Known Unknowns

Currently I am confused about the `loadings.A` and `loadings.B` objects that are returned by the `spcr` function. I am not sure if I need to use both to reconstruct data. If both are needed how are they used (I am guessing it's something like AXB^T).

When performing PCR, model will be of form $y = \beta_0 + \beta_1 * \lambda_1 + \beta_2 * \lambda_2 ...$ where λ_i are principal components. For test data, do we calculate the principal components and then plug them in or "reconstruct" the original principal components using loading matrices. I would guess it's the second as the principal components for test data may vary widely.

How do we select values of `lambda.B` and `lambda.gamma` as to optimize model performance?

What are the L_1 and L_2 norms? - These $L_1 = \sum_{i=1}^{p} |\beta_i|$ and $L_2 = \sqrt{\sum_{i=1}^{p} \beta_i^2}$. The L_1 norm will force some values of β_1 to 0 which the L_2 norm won't.

How is the Monte Carlo Method used to test model performance?

With large datasets does LASSO/L_1 norm tend to work better as it will remove highly correlated predictors from a model? Assuming that large datasets will tend to have a number of highly correlated predictors.

Function $tr\{\}$ is trace of matrix.

Why do PCR/SPCR models sometimes have a higher valued principle components that end up explaining a large amount of variance in the predictor? Why don't these show up sooner?

I am a bit confused as to how PLS compares to PCR. Why doesn't it minimize a matrix equation?

Way to impose L_1 norm or something similar without needing to find matrix inverse, max/min problem.

PLS with variable selection, e.g variable selection occurs and then model is built instead of folding variable selection into algorithm for building model.

2/17

Why Q^2? What justifies using just $R^2 - Q^2$?

Could we use cross validation to build models instead? Why Q^2 and partial least squares? Interaction terms? WOuld model improve with added nonlinear interaction terms?