

Package: mlf (via r-universe)

September 13, 2024

Type Package

Title Machine Learning Foundations

Version 1.2.1

Date 2018-06-21

Maintainer Kyle Peterson <petersonkdon@gmail.com>

Description Offers a gentle introduction to machine learning concepts for practitioners with a statistical pedigree: decomposition of model error (bias-variance trade-off), nonlinear correlations, information theory and functional permutation/bootstrap simulations. Székely GJ, Rizzo ML, Bakirov NK. (2007). <doi:10.1214/009053607000000505>. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. (2011). <doi:10.1126/science.1205438>.

Imports stats, utils

URL <http://mlf-project.us/>

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Kyle Peterson [aut, cre]

Date/Publication 2018-06-25 08:01:20 UTC

Repository <https://kdpeterson51.r-universe.dev>

RemoteUrl <https://github.com/cran/mlf>

RemoteRef HEAD

RemoteSha 7bd3fc08bf7d3279a89d442e2697ae4b912c810b

Contents

boot	2
bvto	3
distcorr	3
entropy	4
get_bias	4
get_mse	5
get_var	6
jointentropy	6
kld	7
mi	8
mic	8
perm	9
Index	10

boot	<i>Bootstrap Confidence Intervals via Resampling</i>
------	--

Description

Provides nonparametric confidence intervals via percentile-based resampling for given mlf function.

Usage

```
boot(x, y, func, reps, conf.int)
```

Arguments

x, y	numeric vectors of data values
func	specify mlf function
reps	(optional) number of resamples. Defaults to 500
conf.int	(optional) numeric value indicating level of confidence. Defaults to 0.90.

Examples

```
# Sample data
a <- rnorm(25, 80, 35)
b <- rnorm(25, 100, 50)

mlf::mic(a, b)
mlf::boot(a, b, mic)
```

bvto	<i>Bias-Variance Trade-Off</i>
------	--------------------------------

Description

Provides estimated error decomposition from model predictions (mse, bias, variance).

Usage

```
bvto(truth, estimate)
```

Arguments

truth	test data vector or baseline accuratruth to test against.
estimate	predicted vector

Examples

```
# Sample data
test <- rnorm(25, 80, 35)
predicted <- rnorm(25, 80, 50)

mlf::bvto(test, predicted)
```

distcorr	<i>Distance Correlation</i>
----------	-----------------------------

Description

Provides pairwise correlation via distance covariance normalized by distance standard deviation. Allows for non-linear dependencies.

Usage

```
distcorr(x, y)
```

Arguments

x, y	numeric vectors of data values
------	--------------------------------

References

Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. *Ann Stat.* 2007. 35(6):2769-2794.

Examples

```
# Sample data
a <- rnorm(25, 80, 35)
b <- rnorm(25, 100, 50)

mlf::distcorr(a, b)
```

entropy

Entropy

Description

Estimates uncertainty in univariate probability distribution.

Usage

```
entropy(x, bins)
```

Arguments

x numeric or discrete data vector
bins specify number of bins if numeric or integer data class.

Examples

```
# Sample numeric vector
a <- rnorm(25, 80, 35)
mlf::entropy(a, bins = 2)

# Sample discrete vector
b <- as.factor(c(1,1,1,2))
mlf::entropy(b)
```

get_bias

Bias

Description

Estimates squared bias by decomposing model prediction error.

Usage

```
get_bias(truth, estimate)
```

Arguments

truth test data vector or baseline accuracy to test against.
estimate predicted vector

Examples

```
# Sample data  
test <- rnorm(25, 80, 35)  
predicted <- rnorm(25, 80, 50)  
  
mlf::get_bias(test, predicted)
```

get_mse *Mean Squared Error*

Description

Estimates mean squared error from model predictions.

Usage

```
get_mse(truth, estimate)
```

Arguments

truth test data vector or baseline accuracy to test against.
estimate predicted vector

Examples

```
# Sample data  
test <- rnorm(25, 80, 35)  
predicted <- rnorm(25, 80, 50)  
  
mlf::get_mse(test, predicted)
```

get_var *Variance*

Description

Estimates squared variance by decomposing model prediction error.

Usage

```
get_var(estimate)
```

Arguments

estimate predicted vector

Examples

```
# Sample data
test <- rnorm(25, 80, 35)
predicted <- rnorm(25, 80, 50)

mlf::get_var(predicted)
```

jointentropy *Joint Entropy*

Description

Estimated difference between two probability distributions.

Usage

```
jointentropy(x, y, bins)
```

Arguments

x, y numeric or discrete data vectors
bins specify number of bins

Examples

```
# Sample numeric vector
a <- rnorm(25, 80, 35)
b <- rnorm(25, 90, 35)
mlf::jointentropy(a, b, bins = 2)

# Sample discrete vector
a <- as.factor(c(1,1,2,2))
b <- as.factor(c(1,1,1,2))
mlf::jointentropy(a, b)
```

kld

Kullback-Leibler Divergence

Description

Provides estimated difference between individual entropy and cross-entropy of two probability distributions.

Usage

```
kld(x, y, bins)
```

Arguments

x, y	numeric or discrete data vectors
bins	specify number of bins

Examples

```
# Sample numeric vector
a <- rnorm(25, 80, 35)
b <- rnorm(25, 90, 35)
mlf::kld(a, b, bins = 2)

# Sample discrete vector
a <- as.factor(c(1,1,2,2))
b <- as.factor(c(1,1,1,2))
mlf::kld(a, b)
```

mi *Mutual Information*

Description

Estimates Kullback-Leibler divergence of joint distribution and the product of two respective marginal distributions. Roughly speaking, the amount of information one variable provides about another.

Usage

```
mi(x, y)
```

Arguments

x, y numeric or discrete data vectors

Examples

```
# Sample data
a <- rnorm(25, 80, 35)
b <- rnorm(25, 100, 50)

mlf::mi(a, b)
```

mic *Maximal Information Criterion*

Description

Information-theoretic approach for detecting non-linear pairwise dependencies. Employs heuristic discretization to achieve highest normalized mutual information.

Usage

```
mic(x, y)
```

Arguments

x, y numeric or discrete data vectors

References

Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. *Science*. 2011. 334(6062):1518-1524.

Examples

```
# Sample data
a <- rnorm(25, 80, 35)
b <- rnorm(25, 100, 50)

mlf::mic(a, b)
```

perm

Permutation Test

Description

Provides nonparametric statistical significance via sample randomization.

Usage

```
perm(x, y, func, reps)
```

Arguments

x, y	numeric vectors of data values
func	specify mlf function: (distcorr or mic).
reps	(optional) number of resamples. Defaults to 500.

Examples

```
# Sample data
a <- rnorm(25, 80, 35)
b <- rnorm(25, 100, 50)

mlf::mic(a, b)
mlf::perm(a, b, mic)
```

Index

boot, 2

bvto, 3

distcorr, 3

entropy, 4

get_bias, 4

get_mse, 5

get_var, 6

jointentropy, 6

kld, 7

mi, 8

mic, 8

perm, 9