Canonical correlation analysis

Description

Canonical correlation analysis, following Brian McArdle's unpublished graduate course notes, plus improvements to allow the calculations in the case of very sparse and collinear matrices.

Usage

```
CCorA(Y, X1, X2=NULL, stand.Y=FALSE, stand.X1=FALSE, stand.X2=FALSE,
      print.plot=TRUE, print.obj=FALSE))
```

Arguments

Y	left matrix
X1	right matrix
X2	control for the linear effect (over X1) of an optional matrix of covariables
stand.Y	Y will be standardized if TRUE
stand.X1	X1 will be standardized if TRUE
stand.X2	X2 will be standardized if TRUE
print.plot	two biplots will be produced if TRUE
print.obj	object will be represented in biplots if TRUE

Details

Canonical correlation analysis (Hotelling 1936) seeks linear combinations of the variables of Y that are maximally correlated to linear combinations of the variables of X1. The analysis estimates the relationships and displays them in graphs. In the present function, the linear effects of a matrix X2 (optional) on X1 can be controlled for.

Algorithmic notes --

All data matrices are replaced by their PCA object scores, computed by SVD.

1. The blunt approach would be to read the three matrices, compute the covariance matrices, then the matrix [S12 %*% inv(S22) %*% t(S12) %*% inv(S11)]. Its trace is Pillai's trace statistic.

This approach may fail, however, when there is heavy multicollinearity in very sparse data matrices, as it is the case in 4th-corner inflated data matrices for example. The safe approach is to replace all data matrices by their PCA object scores.

2. Inversion by 'solve' is avoided. Computation of inverses is done by SVD in most cases.

3. Regression by OLS is also avoided. Regression residuals are computed by QR decomposition.

Value

Function **CCorA** returns a list containing the following results and matrices:

Pillai	Pillai's trace statistic = sum of canonical eigenvalues.
EigenValues	Canonical eigenvalues. They are the squares of the canonical correlations.
CanCorr	Canonical correlations.
Mat.ranks	Ranks of matrices Y and X1 (possibly after controlling for X2).
RDA.Rsquares	Bimultivariate redundancy coefficients (R-squares) of RDAs of Y|X1 and X1|Y.
RDA.adj.Rsq	RDA.Rsquares adjusted for n and number of explanatory variables.
AA	Scores of Y variables in Y biplot.
BB	Scores of X1 variables in X1 biplot.
Cy	Object scores in Y biplot.
Cx	Object scores in X1 biplot.

Reference

Hotelling, H. 1936. Relations between two sets of variates. Biometrika 28: 321-377.

Author

Pierre Legendre, Departement de Sciences Biologiques, Universite de Montreal.

Examples

```
# Example using random numbers
mat1 <- matrix(rnorm(60),20,3)
mat2 <- matrix(rnorm(100),20,5)
CCorA.out <- CCorA(mat1, mat2)

# Example using intercountry life-cycle savings data, 50 countries
# For information, type '?LifeCycleSavings'
pop <- LifeCycleSavings[, 2:3]
oec <- LifeCycleSavings[, -(2:3)]
CCorA.out <- CCorA(pop, oec)
```