Package ‘evtclass’

November 16, 2018

Title Extreme Value Theory for Open Set Classification - GPD and GEV Classifiers

Version 1.0

Description Two classifiers for open set recognition and novelty detection based on extreme value theory. The first classifier is based on the generalized Pareto distribution (GPD) and the second classifier is based on the generalized extreme value (GEV) distribution. For details, see Vignotto, E., & Engelke, S. (2018) <arXiv:1808.09902>.

Depends R (>= 3.4.0)
License GPL-3
Encoding UTF-8
LazyData true
RoxygenNote 6.1.0.9000
Imports RANN, evd, fitdistrplus
NeedsCompilation no
Author Edoardo Vignotto [aut, cre] (<https://orcid.org/0000-0001-9870-8020>)
Maintainer Edoardo Vignotto <edoardo.vignotto@unige.ch>
Repository CRAN
Date/Publication 2018-11-16 16:40:11 UTC

R topics documented:

gevcTest ................................................................. 2
gevcTrain ............................................................... 3
gpdcTest ............................................................... 4
gpdcTrain ............................................................... 5
LETTER ................................................................. 6

Index 8
gevcTest

Description

This function is used to evaluate a test set for a pre-trained GEV classifier. It can be used to perform open set classification based on the generalized Pareto distribution.

Usage

gevcTest(train, test, pre, prob = TRUE, alpha)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td>a data matrix containing the train data. Class labels should not be included.</td>
</tr>
<tr>
<td>test</td>
<td>a data matrix containing the test data.</td>
</tr>
<tr>
<td>pre</td>
<td>a numeric vector of parameters obtained with the function gevTrain.</td>
</tr>
<tr>
<td>prob</td>
<td>logical indicating whether p-values should be returned.</td>
</tr>
<tr>
<td>alpha</td>
<td>threshold to be used if prob is equal to FALSE. It must be between 0 and 1.</td>
</tr>
</tbody>
</table>

Details

For details on the method and parameters see Vignotto and Engelke (2018).

Value

If prob is equal to TRUE, a vector containing the p-values for each point is returned. A high p-value results in the classification of the corresponding test data as a known point, since this hypothesis cannot be rejected. If the p-value is small, the corresponding test data is classified as an unknown point. If prob is equal to TRUE, a vector of predicted values is returned.

Author(s)

Edoardo Vignotto
<edoardo.vignotto@unige.ch>

References


See Also

gevTrain
**gevcTrain**

**Examples**

```r
trainset <- LETTER[1:15000,]
testset <- LETTER[-(1:15000), -1]
knowns <- trainset[trainset$class == 1, -1]
gevClassifier <- gevctrain(train = knowns)
predicted <- gevctest(train = knowns, test = testset, pre = gevClassifier)
```

---

**Description**

This function is used to train a GEV classifier. It can be used to perform open set classification based on the generalized extreme value distribution.

**Usage**

```r
gevcTrain(train)
```

**Arguments**

- `train` a data matrix containing the train data. Class labels should not be included.

**Details**

For details on the method and parameters see Vignotto and Engelke (2018).

**Value**

A numeric vector of two elements containing the estimated parameters of the fitted reversed Weibull.

**Note**

Data are not scaled internally; any preprocessing has to be done externally.

**Author(s)**

Edoardo Vignotto
<edoardo.vignotto@unige.ch>

**References**


**See Also**

`gevcTest`
Examples

```r
trainset <- LETTER[1:15000,]
knowns <- trainset[trainset$class==1, -1]
gevClassifier <- gevcTrain(train = knowns)
gpdctest
```

**Description**

This function is used to evaluate a test set for a pre-trained GPD classifier. It can be used to perform open set classification based on the generalized Pareto distribution.

**Usage**

```r
gpdctest(train, test, pre, prob = TRUE, alpha = 0.01)
```

**Arguments**

- `train`: data matrix containing the train data. Class labels should not be included.
- `test`: a data matrix containing the test data.
- `pre`: a list obtained with the function `gevcTrain`.
- `prob`: logical indicating whether p-values should be returned.
- `alpha`: threshold to be used if `prob` is equal to FALSE. It must be between 0 and 1.

**Details**

For details on the method and parameters see Vignotto and Engelke (2018).

**Value**

If `prob` is equal to TRUE, a vector containing the p-values for each point is returned. A high p-value results in the classification of the corresponding test data as a known point, since this hypothesis cannot be rejected. If the p-value is small, the corresponding test data is classified as an unknown point. If `prob` is equal to FALSE, a vector of predicted values is returned.

**Author(s)**

Edoardo Vignotto
<edoardo.vignotto@unige.ch>

**References**

gpdcTrain

See Also

   gpdcTrain

Examples

   trainset <- LETTER[1:15000,]
   testset <- LETTER[-(1:15000), -1]
   knowns <- trainset[trainset$class==1, -1]
   gpdClassifier <- gpdcTrain(train = knowns, k = 10)
   predicted <- gpdcTest(train = knowns, test = testset, pre = gpdClassifier)

---

Description

   This function is used to train a GPD classifier. It can be used to perform open set classification
   based on the generalized Pareto distribution.

Usage

   gpdcTrain(train, k)

Arguments

   train a data matrix containing the train data. Class labels should not be included.
   k the number of upper order statistics to be used.

Details

   For details on the method and parameters see Vignotto and Engelke (2018).

Value

   A list of three elements.

   pshapes the estimated rescaled shape parameters for each point in the training dataset.
   balls the estimated radius for each point in the training dataset.
   k the number of upper order statistics used.

Note

   Data are not scaled internally; any preprocessing has to be done externally.

Author(s)

   Edoardo Vignotto
   <edoardo.vignotto@unige.ch>
References


See Also

gpdcTest

Examples

```r
trainset <- LETTER[1:15000,]
knowns <- trainset[trainset$class==1, -1]
gpdcClassifier <- gpdcTrain(train = knowns, k = 10)
```

Description

A dataset containing 16 features extracted from 20000 handwritten characters.

Usage

LETTER

Format

A data frame with 20000 rows and 17 variables:

- **class**: class labels
- **V1**: first extracted feature
- **V2**: second extracted feature
- **V3**: third extracted feature
- **V4**: 4th extracted feature
- **V5**: 5th extracted feature
- **V6**: 6th extracted feature
- **V7**: 7th extracted feature
- **V8**: 8th extracted feature
- **V9**: 9th extracted feature
- **V10**: 10th extracted feature
- **V11**: 11th extracted feature
- **V12**: 12th extracted feature
- **V13**: 13th extracted feature
- **V14**: 14th extracted feature
- **V15**: 15th extracted feature
- **V16**: 16th extracted feature
LETTER

Source

Index

*Topic datasets
  LETTER, 6

gevcTest, 2, 3
gevcTrain, 2, 3
gpdcTest, 4, 6
gpdcTrain, 4, 5, 5

LETTER, 6