grpregusim-vignette

testMethods Function

The grpregusim package provides functions to run simulations using lasso, ridge, group lasso, thresholded group lasso, elastic net, sparse group lasso, and group bridge methods. The main function `testMethods` generates design matrix with user-specified grouping structure and runs user-specified group regularization methods. The function takes in nine arguments:

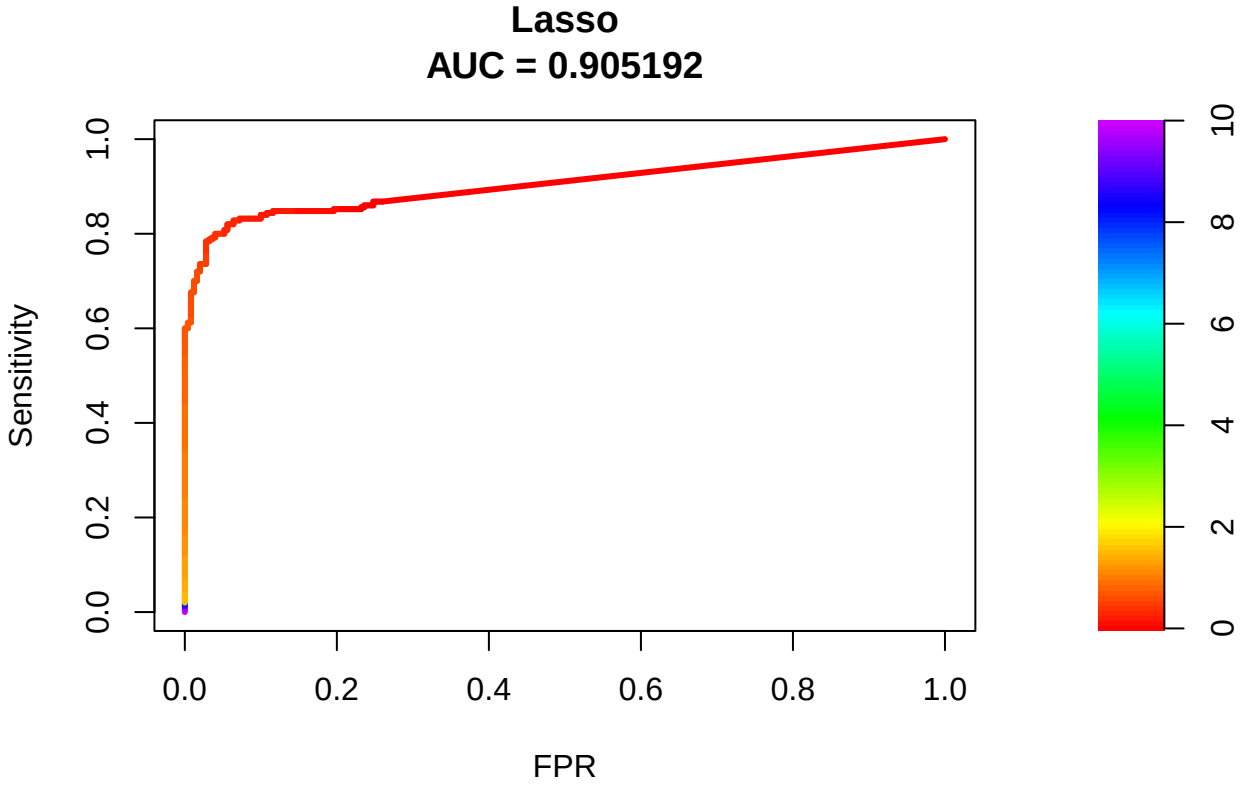
- `a`: Within group correlation (value from -1 to 1) for data generation
- `b`: Between group correlation (value from -1 to 1) for data generation
- `n`: Sample size for data generation
- `p`: Number of input variables for data generation
- `ng`: Number of groups for data generation
- `truegroup`: True grouping structure of input variables (p-dim vector with numerical true group labels)
- `withinGroupNoise`: Proportion of coefficients that switch from category 0 to category 1
- `chooseBetas`: True or false variable that indicates to take only last ng coefficients from each group to run methods on
- `methods`: Specified method ("lasso","ridge","group lasso","threshold group lasso","sparse group lasso","elastic net","group bridge")

An example using this function to do lasso regression:

```
library(grpregusim)
y=testMethods(a=0.9,b = 0.3,n=600,p=500,ng=10,methods="lasso"
)
```

```
## [1] 600 500
## [1] "Lasso================================="
## L2diff= 20.22543
## Sign match percentage = 0.866
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 223  40
##          1  27 210
##
##                Accuracy : 0.866
##                  95% CI : (0.833, 0.8946)
##     No Information Rate : 0.5
```

```
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.732
##
##  Mcnemar's Test P-Value : 0.1426
##
##             Sensitivity : 0.8920
##             Specificity : 0.8400
##          Pos Pred Value : 0.8479
##          Neg Pred Value : 0.8861
##              Prevalence : 0.5000
##          Detection Rate : 0.4460
##    Detection Prevalence : 0.5260
##       Balanced Accuracy : 0.8660
##
##        'Positive' Class : 0
##
```



L2diff Function

The function will return the L2 difference of the true betas and the predicted betas of the user-specified method, a sign matching percentage, sensitivity/specificity information, and an ROC/AUC curve. The L2 difference is calculated using the `L2diff` function which calculates the l2 difference between two vectors of the same length. The function takes in 2 arguments:

- `a`: A vector
- `b`: A vector An example of this function:

```
L2diff(c(1,2,3,4),c(3,6,9,3))
```

```
## [1] 7.549834
```

SignMatch Function

The `SignMatch` function takes pairs of elements between two vectors and sees whether or not both elements are in the same category (abs value less than 0.2, abs value greater than 0.2) and calculates percentage of pairs that are in the same category. The function takes in 2 arguments:

- `a`: A vector
- `b`: A vector

An example of this function:

```
SignMatch(c(1,2,3,4),c(3,6,9,3))
```

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

betaRound Function

The `betaRound` function rounds each element in the vector (to 0 or 1) based off of a threshold (abs value <0.2). Takes in a vector and returns the rounded vector.

- `a`: A vector

An example of this function:

```
betaRound(c(1.30,0.02,-1.17))
```

```
## [1] 1 0 1
```

thresh_within Function

The `thresh_within` function runs thresholded group lasso on group lasso output. Created by: Sumanta Basu. The function takes in 3 arguments:

- `grp.ind`: a p-dim vector of group indices, e.g. 1, 1, 1, 2, 2, 3, 3, 3, 4, 4
- `est`: a p-dim vector (beta-hat) containing the group lasso solution
- `delta`: scalar, controlling the amount of threshold

```
thresh_within(grp.ind = c(0,0,1,1,2,2,3), est =as.matrix(c(1,2,3,4,7,6,5),ncol=1),delta = 0.01)
```

```
##      [,1]
## [1,]    1
## [2,]    2
## [3,]    3
## [4,]    4
## [5,]    7
## [6,]    6
## [7,]    5
```

The function returns the predicted thresholded group lasso values.