

Package ‘LassoGroupProject’

February 28, 2018

Title Replicate results of the project by the Lasso-Group

Version 1.0

Author Pablo Rodriguez Mira [aut, cre], Nicolas Zimmer [aut], Chong Miao [aut]

Maintainer Pablo Rodriguez Mira <rodriguez.mira.pablo@gmail.com>

Description This package was created to replicate the results of the project of the Lasso-group in a straightforward manner.

Depends R (>= 3.4.3), ggplot2

Imports glmnet, lars, tictoc, MASS, reshape, stargazer

License MIT

Encoding UTF-8

LazyData True

RoxygenNote 6.0.1.9000

R topics documented:

| | |
|--------------------------|-----------|
| ABias_fun | 2 |
| AdalassoCV | 2 |
| AMSPE_fun | 3 |
| AVar_fun | 3 |
| ModelSelSim | 4 |
| PredContest | 5 |
| SP_Application | 6 |
| sp_data | 6 |
| Index | 10 |

ABias_fun

Compute the Squared Average Bias of an estimator

Description

Computes the Squared Average Bias (Squared ABias) of an estimator in a Monte Carlo simulation study.

Usage

```
ABias_fun(modelFit, truth)
```

Arguments

| | |
|----------|--|
| modelFit | A matrix of fitting values of dimensions (number of observations x number of simulations). |
| truth | The true values of the data generating process, i.e., $x_i^T \beta$. |

Value

A numeric value.

AdalassoCV

Adaptive Lasso with automatic γ selection by cross-validation

Description

Computes the best (λ, γ) pair for the adaptive Lasso by cross-validation.

Usage

```
AdalassoCV(X, y, gammaSeq = c(0.5, 1, 2, 5), standardize = FALSE,
  lambdaOption = "min", initEst = "OLS")
```

Arguments

| | |
|--------------|--|
| X | A matrix of regressors. |
| y | The response variable. |
| gammaSeq | A sequence of gamma values to cross-validate. |
| standardize | Should the regressors be standardized? |
| lambdaOption | Should the lambda be the minimizer of the cross-validated error curve (min) or should the lambda be chosen such that its corresponding cross-validated error is only 1-standard-deviation above the minimum error (1se)? |
| initEst | Initial estimator for the adaptive Lasso. Either OLS (OLS) or ridge regression (ridge). |

Value

The tuple (λ, γ) of minimizers of the cross-validated error curve for the adaptive Lasso together with the estimated model.

| | |
|-----------|--|
| AMSPE_fun | <i>Compute the Average Mean Squared Prediction Error (AMSPE) of an estimator</i> |
|-----------|--|

Description

Computes the Average Mean Squared Prediction Error (AMSPE) of an estimator in a Monte Carlo simulation study.

Usage

```
AMSPE_fun(modelFit, truth)
```

Arguments

| | |
|----------|--|
| modelFit | A matrix of fitting values of dimensions (number of observations x number of simulations). |
| truth | The true values of the data generating process, i.e., $x_i^T \beta$. |

Value

A numeric value

| | |
|----------|---|
| AVar_fun | <i>Compute the Average Variance of an estimator</i> |
|----------|---|

Description

Computes the Average Variance (AV) of an estimator in a Monte Carlo simulations study.

Usage

```
AVar_fun(modelFit)
```

Arguments

| | |
|----------|---|
| modelFit | A matrix of fitting values of dimensions (number of observations x number of simulations) |
|----------|---|

Value

A numeric value

ModelSelSim*Replicate the results of the model selection consistency study*

Description

Replicate the results of the Monte Carlo Simulation for the model selection consistency study. The default values are set such that the results you obtain are the same as in our term paper.

Usage

```
ModelSelSim(path = getwd(), scenario = c("big", "small"), nObs = 1000,  
  p = 50, nSim = 100, nDesigns = 100, sigma = 0.1, adaGamma = 2,  
  seed = 481)
```

Arguments

| | |
|-----------------------|---|
| <code>path</code> | The path to which you want the results to be exported to. |
| <code>scenario</code> | Big or small coefficients scenarios to be simulated. |
| <code>nObs</code> | Number of observations. |
| <code>p</code> | Number of variables. |
| <code>nSim</code> | Number of simulations. |
| <code>nDesigns</code> | Number of design matrices. |
| <code>sigma</code> | Standard deviation of the noise for the regression model. |
| <code>adaGamma</code> | Gamma parameter for the adaptive Lasso. Default is 2. |
| <code>seed</code> | Seed for the simulations. |

Value

A (Two) pdf-file(s) exported to your path. Moreover, the results are also saved in a list if you assign a variable to the function as in the example below.

Examples

```
results <- ModelSelSim() # It takes around 1 hour with default arguments
```

Description

Replicate the results of the Monte Carlo Simulation for the prediction contest. The default values are set such that the results you obtain are the same as in our term paper.

Usage

```
PredContest(path = getwd(), pList = c(50, 100), sigmaList = c(1, 5, 25),  
  nObs = 500, nSim = 100, adaList = c(0.01, 0.1, 0.5, 1, 2, 5, 7),  
  seed = 823)
```

Arguments

| | |
|-----------|--|
| path | The path to which you want the results to be exported to. |
| pList | Vector with the choices of number of variables. |
| sigmaList | Vector with the choices of noise levels. |
| nObs | Number of observations. |
| nSim | Number of simulations. |
| adaList | Sequence of gamma parameters for the adaptive Lasso to cross-validate. |
| seed | Seed for the simulations. |

Details

This function creates a new directory in the directory specified in the 'path' argument named 'Prediction_Contest'. This directory in turn include two subdirectories, 'Plots' and 'Tables'. Finally, in the 'Tables' directory are included two directories more, 'HTML' for the result tables in html-format and 'LaTeX' for the result tables in LaTeX format.

Value

(Number of variable choices x Number of noise levels) plots and tables in LaTeX and html. Moreover, the results are also saved in a list if you assign a variable to the function as in the example below. The results are also exported to the directory specified in the 'path' argument.

Examples

```
results <- PredContest() # It takes long time with default arguments (around 5 hours)
```

| | |
|----------------|--|
| SP_Application | <i>Replicate the results of the stock prices application</i> |
|----------------|--|

Description

Replicate the results of the real data application with the stock price dataset.

Usage

```
SP_Application(path = getwd(), adaList = c(0.01, 0.1, 0.5, 1, 2, 3),
  mccv = TRUE, cvIter = 100, seed = 123)
```

Arguments

| | |
|---------|--|
| path | The path to which you want the results to be exported to. |
| adaList | Sequence of gamma parameters for the adaptive Lasso to cross-validate. |
| mccv | Should Monte Carlo Cross Validation be applied for evaluation? |
| cvIter | Number of iterations for the Monte Carlo Cross Validation. |
| seed | Seed for the simulations. |

Details

This function creates a new directory, 'Application_SP' where the results will be saved.

Value

Result tables and plots.

Examples

```
results <- SP_Application()
```

| | |
|---------|---|
| sp_data | <i>Dataset for the application part: Prices and fundamentals.</i> |
|---------|---|

Description

A dataset containing the prices and other attributes of stocks from 231 companies. Prices were fetched from Yahoo Finance, fundamentals are from Nasdaq Financials, extended by some fields from EDGAR SEC databases.

Format

A data frame with 231 rows and 76 variables:

Accounts.Payable Accounts Payable, in US dollars

Accounts.Receivable Accounts Receivable, in US dollars

Add.l.income.expense.items Add'l income or expense items, in US dollars

After.Tax.ROE Return on Equity = Net Income/Shareholder's Equity, in ratio

avg_price Log of the average price from one month before until one month after the financial statements date, in log value of the original US dollars price

Capital.Expenditures Capital Expenditures, in US dollars

Capital.Surplus Capital Surplus, in US dollars

Cash.Ratio The ratio of a company's total cash and cash equivalents to its current liabilities, in ratio

Cash.and.Cash.Equivalents Cash and Cash Equivalents, in US dollars

Changes.in.Inventories Changes in Inventories, in US dollars

Common.Stocks Common Stocks shares, in shares

Cost.of.Revenue Cost of Revenue, in US dollars

Current.Ratio Current assets divided by current liabilities, in ratio

Deferred.Asset.Charges A deferred charge is an expenditure that is paid for in one accounting period, but for which the underlying asset will not be entirely consumed until one or more future periods have been completed, in US dollars

Deferred.Liability.Charges Deferred Liability Charges, in US dollars

Depreciation Depreciation, in US dollars

Earnings.Before.Interest.and.Tax Earnings Before Interest and Tax, in US dollars

Earnings.Before.Tax Earnings Before Tax, in US dollars

Earnings.Per.Share The portion of a company's profit allocated to each outstanding share of common stock, in US dollars

Estimated.Shares.Outstanding The weighted average of outstanding shares is a calculation that incorporates any changes in the amount of outstanding shares over a reporting period, in US dollars

Effect.of.Exchange.Rate Effect of Exchange Rate, in US dollars

Equity.Earnings.Loss.Unconsolidated.Subsidiary Equity Earnings Loss Unconsolidated Subsidiary, in US dollars

Fixed.Assets Fixed Assets, in US dollars

Goodwill Goodwill, in US dollars

Gross.margin The difference between revenue and cost of goods sold (COGS) divided by revenue, in ratio

Gross.Profit Gross Profit, in US dollars

Income.Tax Income Tax, in US dollars

Intangible.Assets Intangible Assets, in US dollars

Interest.Expense Interest Expense, in US dollars

Inventory The raw materials, work-in-process products and finished goods that are considered to be the portion of a business's assets that are ready or will be ready for sale, in US dollars

Investments Investments, in US dollars

Liabilities Liabilities, in US dollars

Long.Term.Debt Long Term Debt, in US dollars

Long-term.investments The name of an asset account that includes all investments that are not expected to be liquidated within the next twelve months, in US dollars

Minority.Interest Minority Interest, in US dollars

MiscStocks Misc Stocks Options Warrants, in US dollars

Net.Borrowings The difference between the amount that a company has borrowed and the amount of cash that it has, in US dollars

Net.Cash.Flow Net Cash Flow, in US dollars

Net.Cash.Flow-Operating Net Cash Flow Operating, in US dollars

Net.Cash.Flows-Financing Net Cash Flows Financing , in US dollars

Net.Cash.Flows-Investing Net Cash Flows Investing , in US dollars

Net.Income Net Income, in US dollars

Net.Income.Adjustments Net Income Adjustments, in US dollars

Net.Income.Applicable.to.Common.Shareholders Net Income Applicable to Common Shareholders, in US dollars

Net.Income-Cont.Operations Net Income ContOperations, in US dollars

Net.Receivables Net Receivables, in US dollars

Non.Recurring.Items Non Recurring Items, in US dollars

Operating.Income Operating Income, in US dollars

Operating.margin Measurement of what proportion of a company's revenue is left over after paying for variable costs of production such as wages, raw materials, etc, in ratio

Other.Assets Other Assets, in US dollars

Other.Current.Assets Other Current Assets, in US dollars

Other.current.liabilities Line item in the balance sheet, which aggregates several current liability accounts that are too minor to report separately, in US dollars

Other.Equity Other Equity, in US dollars

Other.Financing.Activities Other Financing Activities, in US dollars

Other.Investing.Activities Other Investing Activities, in US dollars

Other.Liabilities Other Liabilities, in US dollars

Other.Operating.Activities Other Operating Activities, in US dollars

Other.Operating.Items Other Operating Items, in US dollars

Pre.Tax.Margin Pre Tax Margin, in ratio

Pre.Tax.ROE Pre Tax ROE, in US dollars

Profit.Margin Profit Margin, in ratio

Quick.Ratio Quick Ratio, ratio

Research.and.Development Research and Development, in US dollars

Retained.earnings The percentage of net earnings not paid out as dividends, but retained by the company to be reinvested in its core business, or to pay debt, in US dollars

Sale.and.Purchase.of.Stock Sale and Purchase of Stock, in US dollars

Sales.General.and.Admin Sales, general and administrative expense, in US dollars

Short.Term.Debt...Current.Portion.of.Long.Term.Debt Short Term Debt/Current Portion of Long Term Debt, in ratio

Short.Term.Investments Short Term Investments, in US dollars

Total.Assets Total Assets, in US dollars

Total.Current.Assets Total Current Assets, in US dollars

Total.Current.Liabilities Total Current Liabilities, in US dollars

Total.Equity Total Equity, in US dollars

Total.Liabilities Total Liabilities, in US dollars

Total.Liabilities.Equity Total Liabilities Equity, in US dollars

Total.Revenue Total Revenue, in US dollars

Treasury.stocks Shares that were once a part of the float and shares outstanding but were subsequently repurchased by the company and decommissioned. These stocks do not have voting rights and do not pay any distributions, in US dollars

Source

<https://www.kaggle.com/dgawlik/nyse/data>

Index

*Topic **LassoGroupProject**

- ABias_fun, [2](#)
- AdalassoCV, [2](#)
- AMSPE_fun, [3](#)
- AVar_fun, [3](#)
- ModelSelSim, [4](#)
- PredContest, [5](#)
- SP_Application, [6](#)

- ABias_fun, [2](#)
- AdalassoCV, [2](#)
- AMSPE_fun, [3](#)
- AVar_fun, [3](#)

- ModelSelSim, [4](#)

- PredContest, [5](#)

- SP_Application, [6](#)
- sp_data, [6](#)