Corrections for Review

Raju Rimal

Input Parameters

Lets define,

- $W_j, j = 1, \ldots m$ be **Response Components**
- $Y_j, j = 1, \ldots m$ be **Response Variables**
- $Z_i, i = 1, \ldots p$ be **Predictor Components**
- $X_i, i = 1, \ldots p$ be **Predictor variables**

Parameter	*Description*
Training Samples (n) `n = 100`	100 training samples
No. of Predictor Variables (p) `p = 20`	20 predictor variables
No. of Response (m) `m = 3`	3 response variables
No. of relevant Predictor variables for each response components `q = c(10, 10)`	10 Y's are relevant for W_1 10 Y's are relevant for W_2
Position of relevant predictor components `relpos = list(c(1, 4), c(2, 3))`	Z_1 and Z_4 are relevant for W_1 Z_2 and Z_3 are relevant for W_2
Combination of response components during rotation	Y_1 is obtained from W_1 only

Parameter	*Description*
`ypos = list(1, c(2, 3))`	Y_2 and Y_3 are obtained from rotating W_1 and W_3 (uninformative)
Coefficient of determination for each response components `R2 = c(0.6, 0.8)`	W1 and Y1 has coefficient of determination of 0.6 W2 and W3 (uninformative) (Y2 and Y3) has coefficient of determination of 0.8
Decay factor of multicollinearity (γ) `gamma = 0.9`	The value of `gamma` starts from 0 and goes 1 and sometimes more Higher the value of `gamma` larger the multicollinearity in X

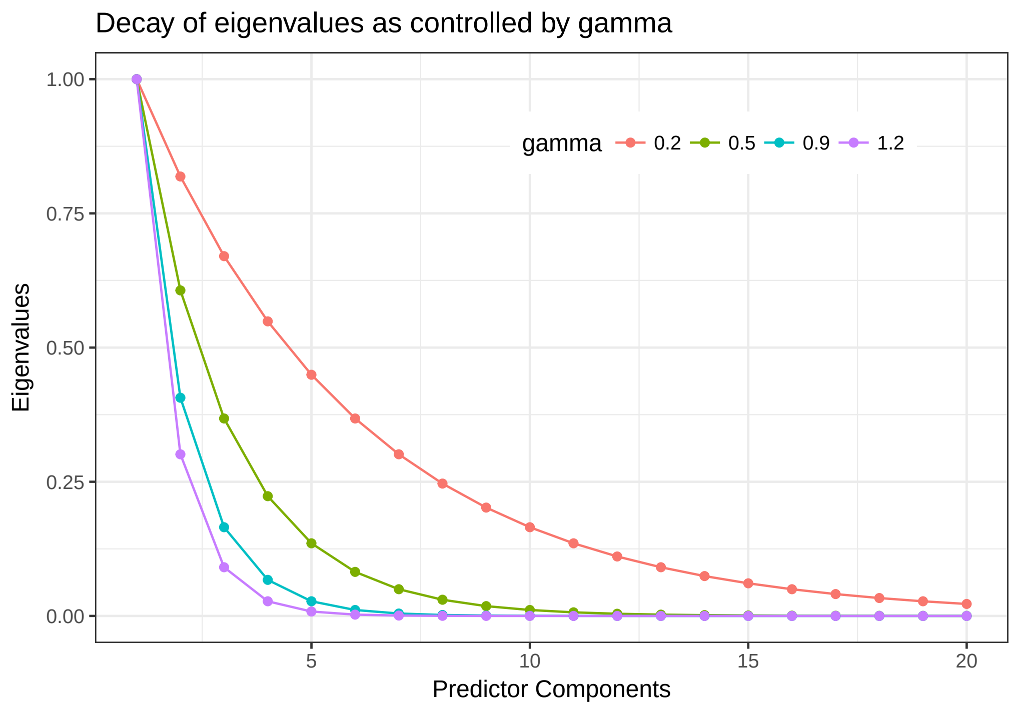
Simulation Assumptions

Decay of eigenvalues

The simulation has assumed that the eigenvalues of Σ_{xx} decreases exponentially by the factor controlled by γ. The expression can be written as,

$$\text{eigenvalues} = e^{-\gamma(i-1)}$$

Higher the value of `gamma` (γ) more collinearity can be seen in predictor variables and vice versa.



Equal variance of response components $(W_t W = I)$

W's are principle components of response variable Y. They are supposed to be independent, i.e. $\text{cov}(W_i, W_j) = 0$, which is natural however for simplicity of simulation procedure, the variance of W are considered to be equals to 1, i.e. $\text{var}(W_i) = 1 \forall i = 1, 2, \ldots m$.

Due to this assumption, user should expect to have the output response components to have same amount of variation. Using example from table above, W_1, W_2 (informative) and W_3 (non-informative) all have variance equals to 1. When W_2 and W_3 is rotated (`ypos = list(1, c(2, 3))`), it mix-up the information content in W_2 with large about of noise than compared to W_1. So the user will obtain $R^2 = 0.6$ for Y_1 while the $R^2 = 0.8$ for W_2 will be divided into Y_2 and Y_3 with $R^2 = 0.4$ each and the remaining will be the noise variance.

Non-overlapping relevant response components

This restriction assumed that any two response components does not share same relevant predictor components. This assumption was used to find the plausible covariance between W's and Z's. Without this assumption, the search of covariance that satisfies all the user supplied properties for simulated data becomes extensive and time consuming.

R package and Shiny Application

The Shiny Application and R package can obtained from GitHub and soon will be available in CRAN. The documentation for using and installing them is available at: http://github.com/therimalaya/...

Other Comments

- A generalized application for monte-carlo experiment by varying many properties at a time
- Wide matrices
- Only change few parameters fixing (randomizing) others
- Could not find the "recent paper" *generating multivariate random data* published in "same journal"

Conclusion

Whether comparing methods or assessing and understanding properties of any methods, tools or procedure; simulation data allows controlled tests for researchers. However, researchers spend enormous amount of time for creating such simulation tool so that they can obtain particular nature of data. This simulation tool will try to fulfill that lack. I hope this tool along with R-package and the easy to use shiny web interface will become an assistive tool for researcher.