

# Package ‘maxTPR’

July 26, 2017

**Type** Package

**Title** Maximizing the TPR for a Specified FPR

**Version** 0.1.0

**Date** 2017-07-26

**Author** Allison Meisner

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

**Description** Estimates a linear combination of predictors by maximizing a smooth approximation to the estimated true positive rate (TPR; sensitivity) while constraining a smooth approximation to the estimated false positive rate (FPR; 1-specificity) at a user-specified level.

**License** GPL-2

**LazyData** TRUE

**Imports** aucm, Rsolnp

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-07-26 21:41:21 UTC

## R topics documented:

maxTPR . . . . .	1
<b>Index</b>	<b>5</b>

---

maxTPR

*Maximizing the TPR for a Specified FPR*

---

## Description

Often in the risk prediction setting, there is interest in combining several predictors (e.g., biomarkers) into a single tool for prognosis, diagnosis or screening. One way to accomplish this is by targeting a measure of predictive capacity. In many cases, there is interest in the true positive rate (TPR; sensitivity) for a clinically meaningful false positive rate (FPR; 1-specificity). This function estimates a linear combination of predictors by maximizing a smooth approximation to the empirical TPR (sTPR) while constraining a smooth approximation to the empirical FPR (sFPR). Furthermore, since the TPR and FPR are determined both by the linear combination and the threshold (i.e., TPR is the proportion of diseased individuals whose linear combination value exceeds some threshold), this function estimates the combination and the threshold simultaneously. Estimates from robust logistic regression, specifically the method of Bianco and Yohai (implemented via the `aucm` package), are used as starting values for the linear combination.

## Usage

```
maxTPR(data, tval, initialval="rGLM", alpha = 0.5, approxh = 0.5,
        tolval = 1e-4, stepsz = 1e-5, multiplier = 2)
```

## Arguments

<code>data</code>	An object of class 'data.frame' where the first column contains the outcome (disease) indicator (1 for diseased, 0 for non-diseased), and the subsequent columns are the predictors. Note that missing observations are allowed, but they will be automatically removed. All columns of data must be numeric. The columns of data will be (re)named "D" for the first column and "V1", "V2", ... for the subsequent (predictor) columns.
<code>tval</code>	The acceptable FPR value. The method constrains the smooth approximation to the FPR to be less than or equal to <code>tval</code> ; see <code>alpha</code> below.
<code>initialval</code>	Starting values of the predictor combination for the smooth TPR maximization 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.
<code>alpha</code>	To improve performance, a small buffer may be added to <code>tval</code> . The parameter <code>alpha</code> controls the size of this buffer, relative to the number of controls (individuals without the disease). The default value is <code>alpha=0.5</code> , meaning that the default buffer is $0.5/(\text{number of controls})$ so the sFPR is constrained to be less than or equal to $tval + 0.5/(\text{number of controls})$ .
<code>approxh</code>	The tuning parameter for the smooth approximations is the ratio of the standard deviation of the linear combination (based on the starting values) to $n^{approxh}$ , where $n$ is the sample size. In particular, larger values of <code>approxh</code> will provide a better approximation to the TPR and FPR, though estimation may become unstable if <code>approxh</code> is too large. Default 0.5.
<code>tolval</code>	Controls the tolerance on feasibility and optimality for the optimization procedure (performed by <code>solnp</code> in the <code>Rsolnp</code> package). Default 1e-4.

stepsz	Controls the step size for the optimization procedure (performed by <code>solnp</code> in the <code>Rsolnp</code> package). Default $1e-5$ .
multiplier	Used to provide an initial value for the threshold to the optimization procedure. Using the starting values for the linear combination (based on robust logistic regression), a reasonable choice for this initial value is the threshold such that $sFPR = tval$ . This value can be found by using the <code>uniroot</code> function, which requires a range over which to search. The <code>multiplier</code> parameter controls the size of this range; if the range is not wide enough, the error 'f() values at end points not of opposite sign' will be seen, and <code>multiplier</code> should be increased. The size of <code>multiplier</code> will not generally have a large impact on results, though narrower (but valid) ranges may offer slightly better precision in the results from <code>uniroot</code> . Default 2.

### Value

A list with the following components:

sTPRrslt	The results from the smooth TPR maximization procedure, including 'delta' (the threshold estimated by the maximization procedure), 'deltaRE' (the threshold estimated based on quantiles of the combination estimated by the maximization procedure), the estimated combination coefficients, and an indicator of convergence for the optimization procedure.
rGLMrslt	The results from the robust logistic regression model (fit using <code>rlogit</code> ), including 'delta' (the threshold estimated based on quantiles of the combination estimated by robust logistic regression), the estimated combination coefficients, and an indicator of convergence for <code>rlogit</code> . Note that if <code>rlogit</code> fails to converge, these results will be identical to <code>GLMrslt</code> , since in this case, the estimates from standard logistic regression are used in place of those from robust logistic regression. Since the smooth TPR maximization procedure involves constraining the norm of the combination coefficients to be 1 for identifiability, this constraint was also applied to the robust logistic regression results.
GLMrslt	The results from the (standard) logistic regression model, including 'delta' (the threshold estimated based on quantiles of the combination estimated by (standard) logistic regression), the estimated combination coefficients, and an indicator of convergence for <code>glm</code> . Since the smooth TPR maximization procedure involves constraining the norm of the combination coefficients to be 1 for identifiability, this constraint was also applied to the logistic regression results.
Nobs	The number of observations remaining after observations with missing values were removed.

For all three methods, the combination coefficients are reported in the same order as the columns of data.

### References

Meisner, A., Carone, M., Pepe, M., and Kerr, K.F. (2017). Combining biomarkers by maximizing the true positive rate for a fixed false positive rate. UW Biostatistics Working Paper Series, Working Paper 420.

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.

**See Also**

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

**Examples**

```
set.seed(4)
x1 <- rnorm(400)
x2 <- rnorm(400)
y <- rbinom(400, 1, exp(x1+x2)/(1+exp(x1+x2)))
data <- data.frame(y, x1, x2)
maxTPR(data, 0.2)
```

# Index

maxTPR, 1

rlogit, 4

solnp, 4