}$ produced by the FLLat model. Specifically, for a given threshold value T, we can declare location (l, s) as exhibiting CNV if $|\hat{y}_{ls}| \geq T$. The FDR is then defined to be the expected proportion of declared CNVs which are not true CNVs.

The FDR for a fitted FLLat model is estimated in the following manner. Firstly, `n.thresh` threshold values are chosen, equally spaced between 0 and the largest absolute fitted value over all locations (l, s). Then, for each threshold value, the estimated FDR is equal to

$$FDR = \frac{\pi_0 \times V_0}{R}$$

where:

- The quantity R is the number of declared CNVs calculated from the fitted FLLat model, as described above.

- The quantity V_0 is the number of declared CNVs calculated from re-fitting the FLLat model to permuted versions of the data Y. In each permuted data set, the probe locations within each sample are permuted to approximate the null distribution of the data.

- The quantity π_0 is the proportion of true null hypotheses. The default value of 1 will result in conservative estimates of the FDR. If warranted, smaller values of π_0 can be specified.

For more details, please see Nowak and others (2011) and the package vignette.

Value

An object of class FDR with components:

thresh.vals   The threshold values for which each FDR was estimated.

FDRs   The estimated FDR for each value of thresh.vals.

thresh.control The smallest threshold value which controls the estimated FDR at fdr.control.

There is a plot method for FDR objects.

Note

Due to the randomness of the permutations, for reproducibility of results please set the random seed using set.seed before running FLLat.FDR.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

   FLLat

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Run FLLat for J = 5, lam1 = 1 and lam2 = 9.
result <- FLLat(simaCGH,J=5,lam1=1,lam2=9)

## Estimate the FDRs.
result.fdr <- FLLat.FDR(simaCGH,result)

## Plotting the FDRs against the threshold values.
plot(result.fdr)

## The threshold value which controls the FDR at 0.05.
result.fdr$thresh.control
```

FLLat.PVE   *Choosing the Number of Features for the Fused Lasso Latent Feature Model*

Description

   Calculates the percentage of variation explained (PVE) for a range of values of J (the number of features) for the Fused Lasso Latent Feature (FLLat) model. Also plots the PVE against J, which can be used for choosing the value of J.

Usage

```
FLLat.PVE(Y, J.seq=seq(1,min(15,floor(ncol(Y)/2)),by=2), B=c("pc","rand"),
          lams=c("same","diff"), thresh=10^(-4), maxiter=100, maxiter.B=1,
          maxiter.T=1)

## S3 method for class 'PVE'
plot(x, xlab="Number of Features", ylab="PVE", ...)
```

Arguments

Y	A matrix of data from an aCGH experiment (usually in the form of log intensity ratios) or some other type of copy number data. Rows correspond to the probes and columns correspond to the samples.
J.seq	A vector of values of J (the number of features) for which to calculate the PVE. The default values are every second integer between 1 and smaller of either 15 or the number of samples divided by 2.

B	The initial values for the features to use in the FLLat algorithm for each value of J. Can be one of "pc" (the first J principal components of Y) or "rand" (a random selection of J columns of Y). The default is "pc".
lams	The choice of whether to use the same values of the tuning parameters in the FLLat algorithm for each value of J ("same") or to calculate the optimal tuning parameters for each value of J ("diff"). When using the same values, the optimal tuning parameters are calculated once for the default value of J in the FLLat algorithm. The default is "same".
thresh	The threshold for determining when the solutions have converged in the FLLat algorithm. The default is 10^{-4}.
maxiter	The maximum number of iterations for the outer loop of the FLLat algorithm. The default is 100.
maxiter.B	The maximum number of iterations for the inner loop of the FLLat algorithm for estimating the features B. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.
maxiter.T	The maximum number of iterations for the inner loop of the FLLat algorithm for estimating the weights Θ. The default is 1. Increasing this may decrease the number of iterations for the outer loop but may still increase total run time.
x	An object of class PVE, as returned by FLLat.PVE.
xlab	The title for the x-axis of the PVE plot.
ylab	The title for the y-axis of the PVE plot.
...	Further graphical parameters.

Details

This function calculates the PVE for each value of J as specified by J.seq. The PVE is defined to be:

$$PVE = 1 - \frac{RSS}{TSS}$$

where RSS and TSS denote the residual sum of squares and the total sum of squares, respectively. For each value of J, the PVE is calculated by fitting the FLLat model with that value of J.

There are two choices for how the tuning parameters are chosen when fitting the FLLat model for each value of J. The first choice, given by lams="same", applies the FLLat.BIC function just once for the default value of J. The resulting optimal tuning parameters are then used for all values of J in J.seq. The second choice, given by lams="diff", applies the FLLat.BIC function for each value of J in J.seq. Although this second choice will give a more accurate measure of the PVE, it will take much longer to run than the first choice.

When the PVE is plotted against J, as J increases the PVE will begin to plateau after a certain point, indicating that additional features are not improving the model. Therefore, the value of J to use in the FLLat algorithm can be chosen as the point at which the PVE plot begins to plateau.

For more details, please see Nowak and others (2011) and the package vignette.

Value

An object of class PVE with components:

PVEs   The PVE for each value of *J* in J.seq.

J.seq   The sequence of *J* values used.

There is a [plot](#) method for PVE objects.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

[FLLat](#), [FLLat.BIC](#)

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Generate PVEs for J ranging from 1 to the number of samples divided by 2.
result.pve <- FLLat.PVE(simaCGH,J.seq=1:(ncol(simaCGH)/2))

## Generate PVE plot.
plot(result.pve)
```

---

plot.FLLat   *Plots Results from the Fused Lasso Latent Feature Model*

Description

Plots either the estimated features or a heatmap of the estimated weights from a fitted Fused Lasso Latent Feature (FLLat) model (i.e., an object of class FLLat).

Usage

```
## S3 method for class 'FLLat'
plot(x, type=c("features","weights"), f.mar=c(5,3,4,2), f.xlab="Probe",
     w.mar=c(3,5,0,2), samp.names=1:ncol(x$Theta), hc.meth="complete", ...)
```

Arguments

x	A fitted FLLat model. That is, an object of class FLLat, as returned by FLLat.
type	The choice of whether to plot the estimated features $\hat{B}$ or a heatmap of the estimated weights $\hat{\Theta}$. Default is "features".
f.mar	The margins for the plot of each estimated feature.
f.xlab	The label for the x-axis for the plot of each estimated feature.
w.mar	The margins for the heatmap of the estimated weights.
samp.names	The sample names used to label the columns in the heatmap of the estimated weights.
hc.meth	The agglomeration method to be used in the hierarchical clustering of the columns of $\hat{\Theta}$. See hclust.
...	Further graphical parameters, for the plot function when type="features" and for the image function when type="weights".

Details

This function plots the estimated features $\hat{B}$ or a heatmap of the estimated weights $\hat{\Theta}$ from a fitted FLLat model. The features are plotted in order of decreasing total magnitude, where the magnitude is given by $\sum_{l=1}^{L} \hat{\beta}_{lj}^2$ with $\hat{\beta}_{lj}$ for $l = 1, \ldots, L$ denoting the jth estimated feature (column of $\hat{B}$). Similarly, the rows of the heatmap of the estimated weights are re-ordered in the same way. The heatmap also includes a dendrogram of a hierarchical clustering of the samples based on their estimated weights (columns of $\hat{\Theta}$).

For more details, please see Nowak and others (2011) and the package vignette.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

FLLat

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Run FLLat for J = 5, lam1 = 1 and lam2 = 9.
result <- FLLat(simaCGH,J=5,lam1=1,lam2=9)

## Plot the estimated features.
```

```
plot(result)

## Plot a heatmap of the estimated weights.
plot(result,type="weights")
```

---

predict.FLLat	*Predicted Values and Weights based on the Fused Lasso Latent Feature Model*

Description

Calculates predicted values and weights for a new set of samples using the estimated features from a fitted Fused Lasso Latent Feature (FLLat) model.

Usage

```
## S3 method for class 'FLLat'
predict(object, newY=NULL, thresh=10^(-4), maxiter.T=100, ...)
```

Arguments

object	A fitted FLLat model. That is, an object of class FLLat, as returned by [FLLat]().
newY	A matrix of new data from an aCGH experiment (usually in the form of log intensity ratios) or some other type of copy number data. Rows correspond to the probes and columns correspond to the samples. The number of probes must match the number of probes in the data used to produce the fitted FLLat model. Note that if newY is not specified, the fitted values from the fitted FLLat model are returned.
thresh	The threshold for determining when the predicted weights have converged. The default is 10^{-4}.
maxiter.T	The maximum number of iterations for the algorithm for calculating the predicted weights. The default is 100.
...	Arguments passed to or from other methods.

Details

Based on the estimated features $\hat{B}$ from a fitted FLLat model, this function predicts the new weights that need to be applied to each feature for predicting a new set of samples Y^*. The predicted weights $\hat{\Theta}^*$ are calculated by minimizing the residual sum of squares:

$$RSS = \left\| Y^* - \hat{B}\Theta^* \right\|_F^2$$

where the L_2 norm of each row of $\hat{\Theta}^*$ is still constrained to be less than or equal to 1. From these predicted weights, the predicted values for the new set of samples are calculated as $\hat{Y}^* = \hat{B}\hat{\Theta}^*$. These predicted values can useful when performing model validation.

Note that for the predictions to be meaningful and useful, the new set of samples Y^* must be similar in scale/magnitude to the original data used in producing the fitted FLLat model. If a new set of samples Y^* are not specified, the function returns the fitted values $\hat{Y}$ and estimated weights $\hat{\Theta}$ from the fitted FLLat model.

For more details, please see Nowak and others (2011) and the package vignette.

Value

A list with components:

pred.Y   The predicted values $\hat{Y}^*$ for the new set of samples, or the fitted values $\hat{Y}$ from the fitted FLLat model.

Theta   The predicted weights $\hat{\Theta}^*$ for the new set of samples, or the estimated weights $\hat{\Theta}$ from the fitted FLLat model.

niter   The number of iterations taken by the algorithm for calculating the predicted weights, or the number of iterations taken by the algorithm for producing the fitted FLLat model.

rss   The residual sum of squares based on the new set of samples, or based on the original data used in producing the fitted FLLat model.

Author(s)

Gen Nowak <gen.nowak@gmail.com>, Trevor Hastie, Jonathan R. Pollack, Robert Tibshirani and Nicholas Johnson.

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

See Also

FLLat

Examples

```
## Load simulated aCGH data.
data(simaCGH)

## Divide the data into a training and test set.
tr.dat <- simaCGH[,1:15]
tst.dat <- simaCGH[,16:20]

## Run FLLat for J = 5, lam1 = 1 and lam2 = 9 on the training set.
result.tr <- FLLat(tr.dat,J=5,lam1=1,lam2=9)

## Calculate fitted values on the training set.
tr.pred <- predict(result.tr)

## Calculate predicted values and weights on the test set using the FLLat
```

```
## model (i.e., the features) fitted on the training set.
tst.pred <- predict(result.tr,newY=tst.dat)

## Plotting predicted values and data for the first sample in the test set.
plot(tst.dat[,1],xlab="Probe",ylab="Y")
lines(tst.pred$pred.Y[,1],col="red",lwd=3)
```

---

simaCGH   *Simulated aCGH Data*

Description

These 20 samples of aCGH data are simulated using the Fused Lasso Latent Feature model with 5
features, as described in Section 4 of Nowak and others (2011).

Usage

```
data(simaCGH)
```

Format

A matrix consisting of 1000 probes (rows) and 20 samples (columns).

References

G. Nowak, T. Hastie, J. R. Pollack and R. Tibshirani. A Fused Lasso Latent Feature Model for
Analyzing Multi-Sample aCGH Data. *Biostatistics*, 2011, doi: 10.1093/biostatistics/kxr012

Examples

```
data(simaCGH)
```

Index