Package ‘SSOSVM’

May 6, 2019

Type Package
Title Stream Suitable Online Support Vector Machines
Version 0.2.1
Date 2019-05-06
Author Andrew Thomas Jones, Hien Duy Nguyen, Geoffrey J. McLachlan
Maintainer Andrew Thomas Jones <andrewthomasjones@gmail.com>
Description Soft-margin support vector machines (SVMs) are a common class of classification models. The training of SVMs usually requires that the data be available all at once in a single batch, however the Stochastic majorization-minimization (SMM) algorithm framework allows for the training of SVMs on streamed data instead Nguyen, Jones & McLachlan(2018)<doi:10.1007/s42081-018-0001-y>. This package utilizes the SMM framework to provide functions for training SVMs with hinge loss, squared-hinge loss, and logistic loss.
License GPL-3
Encoding UTF-8
Imports Rcpp (>= 0.12.13), mvtnorm, MASS
LinkingTo Rcpp, RcppArmadillo
RoxygenNote 6.1.1
Suggests testthat, knitr, rmarkdown, ggplot2, gganimate, gifski
NeedsCompilation yes
Repository CRAN
Date/Publication 2019-05-06 09:10:03 UTC

R topics documented:

- generateSim .................................................. 2
- Hinge .......................................................... 2
- Logistic ....................................................... 3
- SquareHinge .................................................. 4
- SSOSVM ....................................................... 5
- SVMFit ....................................................... 5

Index 7
generateSim  

Generate Simulations

Description

Generate simple simulations for testing of the algorithms.

Usage

generateSim(NN = 10^4, DELTA = 2, DIM = 2, seed = NULL)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>Number of observations. Default is 10^4</td>
</tr>
<tr>
<td>DELTA</td>
<td>Separation of three groups in standard errors. Default is 2.</td>
</tr>
<tr>
<td>DIM</td>
<td>Number of dimensions in data. Default is 2.</td>
</tr>
<tr>
<td>seed</td>
<td>Random seed if desired.</td>
</tr>
</tbody>
</table>

Value

A list containing:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>Coordinates of the simulated points.</td>
</tr>
<tr>
<td>YY</td>
<td>Cluster membership of the simulated points.</td>
</tr>
<tr>
<td>YMAT</td>
<td>YY and XX Combined as a single matrix.</td>
</tr>
</tbody>
</table>

Examples

#100 points of dimension 4.
generateSim(NN=100, DELTA=2, DIM=4)

Hinge

Hinge

Description

Fit SVM with Hinge loss function.

Usage

Hinge(YMAT, DIM = 2L, EPSILON = 1e-05, returnAll = FALSE, rho = 1)
**Logistic**

**Arguments**

- **YMAT** Data. First column is -1 or 1 indicating the class of each observation. The remaining columns are the coordinates of the data points.
- **DIM** Dimension of data. Default value is 2.
- **EPSILON** Small perturbation value needed in calculation. Default value is 0.00001.
- **returnAll** Return all of theta values? Boolean with default value FALSE.
- **rho** Sensitivity factor to adjust the level of change in the SVM fit when a new observation is added. Default value 1.0

**Value**

A list containing:

- **THETA** SVM fit parameters.
- **NN** Number of observation points in YMAT.
- **DIM** Dimension of data.
- **THETA_list** THETA at each iteration (new point observed) as YMAT is fed into the algorithm one data point at a time.
- **OMEGA** Intermediate value OMEGA at each iteration (new point observed).

**Examples**

```r
Ymat <- generateSim(10^4)
h1 <- Hinge(YMAT$YMAT, returnAll=TRUE)

Logistic < Logistic(YMAT, DIM = 2L, EPSILON = 1e-05, returnAll = FALSE, rho = 1)
```

**Description**

Fit SVM with Logistic loss function.

**Usage**

```r
Logistic(YMAT, DIM = 2L, EPSILON = 1e-05, returnAll = FALSE, rho = 1)
```

**Arguments**

- **YMAT** Data. First column is -1 or 1 indicating the class of each observation. The remaining columns are the coordinates of the data points.
- **DIM** Dimension of data. Default value is 2.
- **EPSILON** Small perturbation value needed in calculation. Default value is 0.00001.
- **returnAll** Return all of theta values? Boolean with default value FALSE.
- **rho** Sensitivity factor to adjust the level of change in the SVM fit when a new observation is added. Default value 1.0
Value

A list containing:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>THETA</td>
<td>SVM fit parameters.</td>
</tr>
<tr>
<td>NN</td>
<td>Number of observation points in YMAT.</td>
</tr>
<tr>
<td>DIM</td>
<td>Dimension of data.</td>
</tr>
<tr>
<td>THETA_list</td>
<td>THETA at each iteration (new point observed) as YMAT is fed into the algorithm one data point at a time.</td>
</tr>
<tr>
<td>CHI</td>
<td>Intermediate value CHI at each iteration (new point observed).</td>
</tr>
</tbody>
</table>

Examples

```r
YMAT <- generateSim(1e^4)
l1<-logistic(YMAT$YMAT,returnAll=TRUE)
```

Description

Fit SVM with Square Hinge loss function.

Usage

```r
SquareHinge(YMAT, DIM = 2L, EPSILON = 1e-05, returnAll = FALSE, rho = 1)
```

Arguments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YMAT</td>
<td>Data. First column is -1 or 1 indicating the class of each observation. The remaining columns are the coordinates of the data points.</td>
</tr>
<tr>
<td>DIM</td>
<td>Dimension of data. Default value is 2.</td>
</tr>
<tr>
<td>EPSILON</td>
<td>Small perturbation value needed in calculation. Default value is 0.00001.</td>
</tr>
<tr>
<td>returnAll</td>
<td>Return all of theta values? Boolean with default value FALSE.</td>
</tr>
<tr>
<td>rho</td>
<td>Sensitivity factor to adjust the level of change in the SVM fit when a new observation is added. Default value 1.0</td>
</tr>
</tbody>
</table>

Value

A list containing:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>THETA</td>
<td>SVM fit parameters.</td>
</tr>
<tr>
<td>NN</td>
<td>Number of observation points in YMAT.</td>
</tr>
<tr>
<td>DIM</td>
<td>Dimension of data.</td>
</tr>
<tr>
<td>THETA_list</td>
<td>THETA at each iteration (new point observed) as YMAT is fed into the algorithm one data point at a time.</td>
</tr>
<tr>
<td>PSI</td>
<td>Intermediate value PSI at each iteration (new point observed).</td>
</tr>
</tbody>
</table>
SSOSVM

Examples

\[ \text{YMAT} \leftarrow \text{generateSim}(10^3, \text{DIM}=3) \]
\[ \text{sql} \leftarrow \text{SquareHinge}(\text{YMAT} \_ \text{YMAT}, \text{DIM}=3, \text{returnAll=TRUE}) \]

SSOSVM

**SSOSVM**: A package for online training of soft-margin support vector machines (SVMs) using the Stochastic majorization–minimization (SMM) algorithm.

Description

The SSOSVM package allows for the online training of Soft-margin support vector machines (SVMs) using the Stochastic majorization–minimization (SMM) algorithm. \text{SquareHinge}, \text{Hinge} and \text{Logistic} The function \text{generateSim} can also be used to generate simple test sets.

Author(s)

Andrew T. Jones, Hien D. Nguyen, Geoffrey J. McLachlan

References


SVMFit

**SSOSVM Fit function**

Description

This is the primary function for uses to fit SVMs using this package.

Usage

\[ \text{SVMFit} (\text{YMAT}, \text{method} = \text{"logistic"}, \text{EPSILON} = 1e-05, \text{returnAll} = \text{FALSE}, \text{rho} = 1) \]

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{YMAT}</td>
<td>Data. First column is -1 or 1 indicating the class of each observation. The remaining columns are the coordinates of the data points.</td>
</tr>
<tr>
<td>\text{method}</td>
<td>Choice of function used in SVM. Choices are 'logistic', 'hinge' and 'square-Hinge'. Default value is 'logistic'</td>
</tr>
<tr>
<td>\text{EPSILON}</td>
<td>Small perturbation value needed in calculation. Default value is 0.00001.</td>
</tr>
<tr>
<td>\text{returnAll}</td>
<td>Return all of theta values? Boolean with default value FALSE.</td>
</tr>
<tr>
<td>\text{rho}</td>
<td>Sensitivity factor to adjust the level of change in the SVM fit when a new observation is added. Default value 1.0</td>
</tr>
</tbody>
</table>
Value

A list containing:

- **THETA** SVM fit parameters.
- **NN** Number of observation points in YMAT.
- **DIM** Dimension of data.
- **THETA_list** THETA at each iteration (new point observed) as YMAT is fed into the algorithm one data point at a time.
- **PSI, OMEGA, CHI** Intermediate value for PSI, OMEGA, or CHI (depending on method choice) at each iteration (new point observed).

Examples

```r
Sim <- generateSim(10^4)
m1 <- SVMFit(Sim$YMAT)
```
Index

generateSim, 2
Hinge, 2
Logistic, 3
SquareHinge, 4
SSOSVM, 5
SSOSVM-package (SSOSVM), 5
SVMFit, 5