

Results

Harpeth Lee

4/2/2022

Data Simulation

Our data simulations are done based on the latent variable model that PLS methods assume. We generate \mathbf{X} and \mathbf{Y} using the following equations, $\mathbf{X} = \phi\mathbf{A} + \epsilon_X$ and $\mathbf{Y} = \phi\mathbf{D} + \epsilon_Y$. The \mathbf{A} and \mathbf{D} matrices are fixed, providing most of the structure seen in \mathbf{X} and \mathbf{Y} . ϕ , ϵ_X , and ϵ_Y are randomly generated. The method by which ϕ is generated is less important as we want to approximate ϕ with our final model. ϵ_X , and ϵ_Y are sampled from a normal distribution.

Although ϕ is generated using the same method as ϵ_X , and ϵ_Y , it is important to note that ϵ_X and ϵ_Y are noise while ϕ is not. Ideally, our models will be able to identify ϕ while ignoring ϵ_X and ϵ_Y . ϕ is generated using the `mvrnorm` function in the `MASS` package in order to make sure columns are uncorrelated. ϵ_X and ϵ_Y are generated using `rnorm` as this will cause some random correlation in the noise.