Week 14 – Dimensionality reduction

L14-06. Partial least square (PLS) and Canonical correlation analysis (CCA) (What & When)

Donghee and Jungwoo

PCA: Decomposition of data covariance matrix!

ICA: Decomposing data to spatially or temporally independent components!

NNMF: Decomposing matrix with "**non-negative**" elements

All non-supervised dimension reduction…

PLS and CCA

are supervised dimension reduction! which allows multivariate dependent variables!

Use information of dependent variable to reduce dimension!

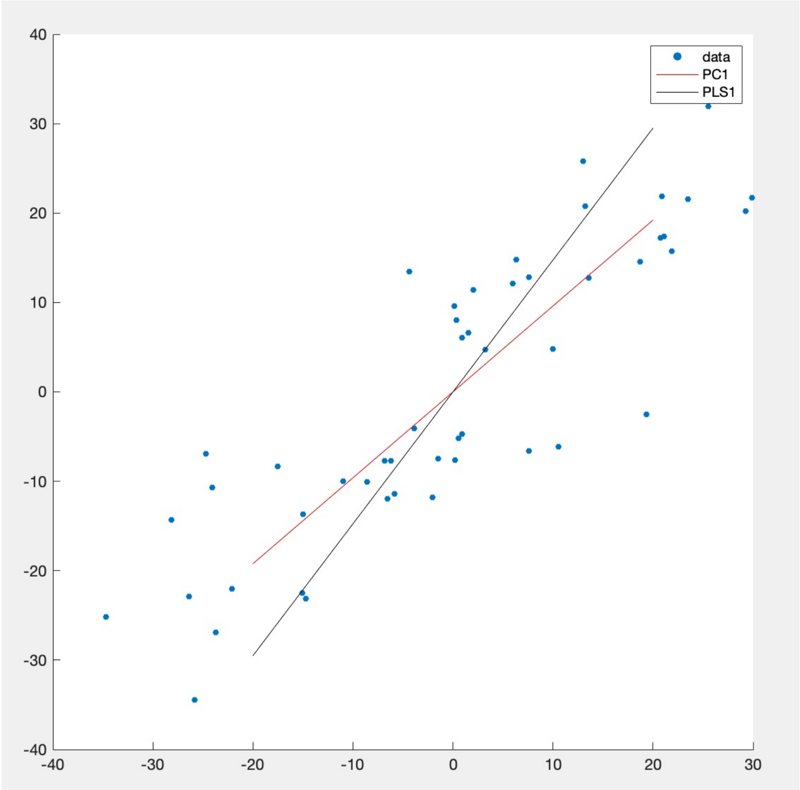
Partial least Square regression

Finding latent spaces that maximize **covariance between X and Y**, rather than variance of X which PCA does.

$$\max\ var(Xw) \rightarrow \max cov(Xw, Yz)$$

PCA PLS

(w and z are coeffients of each X and Y)



In the right example, there are two axis where data can be projected,
So, if data is projected on to PC1 (Xw), it retains the maximum variance of data than any other axis.

If it is projected on to PLS1 (Xw), it explains the maximum variance of projected Y (Yz) which is not shown in the graph.

Partial least Square regression

How…?

Insights from PCA…

In PCA…we can obtain PCs from singular value decomposition (SVD)

$$X = U\Sigma V^T$$

If *X* is p x n, columns of *U* are principal compnents!

where X is mean centered…

Note: Usually we're interested in *spatial* PCs,
but columns of *V* can be interpreted as *temporal* PCs

Note: $U\Sigma$ can be considered as "*functional gradient*"

How…?

Same logic can be applied to covariance matrix

$$XX^T = U\Sigma V^T V\Sigma^T U^T$$

$$XX^T = U\Lambda U^T$$

XX^T divided by N-1 yields covariance matrix

This is exactly **Eigen Value Decomposition (EVD)**

SVD of X of EVD of covariance matrix gives you principal components!

where X is mean centered…

How…?

Exactly same way but substituting one X to Y which yields cross-covariance matrix.

X: p x n
Y: q x n

$$\underbrace{XY^T}_{p \times q} = U\Sigma V^T$$

This case, columns of U gives you the components in X space (p)
while columns of V gives you the components in Y space (q)

Those U and V are w and z in the equation we saw!

$$\max cov(Xw, Yz)$$

where X and Y are both mean centered…

Its name has "regression"…where are regression coefficients?

```
for i = 1:numcomp
    [W, ~, Q] = svd(X' * Y, 'econ');
    t = X * W(:, 1); % w
    u = Y * Q(:, 1); % q, if Y is univariate, first u is equal to Y(mean centered) and Q is
always just 1.
    Xweights(:, i) = W(:, 1);
    Yweights(:, i) = Q(:, 1);

    Xscore(:, i) = t;
    Yscore(:, i) = u;
    b(i, :) = u'*t / (t'*t); % b is the regression coefficients. t as the independent, u as
the dependent variable.

    X = X - t * (t' * X) / (t' * t); % regress out t
    Y = Y - t * (t' * Y) / (t' * t);

    % X is "deflated" in each for loop. It means information on the space
    % which has maximum covariance with Y and X are removed. Consequently,
    % data after deflation are moved to orhtogonal space of the t.
    % So, cov(Xscore) yields diagonal matrix.
end
```

t and u here is the projected value of X and Y.

b is the regression coefficients of t regard to u.

and regress out t in X and Y.

https://github.com/didch1789/yanchogosu_toolbox

What's the point?

Say we are to make a model predictive of Y1, Y2 and Y3.

With the PLS, we can make a model predictive of each Ys, but the model includes not only the information of Y itself, but also information in the different Ys.

In other words,
if each Y itself may not fully represents the information in independent variables,
PLS also takes effects of other Ys into consider to make a model.

Examples in Phil Kragel's paper!

NEW!!

Then what is CCA…?

On the Equivalence Between Canonical Correlation Analysis and Orthonormalized Partial Least Squares

Liang Sun†§, Shuiwang Ji†§, Shipeng Yu‡, Jieping Ye†§
†Department of Computer Science and Engineering, Arizona State University
§Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University
‡CAD and Knowledge Solutions, Siemens Medical Solutions USA, Inc.
†§{sun.liang, shuiwang.ji, jieping.ye}@asu.edu; ‡shipeng.yu@siemens.com

dimensional variables. The fundamental difference between CCA and PLS is that CCA maximizes the correlation while PLS maximizes the covariance.

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Covariance divided by standard deviation of X and Y is correlation.

How those kind of "standadization" can yield differenct results with PLS…?

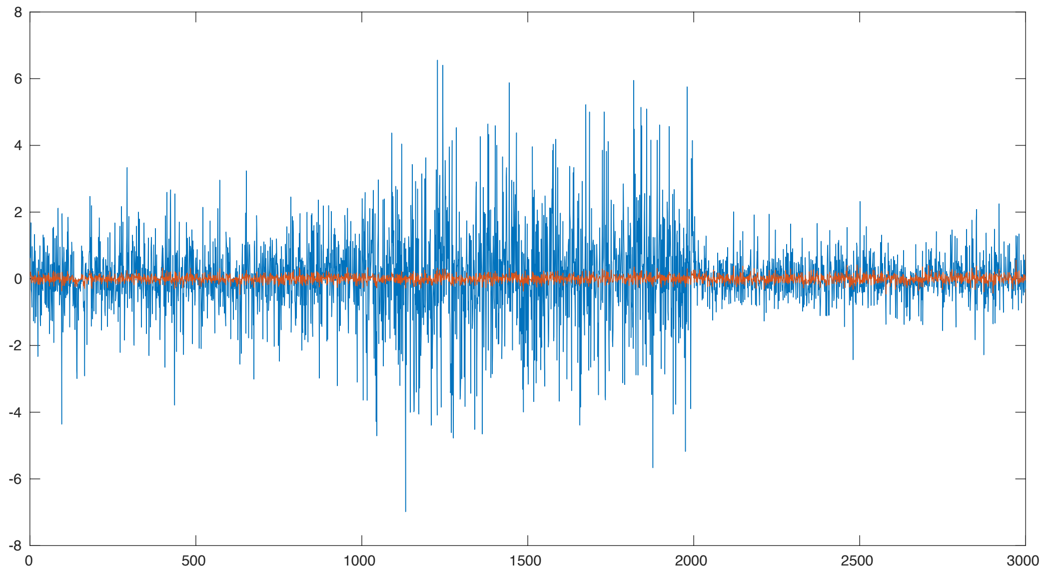
```
X1 = fmri_gen_pseudodat_ycgosu(100, 1000, 'noisetype', 'gaussian');
y = fmri_gen_pseudodat_ycgosu(100,3, 'noisetype', 'gaussian');
N = 100;

X0 = X1 - mean(X1);          y0 = y -mean(y);


covXY = (X0'*y0) ./ (N-1);  corrXY = corr(X0, y0);

covXYflat = covXY(:);        corrXYflat = corrXY(:);

corr(covXYflat, corrXYflat) % 0.8875
```



Pattern of covariance matrix is not fully conserved in correlation matrix.

Not only that…

PLS cost function was…

$$\max cov(w_x{}^T X, YTWy)$$

$$\max_{W_x, W_y} \quad \mathrm{tr}\left(W_x^T X Y^T W_y\right)$$

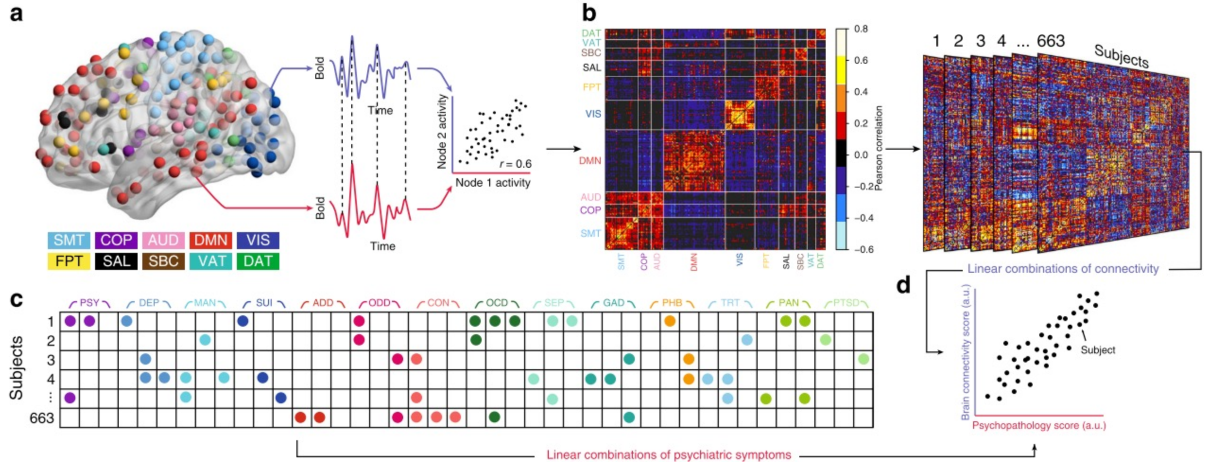
$$\text{subject to} \quad W_x^T X X^T W_x = I, \;\; W_y^T Y Y^T W_y = I$$

Cost function trying to solve is little bit different!

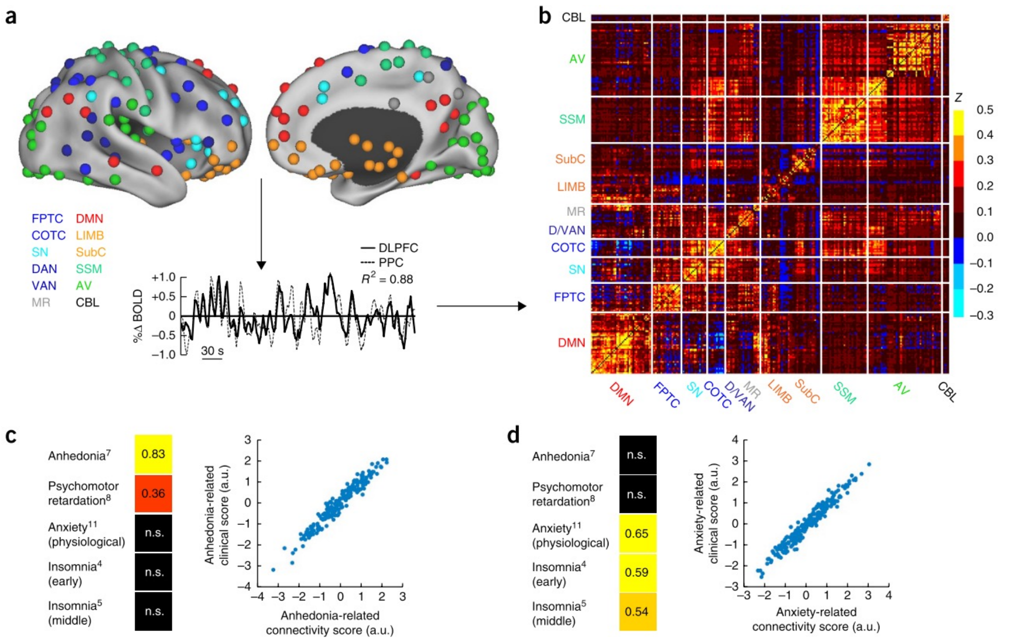
∴ PLS and CCA are common in that it tries to find subspaces in both independent and dependent variables that retains the information of each other.
But, in a little different way.

PLS & CCA

Examples…



Xia et al. 2018



Drysdale et al. 2017

In the actual field…
seems PLS is commonly used for regression method (explaining the variance of depedent variables)
while CCA is used for finding latent dimensions explaining both both indep. and dependent variables.

Important issue!

Since it learned its features supervised by Y…
it is susceptible to *overfitting*…

Need independent data set for testing!!

On stability of Canonical Correlation Analysis and Partial Least Squares with application to brain-behavior associations

Markus Helmer[a], Shaun Warrington[b], Ali-Reza Mohammadi-Nejad[b,c], Jie Lisa Ji[a,d], Amber Howell[a,d], Benjamin Rosand[e], Alan Anticevic[a,d,f], Stamatios N. Sotiropoulos[b,c,g,1], and John D. Murray[a,d,e,1]

[a]Department of Psychiatry, Yale School of of Medicine, New Haven, CT 06511; [b]Sir Peter Mansfield Imaging Centre, School of Medicine, University of Nottingham, Nottingham, NG7 2UH, United Kingdom; [c]National Institute for Health Research (NIHR) Nottingham Biomedical Research Ctr, Queens Medical Ctr, Nottingham, United Kingdom; [d]Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, CT 06511, USA; [e]Department of Physics, Yale University, New Haven, CT 06511, USA; [f]Department of Psychology, Yale University, New Haven, CT 06511, USA; [g]FMRIB, Wellcome Centre for Integrative Neuroimaging, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DU, United Kingdom

Cocoan 101

https://cocoanlab.github.io