Lecture 10

Monday, February 14, 2022 8:35 AM

Admin

1. HW #1 due.
2. HW #2 posted.
   ↳ Mar 2nd.

Partial Least Squares Regression and Kalman Filtering

Partial Least Squares Regression

- previously, we derived "ordinary least squares"

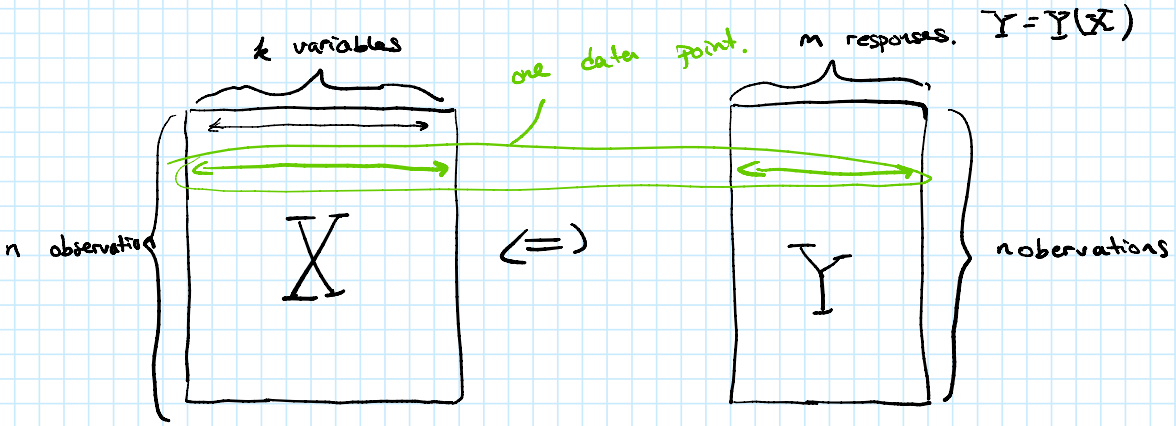
$$y = \varphi^T \theta \leftarrow \text{model structure}$$

$$\hat{\theta} = \left[\sum_{i=1}^{N} \varphi_i \varphi_i^T\right]^{-1} \left[\sum_{i=1}^{N} y_i \varphi_i\right]$$

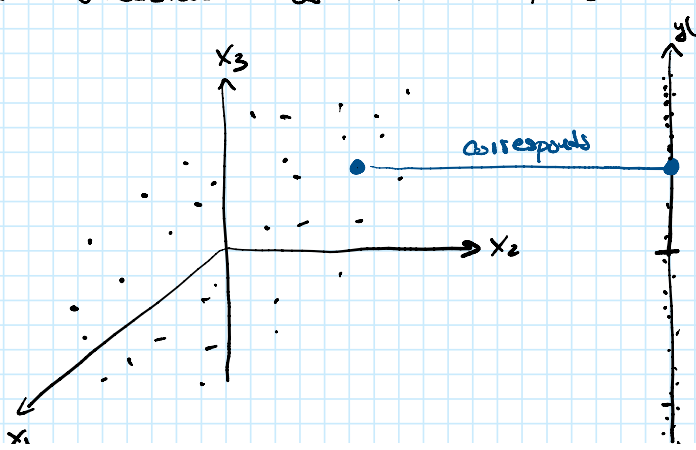
$\hat{\theta}$: k_θ parameters; $\sum_{i=1}^{N} \varphi_i \varphi_i^T$: auto-cov matrix; N: # of observations

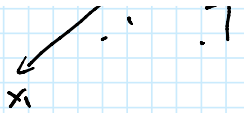
- what if $N < k_\theta$? or $N << k_\theta$?
  - → Cannot uniquely identify $\hat{\theta}$.
  - → we need a structured way to relate inputs & outputs when $N << k$.
    - → one approach is called "principal component regression".

Partial Least Squares Regression (PLSR) — Eriksson, 1993



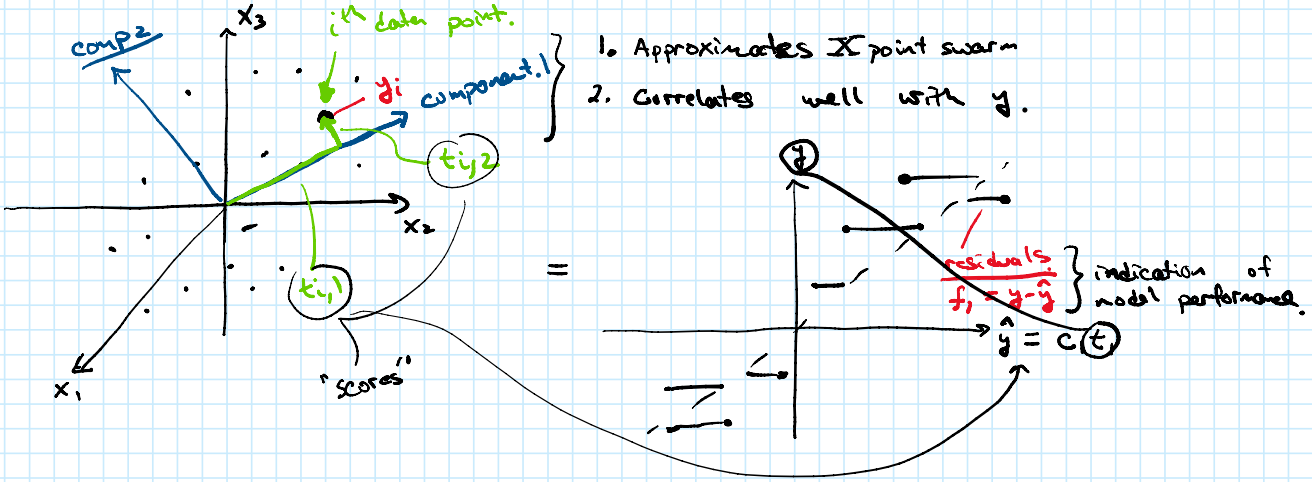
- As a regression problem, the observations X, Y can be understood as two point swarms.



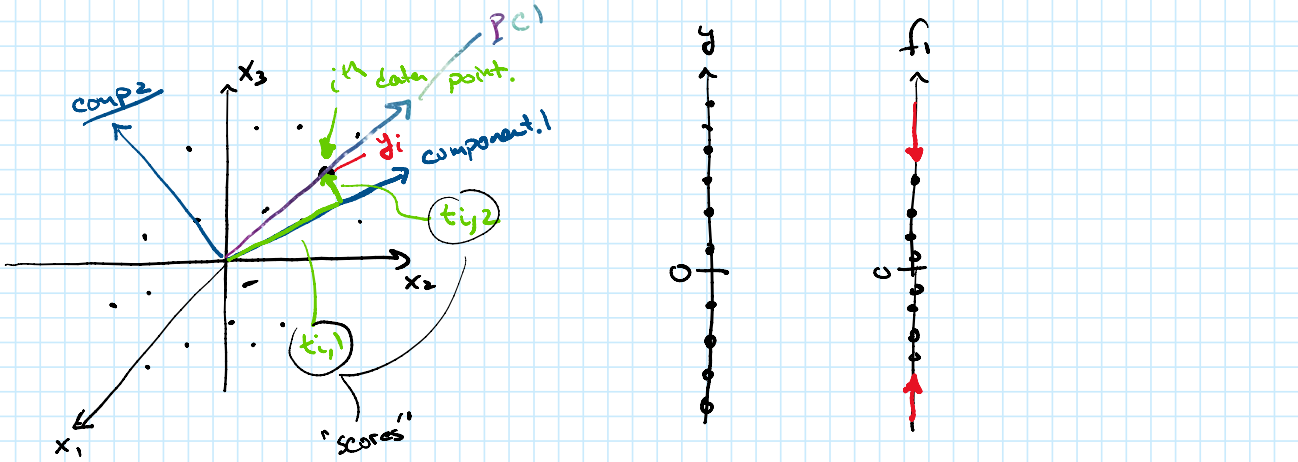


pre-processing:
- Typically z-score the data, as in ICA
- Can selectively scale columns in X or Y to increase importance of that column in the model.

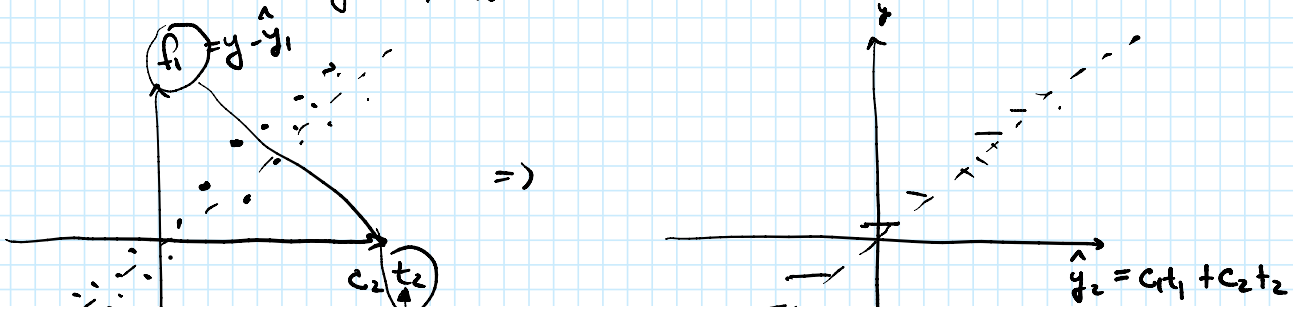
- Computing the 1st PLSR component.

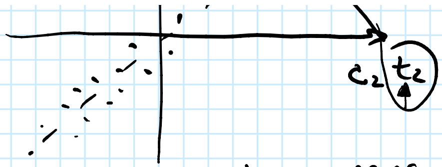


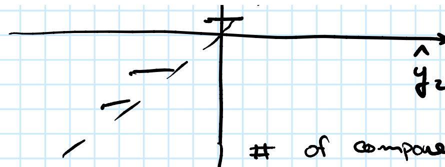
- Computing 2nd PLSR component



- The second component is orthogonal to the 1st, captures as much of X variance as possible, while correlating w/ f_1.

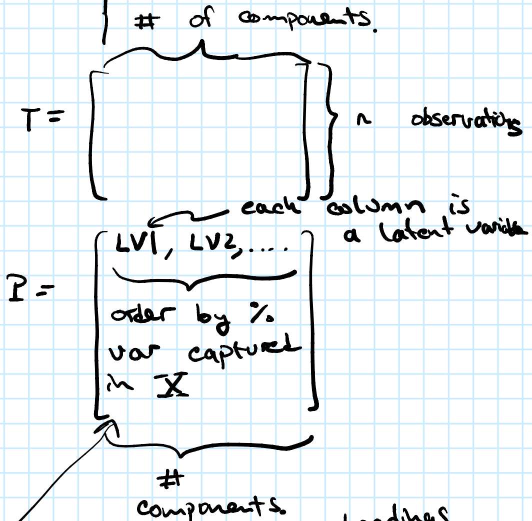




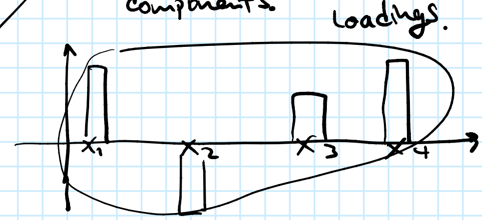


$\hat{y}_2 = c_1 t_1 + c_2 t_2$

- $X = \tilde{X} + TP^T + E$
  - $\tilde{X}$ ← column means
  - E ← residual matrix
  - T ← scores matrix
  - P ← loadings matrix



$Y = \tilde{Y} + TC^T + G$
- $\tilde{Y}$ ← column means
- G ← residual matrix
- T ← scores matrix
- C ← weighting matrix relating T to Y



- Difference between PCA & PLSR.

  <u>PCA</u>: set $\text{cov}(t_{ij}, t_{ik}) = 0,\ j \neq k$
  $\Rightarrow$ maximizes $\text{var}(t_{ij})$

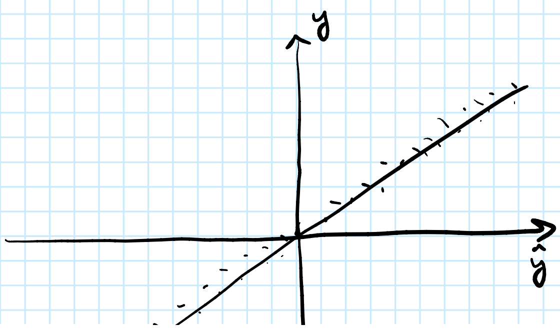
  <u>PLSR</u>: maximizes a combination of
  1. $\text{Cov}(t_{ij}, y)$ & $\text{var}(t_{ij})$

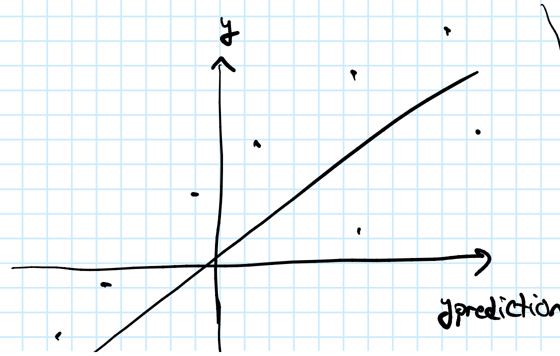
- PLS often computed via Non-iterative partial least squares Regression (NIPALS)
  → supplemental Reading: simple explanation of PLS.
  → ropls (R), plsregress in MATLAB.

- Can be used for:
  - continuous variables → partial least squares regression (PLSR)
  - discretete variables → partial least squares discriminant analysis (PLSDA) or Discriminant - PLSR.

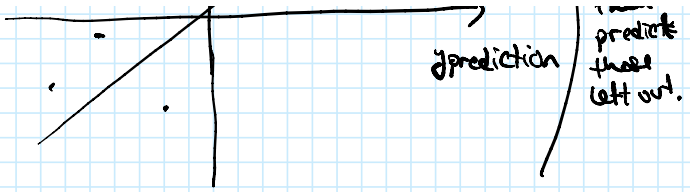
- If using for modeling, a key considration is over-fitting
  $\Rightarrow$ too many components fit noise.

  $\Rightarrow$ can use a leave K-out approach to determine predictive capacity:





Leave k points out, fit model to rest, then predict those left out.



- will discuss more in information theory section of course.
- How to predict new samples?

1. Use <u>same</u> z-transform as for original model.
2. $T_{pred} = X_{new}(P)$ — from original model
   $Y_{pred} = T_{pred}(C^T)$