Matlab Package: Envelope

May 13, 2012

Contents

Part I
tools

1 bootstrapse

Perform bootstrap to estimate actual standard errors for models in the envelope family.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
bootse = bootstrapse(X, Y, u, B, modelType)
bootse = bootstrapse(X, Y, u, B, modelType, Opts)
```

Input

X: Predictors. The predictors can be univariate or multivariate, discrete or continuous.

For model type for method 'env', 'henv', 'ienv', ' senv', and 'xenv'. X is an n by p matrix, p is the number of predictors.

For model type 'penv', X is A list containing the value of X1 and X2.

- X.X1 (only for 'penv'): Predictors of main interest. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X.X2 (only for 'penv'): Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

u: Dimension of the envelope subspace. The legitimate range of u depends on the model specified.

B: Number of bootstrap samples. A positive integer.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv', 'senv' and 'xenv'.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out the number of bootstrap samples, logical 0 or 1. Default value: 0.

Output

bootse: For 'env', 'henv', 'ienv', 'senv' and 'xenv', an r by p matrix containing the standard errors for elements in β computed by bootstrap. For 'penv', an r by p1 matrix containing the standard errors for β_1 computed by bootstrap.

Description

This function computes the bootstrap standard errors for the regression coefficients or for partial envelope model, the main regression coefficients in the specified model by bootstrapping the residuals.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha);
B = 100;
modelType = 'env';
bootse = bootstrapse(X, Y, u, B, modelType)
```

```
bootse =

    0.2896
    0.4352
    0.3189
    0.5735
    0.2543
    0.5840
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
alpha = 0.01;
u = lrt_penv(X, Y, alpha);
B = 100;
modelType = 'penv';
bootse = bootstrapse(X, Y, u, B, modelType)
```

```
bootse =

    0.0027
    0.0012
    0.0020
    0.0009
```

2 bootstrapse_OLS

Compute bootstrap standard error for ordinary least squares.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
bootse = bootstrapse_OLS(X, Y, B)
```

Input

X: Predictors, an n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

B: Number of bootstrap samples. A positive integer.

Opts: A list containing the optional input parameter. If not defined, the default setting is used.

- Opts.verbose: Flag for print out the number of bootstrap samples, logical 0 or 1. Default value: 0.

Output

bootse: The standard error for elements in β computed by bootstrap. An r by p matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in ordinary least squares by bootstrapping the residuals.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1 : 6);
bootse = bootstrapse_OLS(X, Y, 200)
```

```
bootse =

   10.2168
    8.3940
    9.0503
    9.9677
   14.5822
    5.5874
```

3 modelselectaic

Select the dimension for the envelope family using Akaike information criteria.

Contents

Syntax

```
u = modelselectaic(X, Y, modelType)
u = modelselectaic(X, Y, modelType, Opts)
```

Input

X: Predictors. The predictors can be univariate or multivariate, discrete or continuous.

For model type for method 'env', 'henv', 'ienv', ' senv', and 'xenv'. X is an n by p matrix, p is the number of predictors.

For model type 'penv', X is A list containing the value of X1 and X2.

- X.X1 (only for 'penv'): Predictors of main interest. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X.X2 (only for 'penv'): Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv', 'senv' and 'xenv'.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.

- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the envelope subspace for method 'env', 'henv', 'ienv', 'penv', 'senv', and 'xenv'.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1 : 6);
modelType = 'env';
u = modelselectaic(X, Y, modelType)
```

```
u =

     1
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
modelType = 'penv';
u = modelselectaic(X, Y, modelType)
```

```
u =

     3
```

4 modelselectbic

Select the dimension for the envelope family using Bayesian information criteria.

Contents

Syntax

```
u = modelselectbic(X, Y, modelType)
u = modelselectbic(X, Y, modelType, Opts)
```

Input

X: Predictors. The predictors can be univariate or multivariate, discrete or continuous.

For model type for method 'env', 'henv', 'ienv', ' senv', and 'xenv'. X is an n by p matrix, p is the number of predictors.

For model type 'penv', X is A list containing the value of X1 and X2.

- X.X1 (only for 'penv'): Predictors of main interest. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X.X2 (only for 'penv'): Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv', 'senv' and 'xenv'.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.

- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the envelope subspace for method 'env', 'henv', 'ienv', 'penv', 'senv', and 'xenv'.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1 : 6);
modelType = 'env';
u = modelselectbic(X, Y, modelType)
```

```
u =

     1
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
modelType = 'penv';
u = modelselectbic(X, Y, modelType)
```

```
u =

     1
```

5 modelselectlrt

Select the dimension for the envelope family using likelihood ratio testing procedure.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
u = modelselectlrt(X, Y, alpha, modelType)
u = modelselectlrt(X, Y, alpha, modelType, Opts)
```

Input

X: Predictors. The predictors can be univariate or multivariate, discrete or continuous.

For model type for method 'env', 'henv', 'ienv', and 'xenv'. X is an n by p matrix, p is the number of predictors.

For model type 'penv', X is A list containing the value of X1 and X2.

- X.X1 (only for 'penv'): Predictors of main interest. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X.X2 (only for 'penv'): Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv' and 'xenv'.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the envelope subspace for method 'env', 'henv', 'ienv', 'penv', and 'xenv'. The likelihood ratio resting procedure does not support 'senv', because the scaled envelope models are not nested with the standard model.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1 : 6);
alpha = 0.01;
modelType = 'env';
u = modelselectlrt(X, Y, alpha, modelType)
```

```
u =

     1
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
alpha = 0.01;
modelType = 'penv';
u = modelselectlrt(X, Y, alpha, modelType)
```

```
u =

     1
```

6 prediction

Perform estimation or prediction for models in the envelope family.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
PredictOutput = prediction(ModelOutput, Xnew, infType, modelType)
```

Input

ModelOutput: A list containing the model outputs from fitting the models.

Xnew: The value of X with which to estimate or predict Y.

For 'env', 'henv', 'ienv', 'senv' and 'xenv', it is a p by 1 vector.

For 'penv', it is a list containing the value of X1 and X2.

```
* Xnew.X1 (only for 'penv'): A p1 by 1 vector containing the value of X1.
* Xnew.X2 (only for 'penv'): A p2 by 1 vector containing the value of X2.
```

infType: A string of characters indicting the inference type, the choices can be 'estimation' or 'prediction'.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv', 'senv' and 'xenv'.

Output

PredictOutput: A list containing the results of the inference.

- PredictOutput.value: The fitted value or the prediction value evaluated at Xnew. An r by 1 vector.
- PredictOutput.covMatrix: The covariance matrix of PredictOutput.value. An r by r matrix.
- PredictOutput.SE: The standard error of elements in PredictOutput.value. An r by 1 vector.

Description

This function evaluates the user-specified model, could be 'env', 'henv', 'ienv', 'penv', 'senv' or 'xenv', at new value Xnew. It can perform estimation: find the fitted value when X = Xnew, or prediction: predict Y when X = Xnew. The covariance matrix and the standard errors are also provided.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
modelType = 'env';
u =  modelselectbic(X, Y, modelType);
ModelOutput = env(X, Y, u);
Xnew = X(2, :)';
PredictOutput = predict_env(ModelOutput, Xnew, 'estimation')
[PredictOutput.value, Y(2, :)'] % Compare the fitted value with
the observed value
```

```
PredictOutput =

        value: [6x1 double]
    covMatrix: [6x6 double]
           SE: [6x1 double]


ans =

  474.7135  458.0000
  127.4740  112.0000
  251.2044  236.0000
  380.8280  368.0000
  380.9473  383.0000
   -6.3287  -15.0000
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
modelType = 'penv';
```

```
u =  modelselectbic(X, Y, modelType);
ModelOutput = penv(X, Y, u);
Xnew.X1 = X.X1(1, :)';
Xnew.X2 = X.X2(1, :)';
PredictOutput = predict_penv(ModelOutput, Xnew, 'estimation')
PredictOutput.SE
```

```
PredictOutput =

        value: [4x1 double]
    covMatrix: [4x4 double]
           SE: [4x1 double]


ans =

    1.4680
    0.4234
    0.7145
    0.3161
```

7 testcoefficient

```
This function tests the null hypothesis L * beta * R = A versus the
alternative hypothesis L * beta * R ~= A, where beta is estimated under
the model in the envelope family.
```

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
TestOutput = testcoefficient(ModelOutput, modelType)
TestOutput = testcoefficient(ModelOutput, modelType, TestInput)
```

Input

ModelOutput: A list containing the model outputs from fitting the models.

modelType: A string characters indicting the model, choices can be 'env', 'henv', 'ienv', 'penv', 'senv' and 'xenv'.

TestInput: A list that specifies the null hypothesis, including L, R, and A. If not provided by the user, default values will be used.

- TestInput.L: The matrix multiplied to β on the left. According to different model, it has different size requirement. Default value will be set if the user does not specify.
- TestInput.R: The matrix multiplied to β on the right. According to different model, it has different size requirement. Default value will be set if the user does not specify.
- TestInput.A: The matrix on the right hand side of the equation. Default value will be set if the user does not specify.

Output

TestOutput: A list containing test statistics, degrees of freedom for the reference chi-squared distribution, and the p-value. At the same time, a table is printed out.

- TestOutput.chisqStatistic: The test statistics. A real number.
- TestOutput.df: The degrees of freedom of the reference chi-squared distribution. A positive integer.
- TestOutput.pValue: p-value of the test. A real number in [0, 1].

Description

This function tests for hypothesis $H_0 : L\beta R = A$, versus $H_\alpha : L\beta R \neq A$. The β is estimated by a model in the envelope model. If the user does not specify the values for L, R and A, then the test is equivalent to the standard F test on if $\beta = 0$ (for 'env', 'ienv', 'penv', 'senv' and 'xenv'), or if the group main effects are all zeros (for 'henv'). The test statistics used is vec $(L\beta R - A)\ \hat{\Sigma}^{-1}$ vec $(L\beta R - A)^T$, and the reference distribution is chi-squared distribution with degrees of freedom the same as the length of vec(A).

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha);
ModelOutput = env(X, Y, u);
modelType = 'env';
TestOutout = testcoefficient(ModelOutput, modelType);
```

```
 Test Hypothesis     Chisq Statistic    DF     P-value
----------------------------------------------------------------------
L * beta * R = A         116.230        6      0.0000
----------------------------------------------------------------------
```

```
load fiberpaper.dat
Y = fiberpaper(:, 1 : 4);
Xtemp = fiberpaper(:, 5 : 7);
X.X1 = Xtemp(:, 3);
X.X2 = Xtemp(:, 1 : 2);
alpha = 0.01;
u = lrt_penv(X, Y, alpha);
ModelOutput = penv(X, Y, u);
r = size(Y, 2);
p1 = size(X.X1, 2);
TestInput.L = rand(2, r);
TestInput.R = rand(p1, 1);
TestInput.A = zeros(2, 1);
TestOutout = testcoefficient_penv(ModelOutput, TestInput);
```

Test Hypothesis	Chisq Statistic	DF	P-value
L * beta * R = A	12.598	2	0.0018

Part II

env

8 aic_env

Select the dimension of the envelope subspace using Akaike information criterion.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
u = aic_env(X, Y)
u = aic_env(X, Y, Opts)
```

Input

X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the envelope subspace.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
u = aic_env(X, Y)
```

```
u =

     1
```

9 bic_env

Select the dimension of the envelope subspace using Bayesian information criterion.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
u = bic_env(X, Y)
u = bic_env(X, Y, Opts)
```

Input

X: Predictors. An n by p matrix, p is the number of predictors and n is the number of observations. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the envelope subspace.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
u = bic_env(X, Y)
```

```
u =

     1
```

10 bstrp_env

Compute bootstrap standard error for the envelope model.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
bootse = bstrp_env(X, Y, u, B)
bootse = bstrp_env(X, Y, u, B, Opts)
```

Input

X: Predictors, an n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

u: Dimension of the envelope subspace. A positive integer between 0 and r.

B: Number of bootstrap samples. A positive integer.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out the number of bootstrap samples, logical 0 or 1. Default value: 0.

Output

bootse: The standard error for elements in β computed by bootstrap. An r by p matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the envelope model by bootstrapping the residuals.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha)
```

```
u =

     1
```

```
B = 100;
bootse = bstrp_env(X, Y, u, B)
```

```
bootse =

    0.2893
    0.4260
    0.3523
    0.5628
    0.1675
    0.6192
```

11 dF4env

The first derivative of the objective function for computing the envelope subspace.

Contents

- Syntax
- Input
- Output
- Description

Syntax

```
df = dF4env(R, DataParameter)
```

Input

R: An r by u semi orthogonal matrix, 0<u<=r.

DataParameter: A structure that contains the statistics calculated from the data.

Output

df: An r by u matrix containing the value of the derivative function evaluated at R.

Description

The objective function is derived in Section 4.3 in Cook et al. (2010) by using maximum likelihood estimation. This function is the derivative of the objective function.

12 env

Fit the envelope model.

Contents

- Syntax
- Input
- Output
- Description
- References
- Example

Syntax

```
ModelOutput = env(X, Y, u)
ModelOutput = env(X, Y, u, Opts)
```

Input

X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p.

u: Dimension of the envelope. An integer between 0 and r.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out Grassmann manifold optimization process, logical 0 or 1. Default value: 0.

Output

ModelOutput: A list that contains the maximum likelihood estimators and some statistics.

- ModelOutput.beta: The envelope estimator of the regression coefficients β. An r by p matrix.

- ModelOutput.Sigma: The envelope estimator of the error covariance matrix. An r by r matrix.
- ModelOutput.Gamma: The orthogonal basis of the envelope subspace. An r by u semi-orthogonal matrix.
- ModelOutput.Gamma0: The orthogonal basis of the complement of the envelope subspace. An r by r-u semi-orthogonal matrix.
- ModelOutput.eta: The coordinates of β with respect to Gamma. An u by p matrix.
- ModelOutput.Omega: The coordinates of Sigma with respect to Gamma. An u by u matrix.
- ModelOutput.Omega0: The coordinates of Sigma with respect to Gamma0. An r-u by r-u matrix.
- ModelOutput.alpha: The estimated intercept in the envelope model. An r by 1 vector.
- ModelOutput.l: The maximized log likelihood function. A real number.
- ModelOutput.covMatrix: The asymptotic covariance of vec(β). An rp by rp matrix. The covariance matrix returned are asymptotic. For the actual standard errors, multiply by 1/n.
- ModelOutput.asyEnv: The asymptotic standard error for elements in β under the envelope model. An r by p matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- ModelOutput.ratio: The asymptotic standard error ratio of the standard multivariate linear regression estimator over the envelope estimator, for each element in β. An r by p matrix.
- ModelOutput.np: The number of parameters in the envelope model. A positive integer.
- ModelOutput.n: The number of observations in the data. A positive integer.

Description

This function fits the envelope model to the responses and predictors, using the maximum likelihood estimation. When the dimension of the envelope is between 1 and r-1, we implemented the algorithm in Cook et al. (2010). When the dimension is r, then the envelope model degenerates to the standard multivariate linear regression. When the dimension is 0, it means that X and Y are uncorrelated, and the fitting is different.

References

1. The codes is implemented based on the algorithm in Section 4.3 of Cook et al (2010).
2. The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (http://web.mit.edu/∼ripper/www.sgmin.html).

Example

The following codes will reconstruct the results in the wheat protein data example in Cook et al. (2010).

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha)
```

```
u =

     1
```

```
ModelOutput = env(X, Y, u)
```

```
ModelOutput =

         beta: [6x1 double]
        Sigma: [6x6 double]
        Gamma: [6x1 double]
       Gamma0: [6x5 double]
          eta: 8.5647
        Omega: 7.8762
       Omega0: [5x5 double]
        alpha: [6x1 double]
            l: -850.7592
    covMatrix: [6x6 double]
       asyEnv: [6x1 double]
        ratio: [6x1 double]
           np: 28
            n: 50
```

```
ModelOutput.Omega
```

```
ans =

    7.8762
```

```
eig(ModelOutput.Omega0)
```

```
ans =

   1.0e+03 *

    6.5166
    0.2083
    0.0201
    0.0004
    0.0003
```

```
ModelOutput.ratio
```

```
ans =

   28.0945
   18.4326
   23.6384
   16.3211
   65.8245
    6.4668
```

13 F4env

Objective function for computing the envelope subspace.

Contents

- Syntax
- Input
- Output
- Description

Syntax

```
f = F4env(R, DataParameter)
```

Input

R: An r by u semi orthogonal matrix, 0 < u <= r.

DataParameter: A structure that contains the statistics calculated from the data.

Output

f: A scalar containing the value of the objective function evaluated at R.

Description

The objective function is derived in Section 4.3 of Cook et al. (2010) using maximum likelihood estimation. The columns of the semi-orthogonal matrix that minimizes this function span the estimated envelope subspace.

14 lrt_env

Select the dimension of the envelope subspace using likelihood ratio testing.

Contents

Syntax

```
u = lrt_env(X, Y, alpha)
u = lrt_env(X, Y, alpha, Opts)
```

Input

X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.

Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

Opts: A list containing the optional input parameter, to control the iterations in sg_min. If one or several (even all) fields are not defined, the default settings are used.

- Opts.maxIter: Maximum number of iterations. Default value: 300.
- Opts.ftol: Tolerance parameter for F. Default value: 1e-10.
- Opts.gradtol: Tolerance parameter for dF. Default value: 1e-7.
- Opts.verbose: Flag for print out dimension selection process, logical 0 or 1. Default value: 0.

Output

u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the envelope subspace, with pre-specified significance level α.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha)
```

```
u =

     1
```

15 predict_env

Perform estimation or prediction under the envelope model.

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
PredictOutput = predict_env(ModelOutput, Xnew, infType)
```

Input

ModelOutput: A list containing the maximum likelihood estimators and other statistics inherited from env.

Xnew: The value of X with which to estimate or predict Y. A p by 1 vector.

infType: A string of characters indicting the inference type, the choices can be 'estimation' or 'prediction'.

Output

PredictOutput: A list containing the results of the inference.

- PredictOutput.value: The fitted value or the prediction value evaluated at Xnew. An r by 1 vector.
- PredictOutput.covMatrix: The covariance matrix of PredictOutput.value. An r by r matrix.
- PredictOutput.SE: The standard error of elements in PredictOutput.value. An r by 1 vector.

Description

This function evaluates the envelope model at new value Xnew. It can perform estimation: find the fitted value when X = Xnew, or prediction: predict Y when X = Xnew. The covariance matrix and the standard errors are also provided.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha);
ModelOutput = env(X, Y, u);
Xnew = X(2, :)';
PredictOutput = predict_env(ModelOutput, Xnew, 'estimation')
[PredictOutput.value, Y(1, :)']  % Compare the fitted value with the data
PredictOutput.SE
```

```
PredictOutput =

        value: [6x1 double]
    covMatrix: [6x6 double]
           SE: [6x1 double]


ans =

  474.7135  468.0000
  127.4740  123.0000
  251.2044  246.0000
  380.8280  374.0000
  380.9473  386.0000
   -6.3287  -11.0000


ans =

    4.8892
    4.0227
    4.3237
    4.7470
    6.8186
    2.6948
```

```
PredictOutput = predict_env(ModelOutput, Xnew, 'prediction')
PredictOutput.SE
```

```
PredictOutput =
```

```
        value: [6x1 double]
    covMatrix: [6x6 double]
           SE: [6x1 double]


ans =

  474.7135
  127.4740
  251.2044
  380.8280
  380.9473
   -6.3287


ans =

   34.9161
   28.7280
   30.8775
   33.9006
   48.6945
   19.2448
```

16 testcoefficient_env

`This function tests the null hypothesis L * beta * R = A versus the alternative hypothesis L * beta * R ~= A, where beta is estimated under the envelope model.`

Contents

- Syntax
- Input
- Output
- Description
- Example

Syntax

```
TestOutput = testcoefficient_env(ModelOutput)
TestOutput = testcoefficient_env(ModelOutput, TestInput)
```

Input

ModelOutput: A list containing the maximum likelihood estimators and other statistics inherited from env.

TestInput: A list that specifies the null hypothesis, including L, R, and A. If not provided by the user, default values will be used.

- TestInput.L: The matrix multiplied to β on the left. It is a d1 by r matrix, while d1 is less than or equal to r. Default value: identity matrix I_r.
- TestInput.R: The matrix multiplied to β on the right. It is a p by d2 matrix, while d2 is less than or equal to p. Default value: identity matrix I_p.
- TestInput.A: The matrix on the right hand side of the equation. It is a d1 by d2 matrix. Default value: d1 by d2 zero matrix.

Output

TestOutput: A list containing test statistics, degrees of freedom for the reference chi-squared distribution, and the p-value. At the same time, a table is printed out.

- TestOutput.chisqStatistic: The test statistics. A real number.
- TestOutput.df: The degrees of freedom of the reference chi-squared distribution. A positive integer.
- TestOutput.pValue: p-value of the test. A real number in [0, 1].

Description

This function tests for hypothesis $H_0 : L\beta R = A$, versus $H_\alpha : L\beta R \neq A$. The β is estimated by the envelope model. If the user does not specify the values for L, R and A, then the test is equivalent to the standard F test on if $\beta = 0$. The test statistics used is vec $(L\beta R - A)$ $\hat{\Sigma}^{-1}$ vec $(L\beta R - A)^T$, and the reference distribution is chi-squared distribution with degrees of freedom d1 * d2.

Example

```
load wheatprotein.txt
X = wheatprotein(:, 8);
Y = wheatprotein(:, 1:6);
alpha = 0.01;
u = lrt_env(X, Y, alpha);
ModelOutput = env(X, Y, u);
TestOutout = testcoefficient_env(ModelOutput);
```

```
 Test Hypothesis     Chisq Statistic    DF     P-value
----------------------------------------------------------------------
L * beta * R = A         116.230        6      0.0000
----------------------------------------------------------------------
```

```
r = size(Y, 2);
p = size(X, 2);
TestInput.L = rand(2, r);
TestInput.R = rand(p, 1);
TestInput.A = zeros(2, 1);
TestOutout = testcoefficient_senv(ModelOutput, TestInput);
```

```
 Test Hypothesis     Chisq Statistic    DF     P-value
----------------------------------------------------------------------
L * beta * R = A          61.344        2      0.0000
----------------------------------------------------------------------
```

Part III
henv

17 aic_henv

Select the dimension of the envelope subspace using Akaike information criterion for the heteroscedastic envelope model.

Contents

- Usage
- Description
- Example

Usage

u=aic_henv(X,Y)

Input

- X: Group indicators. An n by p matrix, p is the number of groups. X can only take p different values, one for each group.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the envelope subspace for the heteroscedastic envelope model.

Example

load waterstrider.mat
u=aic_henv(X,Y)

18 bic_henv

Select the dimension of the envelope subspace using Bayesian information criterion for the heteroscedastic envelope model.

Contents

- Usage
- Description
- Example

Usage

u=bic_henv(X,Y)

Input

- X: Group indicators. An n by p matrix, p is the number of groups. X can only take p different values, one for each group.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the envelope subspace for the heteroscedastic envelope model.

Example

load waterstrider.mat
u=bic_henv(X,Y)

19 bstrp_henv

Compute bootstrap standard error for the heteroscedastic envelope model.

Contents

- Usage
- Description
- Example

Usage

bootse=bstrp_henv(X,Y,B,u)

Input

- X: Group indicators. An n by p matrix, p is the number of groups. X can only take p different values, one for each group.
- Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- B: Number of boostrap samples. A positive integer.
- u: Dimension of the envelope subspace. A positive integer between 0 and r.

Output

- bootse: The standard error for elements in β computed by bootstrap. An r by p matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the heteroscedastic envelope model by bootstrapping the residuals.

Example

load waterstrider.mat
u=lrt_henv(X,Y,0.01)
B=100;
bootse=bstrp_henv(X,Y,B,u)

20 dF4henv

The first derivative of the objective funtion for computing the envelope subspace in the heteroscedastic envelope model.

Contents

Usage

df = dF4henv(R,DataParameter)

Input

- R: An r by u semi orthogonal matrix, 0<u<=r.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- df: An r by u containing the value of the derivative function evaluated at R.

Description

The objective function is derived in Section 2.2 in Su and Cook (2012) by using maximum likelihood estimation. This function is the derivative of the objective function.

21 F4henv

Objective funtion for computing the envelope subspace in heteroscedastic envelope model.

Contents

- Usage
- Description

Usage

f = F4henv(R,DataParameter)

Input

- R: An r by u semi orthogonal matrix, 0<u<=r.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- f: A scalar containing the value of the objective function evaluated at R.

Description

The objective function is derived in Section 2.2 of Su and Cook (2012) using maximum likelihood estimation. The columns of the semi-orthogonal matrix that minimizes this function span the estimated envelope subspace in the heteroscedastic envelope model.

22 henv

Fit the heteroscedastic envelope model.

Contents

- Usage
- Description
- References
- Example

Usage

stat=henv(X,Y,u)

Input

- X: Group indicators. An n by p matrix, p is the number of groups. X can only take p different values, one for each group.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p.
- u: Dimension of the envelope. An integer between 0 and r.

Output

stat: A list that contains the maximum likelihood estimators and some statistics.

- stat.mu: The heteroscedastic envelope estimator of the grand mean. A r by 1 vector.
- stat.mug: The heteroscedastic envelope estimator of the group mean. A r by p matrix, the ith column of the matrix contains the mean for the ith group.
- stat.Yfit: A n by r matrix, the ith row gives the group mean of the group that the ith observation belongs to. As X is just a group indicator, and is not ordinal, stat.mug alone does not tell which group corresponds to which group mean.
- stat.Gamma: The orthogonal basis of the envelope subspace. An r by u semi-orthogonal matrix.
- stat.Gamma0: The orthogonal basis of the complement of the envelope subspace. An r by r-u semi-orthogonal matrix.
- stat.beta: The heteroscedastic envelope estimator of the group mean effect. An r by p matrix, the ith column of the matrix contains the main effect for the ith group.

- stat.Sigma: The heteroscedastic envelope estimator of the error covariance matrix. A three dimensional matrix with dimension r, r and p, stat.Sigma(:,:,i) contains the estimated covariance matrix for the ith group.
- stat.eta: The coordinates of β with respect to Gamma. An u by p matrix, the ith column contains the coordinates of the main effect of the ith group with respect to Gamma.
- stat.Omega: The coordinates of Sigma with respect to Gamma. An u by u by p matrix, stat.Omega(:,:,i) contains the coordinates of the covariance matrix of the ith group with respect to Gamma.
- stat.Omega0: The coordinates of Sigma with respect to Gamma0. An r-u by r-u matrix.
- stat.l: The maximized log likelihood function. A real number.
- stat.np: The number of parameters in the heteroscedastic envelope model. A positive integer.
- stat.asyHenv: The asymptotic standard errors for elements in *beta* under the heteroscedastic envelope model. An r by p matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- stat.ratio: The asymptotic standard error ratio of the standard multivariate linear regression estimator over the heteroscedastic envelope estimator. An r by p matrix, the (i, j)th element in stat.ratio is the elementwise standard error ratio for the ith element in the jth group mean effect.

Description

This function fits the heteroscedatic envelope model to the responses and predictors, using the maximum likehood estimation. When the dimension of the envelope is between 1 and r-1, we implemented the algorithm in Su and Cook (2012). When the dimension is r, then the envelope model degenerates to the standard multivariate linear model for comparing group means. When the dimension is 0, it means there is not any group effect, and the fitting is different.

References

- The codes is implemented based on the algorithm in Section 2.2 of Su and Cook (2012).
- The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (http://web.mit.edu/~ripper/www.sgmin.html).

Example

The following codes produce the results of the waterstrider example in Su and Cook (2011).

```
load waterstrider.mat
u=lrt_henv(X,Y,0.01)
stat=henv(X,Y,u)
stat.ratio
```

23 lrt_henv

Select the dimension of the envelope subspace using likelihood ratio testing for the heteroscedastic envelope model.

Contents

- Usage
- Description
- Example

Usage

u=lrt_henv(X,Y,alpha)

Input

- X: Group indicators. An n by p matrix, p is the number of groups. X can only take p different values, one for each group.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

Output

- u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the envelope subspace in heteroscedastic envelope model, with prespecified significance level α.

Example

load waterstrider.mat
u=lrt_henv(X,Y,0.01)

Part IV
ienv

24 aic_ienv

Select the dimension of the inner envelope subspace using Akaike information criterion.

Contents

- Usage
- Description
- Example

Usage

u=aic_ienv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors and n is the number of observations. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.

Output

- u: Dimension of the inner envelope. An integer between 0 and p or equal to r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the inner envelope subspace.

Example

load irisf.mat
u=aic_ienv(X,Y)

25 bic_ienv

Select the dimension of the inner envelope subspace using Bayesian information criterion.

Contents

- Usage
- Description
- Example

Usage

u=bic_ienv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors and n is the number of observations. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.

Output

- u: Dimension of the inner envelope. An integer between 0 and p or equal to r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the inner envelope subspace.

Example

load irisf.mat
u=bic_ienv(X,Y)

26 bstrp_ienv

Compute bootstrap standard error for the inner envelope model.

Contents

- Usage
- Description
- Example

Usage

bootse=bstrp_ienv(X,Y,B,u)

Input

- X: Predictors, an n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- B: Number of boostrap samples. A positive integer.
- u: Dimension of the inner envelope. An integer between 0 and p or equal to r.

Output

- bootse: The standard error for elements in β computed by bootstrap. An r by p matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the inner envelope model by bootstrapping the residuals.

Example

load irisf.mat

u=bic_ienv(X,Y)
B=100;
bootse=bstrp_ienv(X,Y,B,u)

27 dF4ienv

First derivative of the objective funtion for computing the inner envelope subspace.

Contents

- Usage
- Description

Usage

f = dF4ienv(R,DataParameter)

Input

- R: An r by u semi-orthogonal matrix, 0<u<=p.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- dF: The first derivative of the objective function for computing the inner envelope subspace. An r by u matrix.

Description

This first derivative of F4ienv obtained by matrix calculus calculations.

28 F4ienv

Objective funtion for computing the inner envelope subspace

Contents

- Usage
- Description

Usage

f = F4ienv(R,DataParameter)

Input

- R: An r by u semi orthogonal matrix, 0<u<=p.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- f: A scalar containing the value of the objective function evaluated at R.

Description

The objective function is derived in Section 3.3 in Su and Cook (2012) by using maximum likelihood estimation. The columns of the semi-orthogonal matrix that minimizes this function span the estimated inner envelope subspace.

29 ienv

Fit the inner envelope model.

Contents

- Usage
- Description
- References
- Example

Usage

stat=ienv(X,Y,u)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p.
- u: Dimension of the inner envelope. An integer between 0 and p or equal to r.

Output

stat: A list that contains the maximum likelihood estimators and some statistics.

- stat.beta: The envelope estimator of the regression coefficients β. An r by p matrix.
- stat.Sigma: The envelope estimator of the error covariance matrix. An r by r matrix.
- stat.Gamma1: The orthogonal basis of the inner envelope subspace. An r by u semi-orthogonal matrix.
- stat.Gamma0: The orthogonal basis of the complement of the inner envelope subspace. An r by r-u semi-orthogonal matrix.
- stat.eta1: The transpose of the coordinates of β with respect to Gamma1. An p by u matrix.
- stat.B: An (r-u) by (p-u) semi-orthogonal matrix, so that (Gamma, Gamma0*B) spans β.
- stat.eta2: The transpose of the coordinates of β with respect to Gamma0. An p by (p-u) matrix.
- stat.Omega1: The coordinates of Sigma with respect to Gamma1. An u by u matrix.

- stat.Omega0: The coordinates of Sigma with respect to Gamma0. An r-u by r-u matrix.
- stat.alpha: The estimated intercept in the inner envelope model. An r by 1 vector.
- stat.l: The maximized log likelihood function. A real number.
- stat.asyIenv: Asymptotic standard error for elements in β under the inner envelope model. An r by p matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- stat.ratio: The asymptotic standard error ratio of the stanard multivariate linear regression estimator over the inner envelope estimator, for each element in β. An r by p matrix.
- stat.np: The number of parameters in the inner envelope model. A positive integer.

Description

This function fits the inner envelope model to the responses and predictors, using the maximum likehood estimation. When the dimension of the envelope is between 1 and p-1, we implemented the algorithm in Su and Cook (2012). When the dimension is p, then the inner envelope model degenerates to the standard multivariate linear regression. When the dimension is 0, it means that X and Y are uncorrelated, and the fitting is different.

References

- The codes is implemented based on the algorithm in Su and Cook (2012).
- The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (http://web.mit.edu/∼ripper/www.sgmin.html).

Example

The following codes gives the results of the Fisher's iris data example in Su and Cook (2012).

load irisf.mat

u=bic_env(X,Y)
d=bic_ienv(X,Y)
stat=ienv(X,Y,d)
1-1./stat.ratio

30 lrt_ienv

Select the dimension of the inner envelope subspace using likelihood ratio testing.

Contents

- Usage
- Description
- Example

Usage

u=lrt_ienv(X,Y,alpha)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

Output

- u: Dimension of the inner envelope. An integer between 0 and p or equal to r.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the inner envelope subspace, with prespecified significance level α.

Example

load irisf.mat

alpha=0.01;
u=lrt_ienv(X,Y,alpha)

Part V
penv

31 aic_penv

Select the dimension of the partial envelope subspace using Akaike information criterion.

Contents

- Usage
- Description
- Example

Usage

u=aic_penv(X1,X2,Y)

Input

- X1: Predictors of main interst. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X2: Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the partial envelope subspace.

Example

load T7-7.dat
Y=T7_7(:,1:4);
X=T7_7(:,5:7);

```
X1=X(:,3);
X2=X(:,1:2);
u=aic_penv(X1,X2,Y)
```

32 bic_penv

Select the dimension of the partial envelope subspace using Bayesian information criterion.

Contents

- Usage
- Description
- Example

Usage

u=bic_penv(X1,X2,Y)

Input

- X1: Predictors of main interst. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X2: Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the partial envelope subspace.

Example

```
load T7-7.dat
Y=T7_7(:,1:4);
X=T7_7(:,5:7);
X1=X(:,3);
X2=X(:,1:2);
u=bic_penv(X1,X2,Y)
```

33 bstrp_penv

Compute bootstrap standard error for the partial envelope model.

Contents

- Usage
- Description
- Example

Usage

bootse=bstrp_penv(X1,X2,Y,B,u)

Input

- X1: Predictors of main interst. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X2: Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates. The covariates can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- B: Number of boostrap samples. A positive integer.
- u: Dimension of the partial envelope subspace. A positive integer between 0 and r.

Output

- bootse: The standard error for elements in β_1 computed by bootstrap. An r by p1 matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the partial envelope model by bootstrapping the residuals.

Example

load T7-7.dat
Y=T7_7(:,1:4);
X=T7_7(:,5:7);

```
X1=X(:,3);
X2=X(:,1:2);
alpha=0.01;
u=lrt_penv(X1,X2,Y,alpha)
B=100;
bootse=bstrp_penv(X1,X2,Y,B,u)
```

34 lrt_penv

Select the dimension of the partial envelope subspace using likelihood ratio testing.

Contents

- Usage
- Description
- Example

Usage

u=lrt_penv(X1,X2,Y,alpha)

Input

- X1: Predictors of main interst. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X2: Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates.
- Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.
- alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

Output

- u: Dimension of the partial envelope subspace. An integer between 0 and r.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the partial envelope subspace, with prespecified significance level α.

Example

```
load T7-7.dat
Y=T7_7(:,1:4);
X=T7_7(:,5:7);
X1=X(:,3);
```

```
X2=X(:,1:2);
alpha=0.01;
u=lrt_penv(X1,X2,Y,alpha)
```

35 penv

Fit the partial envelope model.

Contents

- Usage
- Description
- References
- Example

Usage

stat=penv(X1,X2,Y,u)

Input

- X1: Predictors of main interst. An n by p1 matrix, n is the number of observations, and p1 is the number of main predictors. The predictors can be univariate or multivariate, discrete or continuous.
- X2: Covariates, or predictors not of main interest. An n by p2 matrix, p2 is the number of covariates. The covariates can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p1.
- u: Dimension of the partial envelope. An integer between 0 and r.

Output

stat: A list that contains the maximum likelihood estimators and some statistics.

- stat.beta1: The partial envelope estimator of β_1, which is the regression coefficients for X1. An r by p1 matrix.
- stat.beta2: The partial envelope estimator of β_2, which is the regression coefficients for X2. An r by p2 matrix.
- stat.Sigma: The partial envelope estimator of the error covariance matrix. An r by r matrix.
- stat.Gamma: The orthogonal basis of the partial envelope subspace. An r by u semi-orthogonal matrix.
- stat.Gamma0: The orthogonal basis of the complement of the partial envelope subspace. An r by r-u semi-orthogonal matrix.
- stat.eta: The coordinates of β_1 with respect to Gamma. An u by p1 matrix.

- stat.Omega: The coordinates of Sigma with respect to Gamma. An u by u matrix.
- stat.Omega0: The coordinates of Sigma with respect to Gamma0. An r-u by r-u matrix.
- stat.alpha: The estimated intercept in the partial envelope model. An r by 1 vector.
- stat.l: The maximized log likelihood function. A real number.
- stat.asyPenv: Asymptotic standard error for elements in β under the partial envelope model. An r by p1 matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- stat.ratio: The asymptotic standard error ratio of the stanard multivariate linear regression estimator over the partial envelope estimator, for each element in β_1. An r by p1 matrix.
- stat.np: The number of parameters in the envelope model. A positive integer.

Description

This function fits the partial envelope model to the responses Y and predictors X1 and X2, using the maximum likehood estimation. When the dimension of the envelope is between 1 and r-1, we implemented the algorithm in Su and Cook (2011). When the dimension is r, then the partial envelope model degenerates to the standard multivariate linear regression with Y as the responses and both X1 and X2 as predictors. When the dimension is 0, X1 and Y are uncorrelated, and the fitting is the standard multivariate linear regression with Y as the responses and X2 as the predictors.

References

- The codes is implemented based on the algorithm in Section 3.2 of Su and Cook (2012).
- The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (http://web.mit.edu/∼ripper/www.sgmin.html).

Example

The following codes reconstruct the results of the paper and fiber example in Su and Cook (2012).

```
load T7-7.dat
Y=T7_7(:,1:4);
X=T7_7(:,5:7);
X1=X(:,3);
X2=X(:,1:2);
alpha=0.01;
u=lrt_penv(X1,X2,Y,alpha)
```

```
stat=penv(X1,X2,Y,u)
stat.Omega
eig(stat.Omega0)
stat.ratio
```

Part VI
senv

36 aic_senv

Select the dimension of the scaled envelope subspace using Akaike information criterion.

Contents

- Usage
- Description
- Example

Usage

u=aic_senv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors and n is the number of observations. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.

Output

- u: Dimension of the inner envelope. An integer between 0 and r.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the scaled envelope subspace.

Example

```
load('sales.txt')
Y=T9_12(:,4:7);
X=T9_12(:,1:3);
u=aic_senv(X,Y)
```

37 bic_senv

Select the dimension of the scaled envelope subspace using Bayesian information criterion.

Contents

- Usage
- Description
- Example

Usage

u=bic_senv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors and n is the number of observations. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses. The responses must be continuous variables.

Output

- u: Dimension of the inner envelope. An integer between 0 and r.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the scaled envelope subspace.

Example

```
load('sales.txt')
Y=T9_12(:,4:7);
X=T9_12(:,1:3);
u=bic_senv(X,Y)
```

38 bstrp_senv

Compute bootstrap standard error for the scaled envelope model.

Contents

- Usage
- Description
- Example

Usage

bootse=bstrp_senv(X,Y,B,u)

Input

- X: Predictors, an n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- B: Number of boostrap samples. A positive integer.
- u: Dimension of the envelope subspace. A positive integer between 0 and r.

Output

- bootse: The standard error for elements in β computed by bootstrap. An r by p matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the scaled envelope model by bootstrapping the residuals.

Example

```
load('sales.txt')
Y=T9_12(:,4:7);
X=T9_12(:,1:3);
u=bic_ienv(X,Y)
B=20;
bootse=bstrp_senv(X,Y,B,u)
```

39 dF4senv

First derivative of the objective funtion for computing the envelope subspace in the scaled envelope model.

Contents

- Usage
- Description

Usage

f = dF4senv(R,DataParameter)

Input

- R: An r by u semi-orthogonal matrix, 0<u<=p.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- dF: The first derivative of the objective function for computing the envelope subspace. An r by u matrix.

Description

This first derivative of F4senv obtained by matrix calculus calculations.

40 F4senv

Objective funtion for computing the envelope subspace in scaled envelope model.

Contents

- Usage
- Description

Usage

f = F4senv(R,DataParameter)

Input

- R: An r by u semi orthogonal matrix, 0<u<r.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- f: A scalar containing the value of the objective function evaluated at R.

Description

The objective function is derived in Section 4.1 in Cook and Su (2012) using maximum likelihood estimation. The columns of the semi-orthogonal matrix that minimizes this function span the estimated envelope subspace.

41 objfun

Objective funtion for computing the scales in the scaled envelope model.

Contents

- Usage
- Description

Usage

f = objfun(d,Gamma,DataParameter)

Input

- d: An r-1 dimensional column vector containing the scales for the 2nd to the rth responses. All the entries in d are positive.
- Gamma: A r by u semi-orthogomal matrix that spans the envelope subspace or the estimated envelope subspace.
- DataParameter: A structure that contains the statistics calculated form the data.

Output

- f: A scalar containing the value of the objective function evaluated at d.

Description

The objective function is derived in Section 4.1 of Su and Cook (2012) using maximum likelihood estimation.

42 senv

Fit the scaled envelope model.

Contents

- Usage
- Description
- References
- Example

Usage

stat=senv(X,Y,u)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p.
- u: Dimension of the envelope. An integer between 0 and r.

Output

stat: A list that contains the maximum likelihood estimators and some statistics.

- stat.beta: The scaled envelope estimator of the regression coefficients β. An r by p matrix.
- stat.Sigma: The scaled envelope estimator of the error covariance matrix. An r by r matrix.
- stat.Lambda: The matrix of estimated scales. An r by r diagonal matrix with the first diagonal element equal to 1 and other diagonal elements being positive.
- stat.Gamma: The orthogonal basis of the envelope subspace. An r by u semi-orthogonal matrix.
- stat.Gamma0: The orthogonal basis of the complement of the envelope subspace. An r by r-u semi-orthogonal matrix.
- stat.eta: The coordinates of β with respect to Gamma. An u by p matrix.
- stat.Omega: The coordinates of Sigma with respect to Gamma. An u by u matrix.
- stat.Omega0: The coordinates of Sigma with respect to Gamma0. An r-u by r-u matrix.

- stat.alpha: The estimated intercept in the scaled envelope model. An r by 1 vector.
- stat.l: The maximized log likelihood function. A real number.
- stat.asySenv: Asymptotic standard error for elements in β under the scaled envelope model. An r by p matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- stat.ratio: The asymptotic standard error ratio of the standard multivariate linear regression estimator over the scaled envelope estimator, for each element in β. An r by p matrix.
- stat.np: The number of parameters in the scaled envelope model. A positive integer.

Description

This function fits the scaled envelope model to the responses and predictors, using the maximum likehood estimation. When the dimension of the envelope is between 1 and r-1, we implemented the algorithm in Cook and Su (2012). When the dimension is r, then the scaled envelope model degenerates to the standard multivariate linear regression. When the dimension is 0, it means that X and Y are uncorrelated, and the fitting is different.

References

- The codes is implemented based on the algorithm in Section 4.1 of Cook and Su (2012).
- The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (http://web.mit.edu/∼ripper/www.sgmin.html).

Example

The following codes produce the results of the test and performance example in Cook and Su (2012).

```
load('sales.txt')
Y=T9_12(:,4:7);
X=T9_12(:,1:3);
u=bic_env(X,Y)
stat=env(X,Y,u);
1-1./stat.ratio
u=bic_senv(X,Y)
stat=senv(X,Y,u);
stat.Lambda
1-1./stat.ratio
```

Part VII
xenv

aic_xenv

Use Akaike information criterion to select the dimension of the envelope subspace for the reduction on X.

Contents

- Usage
- Description
- Example

Usage

u=aic_xenv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and p.

Description

This function implements the Akaike information criteria (AIC) to select the dimension of the envelope subspace for the reduction on X.

Example

load wheatprotein.txt
X=wheatprotein(:,1:6);
Y=wheatprotein(:,7);
u=aic_xenv(X,Y)

43 bic_xenv

Use Bayesian information criterion to select the dimension of the envelope subspace for the reduction on X.

Contents

- Usage
- Description
- Example

Usage

u=bic_xenv(X,Y)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.

Output

- u: Dimension of the envelope. An integer between 0 and p.

Description

This function implements the Bayesian information criteria (BIC) to select the dimension of the envelope subspace for the reduction on X.

Example

```
load wheatprotein.txt
X=wheatprotein(:,1:6);
Y=wheatprotein(:,7);
u=bic_xenv(X,Y)
```

44 bstrp_xenv

Compute bootstrap standard error of the envelope model for the reduction on X.

Contents

- Usage
- Description
- Example

Usage

bootse=bstrp_xenv(X,Y,B,u)

Input

- X: Predictors, an n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses, an n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- B: Number of boostrap samples. A positive integer.
- u: Dimension of the envelope subspace. A positive integer between 0 and p.

Output

- bootse: The standard error for elements in β computed by bootstrap. An p by r matrix.

Description

This function computes the bootstrap standard errors for the regression coefficients in the envelope model by bootstrapping the residuals. The envelope model here is for the reduction on X.

Example

load wheatprotein.txt
X=wheatprotein(:,1:6);
Y=wheatprotein(:,7);
alpha=0.01;
u=lrt_xenv(Y,X,alpha)

```
B=100;
bootse=bstrp_xenv(X,Y,B,u)
```

45 lrt_xenv

Use likelihood ratio testing to select the dimension of the envelope subspace for the reduction on X.

Contents

- Usage
- Description
- Example

Usage

u=lrt_xenv(X,Y,alpha)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables.
- alpha: Significance level for testing. A real number between 0 and 1, often taken at 0.05 or 0.01.

Output

- u: Dimension of the envelope. An integer between 0 and p.

Description

This function implements the likelihood ratio testing procedure to select the dimension of the envelope subspace for the reduction on X, with pre-specified significance level α.

Example

```
load wheatprotein.txt
X=wheatprotein(:,1:6);
Y=wheatprotein(:,7);
u=lrt_xenv(X,Y,0.01)
```

46 xenv

Fit the envelope model for the reduction on X.

Contents

- Usage
- Description
- References
- Example

Usage

stat=xenv(X,Y,u)

Input

- X: Predictors. An n by p matrix, p is the number of predictors. The predictors can be univariate or multivariate, discrete or continuous.
- Y: Multivariate responses. An n by r matrix, r is the number of responses and n is number of observations. The responses must be continuous variables, and r should be strictly greater than p.
- u: Dimension of the envelope. An integer between 0 and p.

Output

stat: A list that contains the maximum likelihood estimators and some statistics.

- stat.beta: The envelope estimator of the regression coefficients β. An p by r matrix.
- stat.SigX: The envelope estimator of the covariance matrix of X, Σ_X. A p by p matrix.
- stat.Gamma: The orthogonal basis of the envelope subspace. An p by u semi-orthogonal matrix.
- stat.Gamma0: The orthogonal basis of the complement of the envelope subspace. An p by p-u semi-orthogonal matrix.
- stat.eta: The coordinates of β with respect to Gamma. An u by r matrix.
- stat.Omega: The coordinates of Σ_X with respect to Gamma. An u by u matrix.
- stat.Omega0: The coordinates of Σ_X with respect to Gamma0. An p-u by p-u matrix.
- stat.mu: The estimated intercept. An r by 1 vector.
- stat.sigYcX: The estimated conditional covariance matrix of Y given X. An r by r matrix.
- stat.l: The maximized log likelihood function. A real number.

- stat.asyEnv: Asymptotic standard error for elements in β under the envelope model. An r by p matrix. The standard errors returned are asymptotic, for actual standard errors, multiply by 1/sqrt(n).
- stat.ratio: The asymptotic standard error ratio of the standard multivariate linear regression estimator over the envelope estimator, for each element in β. An p by r matrix.
- stat.np: The number of parameters in the envelope model. A positive integer.

Description

This function fits the envelope model to the responses and predictors, using the maximum likehood estimation. When the dimension of the envelope is between 1 and r-1, we implemented the algorithm in Cook et al. (2010). When the dimension is r, then the envelope model degenerates to the standard multivariate linear regression. When the dimension is 0, it means that X and Y are uncorrelated, and the fitting is different.

References

- The codes is implemented based on the algorithm in Section 4.5.1 of Cook et al (2012).
- The Grassmann manifold optimization step calls the package sg_min 2.4.1 by Ross Lippert (

  `http://web.mit.edu/~ripper/www.sgmin.html)`

  .

Example

load wheatprotein.txt
X=wheatprotein(:,1:6);
Y=wheatprotein(:,7);
stat=xenv(X,Y,0);

p=size(X,2);
stat=xenv(X,Y,p);

% When u=p, the envelope model reduces to the ordinary least squares % regression

temp=fit_OLS(X,Y);
temp.SigmaOLS

```
stat.sigYcX
temp.betaOLS'
stat.beta
```

```
stat=xenv(X,Y,5);
```

% To compare with the results obtained by Partial Least Squares, use the command

```
[XL,YL,XS,YS,BETA,PCTVAR,MSE,stats] = plsregress(X,Y,5);
stat.beta
stat.mu
BETA
```