# Package 'cgAUC'

October 12, 2022

**Type** Package

**Title** Calculate AUC-type measure when gold standard is continuous and the corresponding optimal linear combination of variables with respect to it.

**Version** 1.2.1

**Date** 2014-08-24

**Author** Yuan-chin I. Chang, Yu-chia Chang, and Ling-wan Chen

**Maintainer** Yu-chia Chang <curare7177@gmail.com>

**Description** The cgAUC can calculate the AUC-type measure of Obuchowski(2006) when gold standard is continuous, and find the optimal linear combination of variables with respect to this measure.

**License** GPL (>= 2)

**Imports** Rcpp (>= 0.11.2)

**LinkingTo** Rcpp

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2014-08-28 07:01:42

## R topics documented:

cgAUC-package *Calculate AUC when gold standard is continuous with large variables.*

## Description

In this package, the cgAUC is used to calculate the AUC-type measure raised in Obuchowski(2006) when gold standard is continuous.

## Details

| | |
|---|---|
| Package: | cgAUC |
| Type: | Package |
| Version: | 1.2.1 |
| Date: | 2014-08-24 |
| License: | GPL (>=2) |

## Author(s)

Yuan-chin I. Chang, Yu-chia Chang, and Ling-wan Chen

Maintainer: Yu-chia Chang <curare7177@gmail.com>

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
# n = 100; p = 5;
# r.x = matrix(rnorm(n * p), , p) # raw data
# r.z = r.x[ ,1] + rnorm(n) # gold standard
# x = scale(r.x) # standardized of raw data
# z = scale(r.z) # standardized of gold standard
# h = n^(-1 / 2)
# t1 = cgAUC(r.x, r.z, h, delta = 1, auto = FALSE, tau = 1, scale = 1) # the delta be constant
# t1
```

```
# t2 = cgAUC(r.x, r.z, h, delta = 1, auto = TRUE, tau = 1, scale = 1) # the delta be variable
# t2
```

---

cgAUC *Calculate AUC when gold standard is continuous with large variables.*

---

## Description

The cgAUC can calculate the AUC-type measure of Obuchowski(2006) when gold standard is continuous, and find the optimal linear combination of variables with respect to this measure.

## Usage

```
cgAUC(x, z, h, delta = 1, auto = FALSE, tau = 1, scale = 1)
```

## Arguments

| | |
|---|---|
| x | The potential variables. It is a matrix with column of values of a variables. It should be standardized in this application. |
| z | The gold standard variable. It should be standardized. |
| h | The parameter controls the window width of smoothing function. |
| delta | The parameter be used in TGDM. The default value is one. |
| auto | Find the optimal delta in TGDN using cross-validation. If the auto is TRUE. The default is FALSE. |
| tau | The parameter used in TGDM. The default value is one. |
| scale | Scaling data when scale = 1, no scaling data when scale = 0. The default value is 1. |

## Details

In this package, we use the TGDM to find the optimal linear combination of variables in order to maximize the AUC-type measure. Before using this function, all of variables, including gold standard variable, should be standardized first. Below are parameters used in the algorithm:

## Value

| | |
|---|---|
| Rev | When Rev = 0 means l * 1; otherwise, l * -1. |
| l | The estimate of coefficients for the optimal linear combination of variables. |
| theta.sh.h.p | The estimate of the theta of Chang(2012) for the optimal linear combination of variables. |
| theta.sh.h.p.var | The estimate of variance for the theta of Chang(2012). |
| cntin.ri | The estimate of the theta of Chang(2012) for each single vaiable. |
| theta.h.p | The estimate of the theta of Obuchowski(2006) for the optimal linear combination of variables. |

| | |
|---|---|
| theta.h.p.var | The estimate of variance for the theta of Obuchowski(2006). |
| dscrt.ri | The estimate of the theta of Obuchowski(2006) for each single vaiable. |
| delta | The value of delta. |

## Author(s)

Yu-chia Chang

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

# n = 100; p = 5;
# r.x = matrix(rnorm(n * p), , p) # raw data
# r.z = r.x[ ,1] + rnorm(n) # gold standard
# x = scale(r.x) # standardized of raw data
# z = scale(r.z) # standardized of gold standard
# h = n^(-1 / 2)
# t1 = cgAUC(r.x, r.z, h, delta = 1, auto = FALSE, tau = 1, scale = 1) # the delta be constant
# t1
# t2 = cgAUC(r.x, r.z, h, delta = 1, auto = TRUE, tau = 1, scale = 1) # the delta be variable
# t2

## The function is currently defined as
function (x, z, h, delta = 1, auto = FALSE, tau = 1)
{
x = scale(x)
z = scale(z)
conv = FALSE
n = dim(x)[1]
p = dim(x)[2]
cntin.ri = dscrt.ri = rep(0, p)
id = diag(p)
for (i in 1:p) {
dscrt.ri[i] = dscrt(x, z, id[i, ])$theta.h.p
cntin.ri[i] = cntin(x, z, id[i, ], h)$theta.sh.h.p
}
```

```
beta.i = ifelse(cntin.ri > 0.5, 1, -1)
dscrt.ri = ifelse(dscrt.ri > 0.5, dscrt.ri, (1 - dscrt.ri))
cntin.ri = ifelse(cntin.ri > 0.5, cntin.ri, (1 - cntin.ri))
y = x * matrix(beta.i, n, p, byrow = TRUE)
max.x = which(cntin.ri == max(cntin.ri))
theta.sh.h.p = 0
l = id[max.x, ]
while (conv == FALSE) {
d.l = d.theta.sh.h.p(y, z, l, h)
max.d.l = max(d.l)
ind.d.l = ifelse(d.l >= (tau * max.d.l), 1, 0) * d.l
if (auto == TRUE) {
delta = optimal.delta(y, z, l, h, ind.d.l)
}
l = l + delta * ind.d.l
l = l/max(l)
theta.temp = cntin(y, z, l, h)$theta.sh.h.p
ifelse(abs(theta.temp - theta.sh.h.p) < 1e-04, conv <- TRUE, conv <- FALSE)
theta.sh.h.p = theta.temp
}
optimal.dscrt = dscrt(y, z, l)
theta.sh.h.p.var = cntin(y, z, l, h)$var
l = l * beta.i
return(list(l = l, theta.sh.h.p = theta.sh.h.p, theta.sh.h.p.var = theta.sh.h.p.var,
cntin.ri = cntin.ri, theta.h.p = optimal.dscrt$theta.h.p,
theta.h.p.var = optimal.dscrt$var, dscrt.ri = dscrt.ri,
delta = delta))
}
## The function is currently defined as
function (x, z, h, delta = 1, auto = FALSE, tau = 1)
{
    x = scale(x)
    z = scale(z)
    conv = FALSE
    n = dim(x)[1]
    p = dim(x)[2]
    cntin.ri = dscrt.ri = rep(0, p)
    id = diag(p)
    for (i in 1:p) {
        dscrt.ri[i] = dscrt(x, z, id[i, ])$theta.h.p
        cntin.ri[i] = cntin(x, z, id[i, ], h)$theta.sh.h.p
    }
    beta.i = ifelse(cntin.ri > 0.5, 1, -1)
    dscrt.ri = ifelse(dscrt.ri > 0.5, dscrt.ri, (1 - dscrt.ri))
    cntin.ri = ifelse(cntin.ri > 0.5, cntin.ri, (1 - cntin.ri))
    y = x * matrix(beta.i, n, p, byrow = TRUE)
    max.x = which(cntin.ri == max(cntin.ri))
    theta.sh.h.p = 0
    l = id[max.x, ]
    while (conv == FALSE) {
        d.l = d.theta.sh.h.p(y, z, l, h)
        max.d.l = max(d.l)
        ind.d.l = ifelse(d.l >= (tau * max.d.l), 1, 0) * d.l
```

```
        if (auto == TRUE) {
            delta = optimal.delta(y, z, l, h, ind.d.l)
        }
        l = l + delta * ind.d.l
        l = l/max(l)
        theta.temp = cntin(y, z, l, h)$theta.sh.h.p
        ifelse(abs(theta.temp - theta.sh.h.p) < 1e-04, conv <- TRUE,
            conv <- FALSE)
        theta.sh.h.p = theta.temp
    }
    optimal.dscrt = dscrt(y, z, l)
    theta.sh.h.p.var = cntin(y, z, l, h)$var
    l = l * beta.i
    return(list(l = l, theta.sh.h.p = theta.sh.h.p, theta.sh.h.p.var = theta.sh.h.p.var,
        cntin.ri = cntin.ri, theta.h.p = optimal.dscrt$theta.h.p,
        theta.h.p.var = optimal.dscrt$var, dscrt.ri = dscrt.ri,
        delta = delta))
}
```

---

c_cntin *c_cntin*

---

## Description

Continue function, when variable was continue.

## Usage

```
c_cntin(y, z, l, h)
```

## Arguments

| | |
|---|---|
| y | The potential variables. It is a matrix with column of values of a variables. It should be standardized in this application. |
| z | The gold standard variable. It should be standardized. |
| l | Linear combination. A vector. |
| h | The value of h falls into (n^(-1/2), n^(-1/5)). |

## Value

| | |
|---|---|
| theta.sh.h.p | The estimate of the theta of Chang(2012). |
| var | The variance of estimate of the theta of Chang(2012). |

## Author(s)

Yu-chia Chang

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function(y, z, l, h) {
    .Call('cgAUC_c_cntin', PACKAGE = 'cgAUC', y, z, l, h)
}
```

---

`c_dscrt` *c_dscrt*

---

## Description

discrete function, when variable is discrete.

## Usage

```
c_dscrt(y, z, l)
```

## Arguments

| | |
|---|---|
| `y` | The potential variables. It is a matrix with column of values of a variables. It should be standardized in this application. |
| `z` | The gold standard variable. It should be standardized. |
| `l` | Linear combination. A vector. |

## Details

Discrete function, when variable is discrete.

## Value

| | |
|---|---|
| `theta.h.p` | The estimate of theta when variable is discrete. |
| `var` | The variance of estimate of theta. |

### Author(s)

Yu-chia Chang

### References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

### Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function(y, z, l) {
    .Call('cgAUC_c_dscrt', PACKAGE = 'cgAUC', y, z, l)
}
```

---

## c_d_theta_sh_h_p *c_d_theta_sh_h_p*

---

### Description

Compute the c_d_theta_sh_h_p.

### Usage

```
c_d_theta_sh_h_p(y, z, l, h)
```

### Arguments

| | |
|---|---|
| y | The potential variables. It is a matrix with column of values of a variables. It should be standardized in this application. |
| z | The gold standard variable. It should be standardized. |
| l | Linear combination. A vector. |
| h | The value of h falls into (n^(-1/2), n^(-1/5)). |

### Details

Compute the c_d_theta_sh_h_p Come from differential.

## Value

`d.theta.sh.h.p` Theta after differential.

## Author(s)

Yu-chia Chang

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function(y, z, l, h) {
    .Call('cgAUC_c_d_theta_sh_h_p', PACKAGE = 'cgAUC', y, z, l, h)
}
```

---

`c_s_h` *c_s_h*

---

## Description

Smooth function.

## Usage

```
c_s_h(t, h)
```

## Arguments

| | |
|---|---|
| `t` | A value, the difference between any two subjects. |
| `h` | The value of h falls into (n^(-1/2), n^(-1/5)). |

## Details

Smooth function.

## Value

| | |
|---|---|
| s_h | The value of smooth function. |

## Author(s)

Yu-chia Chang

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function(t, h) {
    .Call('cgAUC_c_s_h', PACKAGE = 'cgAUC', t, h)
}
```

---

# optimal.delta *optimal.delta*

---

## Description

Find the optimal delta.

## Usage

```
optimal.delta(y, z, l, h, ind.d.l)
```

## Arguments

| | |
|---|---|
| y | The potential variables. It is a matrix with column of values of a variables. It should be standardized in this application. |
| z | The gold standard variable. It should be standardized. |
| l | Linear combination. A vector. |
| h | The value of h falls into (n^(-1/2), n^(-1/5)). |
| ind.d.l | Void |

## Value

`delta.star` Optimal delta.

## Author(s)

Yu-chia Chang

## References

Chang, YCI. Maximizing an ROC type measure via linear combination of markers when the gold reference is continuous. Statistics in Medicine 2012.
Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Statistics in Medicine 2006; 25:481–493.
Obuchowski N. Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. Statistics in Medicine 2005; 20:3261–3278.
Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Technical Report, Department of Statistics, Stanford University, 2004.

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function (y, z, l, h, ind.d.l)
{
    l.i = matrix(rep(l, times = 50), nrow = 50, byrow = TRUE)
    delta = seq(0, 5, length = 50)
    m = delta %*% t(ind.d.l)
    l.i = l.i + m
    l.i.max = apply(l.i, 1, max)
    l.i = l.i/l.i.max
    theta = rep(0, 50)
    for (i in 2:50) {
        theta[i] = cntin(y, z, l.i[i, ], h)$theta.sh.h.p
    }
    delta.star = delta[which(theta == max(theta))]
    return(delta.star)
  }
```

# Index