

Ebayhelp

For help with program **Ebay** used in paper "Empirical Bayes estimates for large-scale prediction problems."

* In R, the command `source("Ebayprogs.R")` will recall program **Ebay** and also the auxiliary programs called by **Ebay**. These are in list form, and should be copied individually into the R workspace.

* In R, the command `source("Ebaydata.R")` will recall the prostate data microarray example **X0**, the expression level matrix **6033x102**, and the 102-vector indicating the two classes of patients **Y0**.

* In R, issue the command `library(splines)`.

* Then the command `Ebay(ktrunc=0)` will produce a more complete version of Table 2 in the paper.

Here is a description of the key variables in the program call for **Ebay**:

```
function (folds=0, X=X0, Y=Y0, alpha0=0.025, tsw=1, ktrunc=5, corr=1,
Sshrink=-1, bre=120, df=7, mmin=1, mmax=200, R=5, pil=0.5, cat=1, sw=0)
```

* folds :	with folds =0 (default) Ebay returns the full empirical Bayes prediction rule as shown in Tables 2 and 4
	with folds = <i>K</i> , say <i>K</i> =10, Ebay does <i>K</i> -fold cross-validation to estimate the overall error rate and the error rates within the two classes, along with standard error estimates.
* R :	The number of complete cross-validation cycles used to estimate the error rates; R =5 is the default.
* X :	The <i>N</i> by <i>n</i> matrix of expression levels; <i>N</i> genes, <i>n</i> subjects.
* Y :	The dichotomous <i>n</i> -vector showing class identifiers for the patients, "1" for class 1 and "2" for class 2.
* alpha0 :	The nominal target value for the empirical Bayes prediction rule; .025 is the default.
* tsw :	With tsw =1 (default), the <i>t</i> -statistics are transformed to <i>z</i> -values as described in Remark F of Section 8. Use tsw =0 to avoid the transformation.
* ktrunc :	Controls the amount of truncation at the extremes of the <i>z</i> -values as seen in Figure 4. Set ktrunc =0 for no truncation; ktrunc =5 is the default.

* corr:	corr =1 (default) produces the correlation corrections to the nominal error estimates, as described in Section 5.
	corr =0 turns this off.
* Sshrink:	Sshrink =-1 (default) uses the usual sample correlation matrix in the correlation corrections as described in Section 5.
	Sshrink =1 uses the empirical Bayes shrinkage estimate as described in Remark H of Section 8.
* bre:	bre = K divides the range of the z -values into K equal bins for use in the estimation of $f(z)$ described in Remark D.
	$K=120$ is the default. bre can also accept a full description of the binning, e.g., bre =seq(-4.5,4.5,.2).
* df:	The number of degrees of freedom in the glm estimation of $f(z)$; df =7 is the default.
* mmin and mmax:	The minimum and maximum number of steps allowed in the empirical Bayes prediction rule; defaults are mmin =1, mmax =200.
* pil:	The <i>a priori</i> probability on Class 1 as described in Remark B; pil =.5 is the default.