

# Package ‘maxadjAUC’

March 14, 2018

**Type** Package

**Title** Maximizing the Adjusted AUC

**Version** 0.1.0

**Date** 2018-03-13

**Author** Allison Meisner

**Maintainer** Allison Meisner <allison.meisner@gmail.com>

**Description** Fits a linear combination of predictors by maximizing a smooth approximation to the estimated covariate-adjusted area under the receiver operating characteristic curve (AUC) for a discrete covariate. (Meisner, A, Parikh, CR, and Kerr, KF (2017) <<http://biostats.bepress.com/uwbiostat/paper421/>>.)

**License** GPL-2

**LazyData** TRUE

**Imports** survival, aucm, Rsolnp, Hmisc

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-03-14 17:11:06 UTC

## R topics documented:

maxadjAUC . . . . . 1

**Index** . . . . . 5

---

maxadjAUC

*Maximizing the Covariate-Adjusted AUC*

---

## Description

Often there is interest in combining several predictors or biomarkers into a linear combination for diagnosis, prognosis or screening. This can be done by targeting measures of predictive capacity. In the presence of a discrete covariate, such as batch or study center, an appropriate summary of discriminatory performance is the covariate-adjusted area under the receiver operating characteristic curve (AUC), or aAUC. This function estimates a linear combination of predictors by maximizing a smooth approximation to the aAUC.

## Usage

```
maxadjAUC(outcome, predictors, covariate, initialval="rGLM", approxh = 1/3,
conditional=FALSE, tolval = 1e-6, stepsz = 1e-5)
```

## Arguments

outcome	A vector of outcome (disease) indicators for each observation (1 for diseased, 0 for non-diseased). Missing values are not allowed.
predictors	A numeric matrix with one row for each observation and one column for each candidate predictor. Missing values are not allowed. The columns of the matrix will be (re)named "V1", "V2", ....
covariate	A numeric vector of covariate values for each observation. The covariate should have a limited number of values (i.e., it should be a discrete covariate). Missing values are not allowed.
initialval	Starting values of the predictor combination for the SaAUC algorithm. Default value is "rGLM", which means that estimates from robust logistic regression, specifically the method of Bianco and Yohai (implemented via the <code>aucm</code> package), are used as starting values. If any other value of <code>initialval</code> is given, or if robust logistic regression fails to converge, estimates from standard logistic regression are used as starting values. For both robust and standard logistic regression, the covariate will be included as a stratifying variable.
approxh	The tuning parameter for the smooth approximation to the covariate-specific AUC is the ratio of the standard deviation of the linear combination (based on the starting values) to $n_c^{approx}$ , where $n_c$ is the number of observations with covariate value $c$ . In particular, larger values of <code>approxh</code> will provide a better approximation to the AUC, though estimation may become unstable if <code>approxh</code> is too large. Default 1/3.
conditional	A logical value indicating whether standard logistic regression should be conditional if TRUE (i.e., <code>survival::clogit</code> ) or unconditional if FALSE ( <code>stats::glm</code> ). Default is FALSE.
tolval	Controls the tolerance on feasibility and optimality for the optimization procedure (performed by <code>solnp</code> in the <code>Rsolnp</code> package). Default 1e-6.
stepsz	Controls the step size for the optimization procedure (performed by <code>solnp</code> in the <code>Rsolnp</code> package). Default 1e-5.

**Details**

The function seeks to optimize a smooth approximation to the covariate-adjusted AUC,  $SaAUC = \sum_{c=1}^m w_c SAUC_c$  where  $SAUC_c$  are the smooth approximations to the covariate-specific AUC and  $w_c$  are covariate-specific weights for a covariate with  $m$  values in the data.

**Value**

A list will be returned with the following components:

NumCov	The number of covariate strata used after removing concordant strata.
FittedCombs	A list containing four fitted combinations: InitialVal (either robust logistic regression, if <code>initialval="rGLM"</code> , standard unconditional logistic regression, if <code>initialval</code> is not <code>"rGLM"</code> and <code>conditional=FALSE</code> , or standard conditional logistic regression, if <code>initialval</code> is not <code>"rGLM"</code> and <code>conditional=TRUE</code> ), NormGLM (standard unconditional or conditional logistic regression, depending on <code>conditional</code> ), NormrGLM (robust logistic regression), MaxSaAUC (SaAUC approach). All fitted combination vectors are normalized. If robust logistic regression fails to converge, standard unconditional or conditional logistic regression (depending on <code>conditional</code> ) is used instead, and a warning is given.
aAUCTR	A vector of the aAUC in the training data for the four fitted combinations.
varTR	A vector of the variability in the covariate-specific AUCs around the aAUC in the training data for the four fitted combinations.

**Note**

The function automatically removes any covariate strata that are concordant on the outcome (i.e., all 0 or all 1).

Warnings are issued if the SaAUC algorithm does not converge or if robust logistic regression fails to converge.

The standard unconditional or conditional logistic regression algorithm may not converge, producing a warning. If such a convergence failure occurs, the "GLM" results will be affected, as will the "rGLM" results if the robust logistic model also fails to converge, and the "SaAUC" results if `initialval` is not `"rGLM"` or if the robust logistic model fails to converge. Thus, users should be alert to any convergence failures.

**References**

- Bianco, A.M. and Yohai, V.J. (1996) Robust estimation in the logistic regression model. In Robust statistics, data analysis, and computer intensive methods (ed H. Rieder), pp 17-34. Springer.
- Janes, H. and Pepe, M.S. (2009) Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika*, pages 1-12.
- Meisner, A., Parikh, C.R., and Kerr, K.F. (2017). Developing biomarker combinations in multi-center studies via direct maximization and penalization. UW Biostatistics Working Paper Series, Working Paper 421.

**See Also**

[rlogit](#), [solnp](#)

**Examples**

```
expit <- function(x){
  exp(x)/(1+exp(x))
}

set.seed(1)

covar <- rep(c(1:4),each=100)

x1 <- rnorm(400,0,rep(runif(4,0.8,1.2),each=100))
x2 <- rnorm(400,0,rep(runif(4,0.8,1.2),each=100))
x3 <- rnorm(400,0,rep(runif(4,0.8,1.2),each=100))
x4 <- rnorm(400,0,rep(runif(4,0.8,1.2),each=100))

covint <- rep(runif(4,-1.5,1.5), each=100)

y <- rbinom(400,1,expit(covint + 1*x1 - 1*x2 + 1*x3 - 1*x4))
X <- cbind(x1,x2,x3,x4)

output <- maxadjAUC(outcome=y, predictors=X, covariate=covar, initialval="rGLM",
  approxh = 1/3, conditional=FALSE, tolval = 1e-6, stepsz = 1e-5)

output
```

# Index

maxadjAUC, 1

rlogit, 3

solnp, 3