Statistical learning: Advanced

Alain Celisse

SAMM

Paris 1-Panthéon Sorbonne University

`alain.celisse@univ-paris1.fr`

Lecture 1: Components-based methods

—

Master 2 TIDE – Paris 1 – Fall 2021

Outline of the lectures

Successive topics of the coming lectures:

1. Component-based analysis:
   PCA, PLS, CCA,...
2. Beyond quantitative data: Categorical data
   Correpondence Analysis (CA), MCA, FAMD
3. Clustering from dissimilarity measures:
   Similarity, K-means and kernelized K-means,
   Hierarchical clustering
4. Change-point detection
   Parametric models and non-parametric ones
5. Multiple testing and FDR control

Outline of the lecture

- PCA: Principal Component Analysis
- PLS: Partial Least-Squares
- CCA: Canonical Correlation Analysis

Principal Component Analysis (PCA)

Unsupervised situation

Notation

- n individuals
- Each individual i described by d (quantitative) variables $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$

Unsupervised situation

Notation

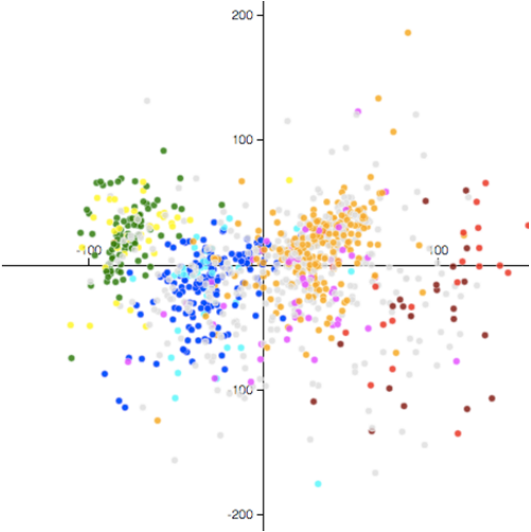
- n individuals
- Each individual i described by d (quantitative) variables $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$

There is no label Y in the PCA story!

Assumption for having PCA work well

Key assumption

Variability carries information about heterogeneity



PCA algorithm

$$\widehat{\mathrm{Var}}(X^\top \theta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta = \frac{1}{n} \left(\boldsymbol{X}\theta\right)^\top \left(\boldsymbol{X}\theta\right) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$$

PCA algorithm

- **Step 1:**
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle \right\}$$

PCA algorithm

$$\widehat{\mathrm{Var}}(X^\top \theta) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta = \frac{1}{n}\left(\boldsymbol{X}\theta\right)^\top \left(\boldsymbol{X}\theta\right) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$$

PCA algorithm

- **Step 1:**
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle \right\}$$

  - $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first principal component

PCA algorithm

$$\widehat{\mathrm{Var}}(X^\top\theta) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta = \frac{1}{n}(\boldsymbol{X}\theta)^\top(\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle$$

PCA algorithm

- **Step 1:**
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle\}$$

  - $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first principal component
- **Step t ($2 \leq t \leq \mathrm{rank}(\boldsymbol{X}) = r$):**
  - Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j\rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle\}$$

PCA algorithm

$$\widehat{\mathrm{Var}}(X^\top\theta) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta = \frac{1}{n}\left(\boldsymbol{X}\theta\right)^\top\left(\boldsymbol{X}\theta\right) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$$

PCA algorithm

- **Step 1:**
  - Find
    $$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\right\}$$
  - $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first principal component
- **Step t $(2 \leq t \leq \mathrm{rank}\left(\boldsymbol{X}\right) = r)$:**
  - Find
    $$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \left\langle \boldsymbol{X}\theta, h^j \right\rangle = 0 \\ j = 1, \ldots, t-1}} \left\{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\right\}$$
  - $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth principal component

PCA algorithm

$$\widehat{\mathrm{Var}}(X^\top \theta) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta = \frac{1}{n}(\boldsymbol{X}\theta)^\top(\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle$$

PCA algorithm

- **Step 1:**
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle\}$$

  - $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first principal component
- **Step t** $(2 \leq t \leq \mathrm{rank}(\boldsymbol{X}) = r)$**:**
  - Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j\rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta\rangle\}$$

  - $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth principal component
- **Output:**

$$\boldsymbol{X} \approx \widetilde{\boldsymbol{X}} = \left[\, h^1; h^2; \ldots; h^r \,\right] = \boldsymbol{X}\left[\, \theta^1; \ldots; \theta^r \,\right]$$

PCA: Focus on the first step

Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first principal component

Interpretation

- $\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle = \|\boldsymbol{X}\theta\|^2 = \theta^\top \boldsymbol{X}^\top \boldsymbol{X}\theta = n\widehat{\mathrm{Var}}(X^\top \theta)$
- h^1 is the first (one-dimensional) approximation to $\boldsymbol{X}$ for describing each individual i by $h_i^1 \in \mathbb{R}$ instead of $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$

PCA: Focus on the other steps

Step t $(2 \leq t \leq r = \mathrm{rank}\,(\boldsymbol{X}))$

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \left\langle \boldsymbol{X}\theta, h^j \right\rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth principal component

PCA: Focus on the other steps

Step t $(2 \leq t \leq r = \operatorname{rank}(\boldsymbol{X}))$

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth principal component

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, (\boldsymbol{X}^\top \boldsymbol{X})\, \theta^j \rangle = 0$ for all $2 \leq j \leq r$
- Any new component must be $\perp$ previous ones
- $(h^1, \ldots, h^r)$: orthogonal family
- Each individual i described by:

$$(h_i^1, h_i^2) \in \mathbb{R}^2 \text{ instead of } \boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$$

(with two components)

Partial Least-Squares (PLS 1)

Supervised situation

Notation

- n individuals
- Each individual i described by d (quantitative) variables $X^1, \ldots, X^d$
- $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$
- Label:

Each individual i described by Y_i

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$

Supervised situation

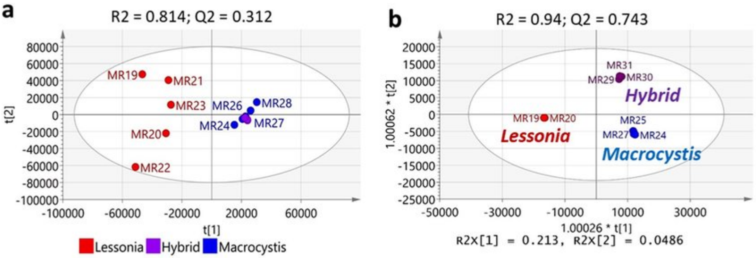
Notation

- n individuals
- Each individual i described by d (quantitative) variables $X^1, \ldots, X^d$
- $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$
- Label:

  Each individual i described by Y_i
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$

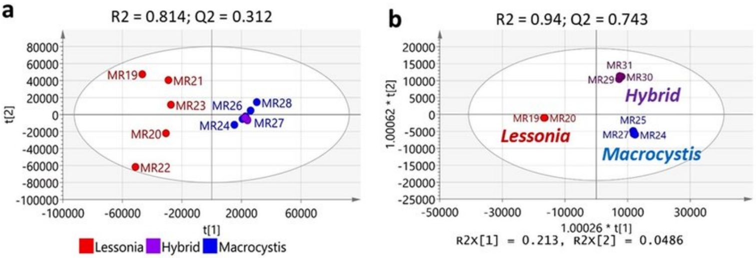
Unlike PCA, the PLS framework is a supervised one!

Comparison PCA/PLS



- **Picture (a):** If one class exhibits more variability than the others, PCA focuses on this class for defining axes $\longrightarrow$ blue and purple points gathered within one cluster

Comparison PCA/PLS



- **Picture (a):** If one class exhibits more variability than the others, PCA focuses on this class for defining axes $\longrightarrow$ blue and purple points gathered within one cluster
- **Picture (b):** With PLS, the axes split labeled classes into well separated clusters

PLS algorithm: Step 1

$$\widehat{\mathrm{Cov}\,(X^\top\theta, Y)} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta \cdot Y_i = \frac{1}{n}\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle$$

(columns of $\boldsymbol{X}$ centered and reduced)
(vector $\boldsymbol{Y}$ is centered)

PLS algorithm: Step 1

$$\widehat{\mathrm{Cov}\left(X^\top\theta, Y\right)} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top\theta \cdot Y_i = \frac{1}{n}\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle$$

(columns of $\boldsymbol{X}$ centered and reduced)
(vector $\boldsymbol{Y}$ is centered)

PLS-1 algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle \right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first PLS component

PLS algorithm: Step 1

$$\widehat{\mathrm{Cov}\left(X^\top\theta, Y\right)} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta \cdot Y_i = \frac{1}{n}\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle$$

(columns of $\boldsymbol{X}$ centered and reduced)
(vector $\boldsymbol{Y}$ is centered)

PLS-1 algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first PLS component

Remarks:

- h^1 maximizes the covariance with $\boldsymbol{Y}$ (explicability)

PLS algorithm: Step 1

$$\widehat{\mathrm{Cov}\left(X^\top\theta, Y\right)} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_{i\cdot}^\top \theta \cdot Y_i = \frac{1}{n}\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle$$

(columns of $\boldsymbol{X}$ centered and reduced)
(vector $\boldsymbol{Y}$ is centered)

PLS-1 algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first PLS component

Remarks:

- h^1 maximizes the covariance with $\boldsymbol{Y}$ (explicability)
- Focuses on covariance ($\neq$ correlation!)
- Not scale invariant
  $\longrightarrow$ centering and reducing columns of $\boldsymbol{X}$

PLS algorithm: Step 2

PLS-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth PLS component

PLS algorithm: Step 2

PLS-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}\,(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth PLS component

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, \left(\boldsymbol{X}^\top \boldsymbol{X}\right) \theta^j \rangle = 0$ for all $2 \leq j \leq r$
- Any new PLS component must be $\perp$ previous ones

PLS algorithm: Step 2

PLS-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth PLS component

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, (\boldsymbol{X}^\top \boldsymbol{X})\, \theta^j \rangle = 0$ for all $2 \leq j \leq r$
- Any new PLS component must be $\perp$ previous ones
- $(h^1, \ldots, h^r)$: orthogonal family

PLS algorithm: Step 2

PLS-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}\,(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\|\theta\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth PLS component

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, \left(\boldsymbol{X}^\top \boldsymbol{X}\right) \theta^j \rangle = 0$ for all $2 \leq j \leq r$
- Any new PLS component must be $\perp$ previous ones
- $(h^1, \ldots, h^r)$: orthogonal family
- Each individual i described by:

$$(h_i^1, h_i^2) \in \mathbb{R}^2 \text{ instead of } \boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$$

(with two components)

Revisiting PLS algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\}$$

- With $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$

Calculating the solution

$$\begin{aligned} & \theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\} \\ \Leftrightarrow \quad & \theta^1 \in Arg \max_{\|\theta\|=1} \left\{\left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle\right\} \\ \Leftrightarrow \quad & \theta^1 = \frac{\boldsymbol{X}^\top \boldsymbol{Y}}{\left\|\boldsymbol{X}^\top \boldsymbol{Y}\right\|} \end{aligned}$$

Revisiting PLS algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle \right\}$$

- With $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$

Calculating the solution

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle \right\}$$

$$\Leftrightarrow \quad \theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^1 = \frac{\boldsymbol{X}^\top \boldsymbol{Y}}{\left\| \boldsymbol{X}^\top \boldsymbol{Y} \right\|}$$

$\longrightarrow$ The first PLS component is $h^1 = \boldsymbol{X}\theta^1 = \frac{\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}}{\left\| \boldsymbol{X}^\top \boldsymbol{Y} \right\|}$

Revisiting PLS algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\}$$

- With $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$

Calculating the solution

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{Y}\rangle\}$$

$$\Leftrightarrow \quad \theta^1 \in Arg \max_{\|\theta\|=1} \left\{\left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y}\right\rangle\right\}$$

$$\Leftrightarrow \quad \theta^1 = \frac{\boldsymbol{X}^\top \boldsymbol{Y}}{\left\|\boldsymbol{X}^\top \boldsymbol{Y}\right\|}$$

$\longrightarrow$ The first PLS component is $h^1 = \boldsymbol{X}\theta^1 = \dfrac{\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}}{\left\|\boldsymbol{X}^\top \boldsymbol{Y}\right\|}$

Interpretation of h^1

$h^1 = \widehat{\boldsymbol{Y}}^1$: first explanation of $\boldsymbol{Y}$ by "the model" $\boldsymbol{X}\theta$ (regression)

Revisiting PLS algorithm: Step 2 ($t = 2$)

Orthogonality constraint $\left(V_1 = \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right)\right)$

$$\langle \boldsymbol{X}\theta, h^1 \rangle = 0 \quad \Leftrightarrow \quad \left\langle \theta, \boldsymbol{X}^\top h^1 \right\rangle = 0$$

$$\Leftrightarrow \quad \theta \in \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right)^\perp = V_1^\perp$$

Revisiting PLS algorithm: Step 2 ($t = 2$)

Orthogonality constraint

$$(V_1 = \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right))$$

$$\left\langle \boldsymbol{X}\theta, h^1 \right\rangle = 0 \quad \Leftrightarrow \quad \left\langle \theta, \boldsymbol{X}^\top h^1 \right\rangle = 0$$

$$\Leftrightarrow \quad \theta \in \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right)^\perp = V_1^\perp$$

Constrained maximization

Under this orthogonality constraint, solving

$$\max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle \right\} \quad \Leftrightarrow \quad \max_{\|\theta\|=1} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\}$$

where $\Pi_{V_1}(\cdot)$ denotes the orthogonal projection onto V_1

Reminder:

$$u - \Pi_{V_1}(u) \in V_1^\perp, \qquad \forall u \in \mathbb{R}^d$$

Revisiting PLS algorithm: Step 2 ($t = 2$)

Solution at $t = 2$

$$
\begin{aligned}
& \theta^2 = Arg \max_{\substack{\|\theta\| = 1 \\ \langle \theta, \boldsymbol{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle \right\} \\
\Leftrightarrow \quad & \theta^2 = Arg \max_{\substack{\|\theta\| = 1 \\ \langle \theta, \boldsymbol{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\} \\
\Leftrightarrow \quad & \theta^2 = \frac{\boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})}{\left\| \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right\|}
\end{aligned}
$$

Revisiting PLS algorithm: Step 2 ($t = 2$)

Solution at $t = 2$

$$\theta^2 = Arg \max_{\substack{\|\theta\| = 1 \\ \langle \theta, \mathbf{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \mathbf{X}^\top \mathbf{Y} \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^2 = Arg \max_{\substack{\|\theta\| = 1 \\ \langle \theta, \mathbf{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \mathbf{X}^\top \mathbf{Y} - \Pi_{V_1}(\mathbf{X}^\top \mathbf{Y}) \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^2 = \frac{\mathbf{X}^\top \mathbf{Y} - \Pi_{V_1}(\mathbf{X}^\top \mathbf{Y})}{\left\| \mathbf{X}^\top \mathbf{Y} - \Pi_{V_1}(\mathbf{X}^\top \mathbf{Y}) \right\|}$$

Reminder:

$$\Pi_{V_1}(\mathbf{X}^\top \mathbf{Y}) = \left\langle \mathbf{X}^\top \mathbf{Y}, \frac{\mathbf{X}^\top h^1}{\left\| \mathbf{X}^\top h^1 \right\|} \right\rangle \frac{\mathbf{X}^\top h^1}{\left\| \mathbf{X}^\top h^1 \right\|}$$

PLS-1 algorithm: Summary

Pseudo-code

Input: $\boldsymbol{X}$, $\boldsymbol{Y}$

- **Initialization** $t = 0$:
$$R^0 = \boldsymbol{X}^\top \boldsymbol{Y}$$
- **Step** $t = 1$:
  - $\theta^1 = \frac{R^0}{\|R^0\|}$ and $h^1 = \boldsymbol{X}\theta^1$
  - $\Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|} \right\rangle \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|}$
  - $R^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$

PLS-1 algorithm: Summary

Pseudo-code

Input: $\boldsymbol{X}$, $\boldsymbol{Y}$

- **Initialization** $t = 0$:
$$R^0 = \boldsymbol{X}^\top \boldsymbol{Y}$$
- **Step** $t = 1$:
  - $\theta^1 = \frac{R^{\mathbf{0}}}{\|R^{\mathbf{0}}\|}$ and $h^1 = \boldsymbol{X}\theta^1$
  - $\Pi_{V_{\mathbf{1}}}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^{\mathbf{1}}}{\|\boldsymbol{X}^\top h^{\mathbf{1}}\|} \right\rangle \frac{\boldsymbol{X}^\top h^{\mathbf{1}}}{\|\boldsymbol{X}^\top h^{\mathbf{1}}\|}$
  - $R^1 = R^0 - \Pi_{V_{\mathbf{1}}}(\boldsymbol{X}^\top \boldsymbol{Y})$
- **For** $t = 2$ **to** $\mathrm{rank}\,(\boldsymbol{X}) = r$:
  - $\theta^t = \frac{R^{t-\mathbf{1}}}{\|R^{t-\mathbf{1}}\|}$ and $h^t = \boldsymbol{X}\theta^t$
  - $\Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|} \right\rangle \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|}$
  - $R^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$

PLS-1 algorithm: Summary

Pseudo-code

Input: $\boldsymbol{X}$, $\boldsymbol{Y}$

- **Initialization** $t = 0$:
$$R^0 = \boldsymbol{X}^\top \boldsymbol{Y}$$
- **Step** $t = 1$:
  - $\theta^1 = \frac{R^0}{\|R^0\|}$ and $h^1 = \boldsymbol{X}\theta^1$
  - $\Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|} \right\rangle \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|}$
  - $R^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$
- **For** $t = 2$ **to** $\mathrm{rank}(\boldsymbol{X}) = r$:
  - $\theta^t = \frac{R^{t-1}}{\|R^{t-1}\|}$ and $h^t = \boldsymbol{X}\theta^t$
  - $\Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|} \right\rangle \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|}$
  - $R^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$

Output: $\left(h^1, h^2, \ldots, h^r\right)$ such that $\quad (r = \mathrm{rank}(\boldsymbol{X}))$

$$\widetilde{\boldsymbol{X}} = \left[\, h^1; h^2; \ldots; h^r \,\right] = \boldsymbol{X} \cdot \left[\, \theta^1; \theta^2; \ldots; \theta^r \,\right]$$

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante
- The $\boldsymbol{X}_{\cdot j}$s have been normalized (centering and reducing)

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante
- The $\boldsymbol{X}_{\cdot j}$s have been normalized (centering and reducing)
- By contrast, the h^js (PLS components):

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante
- The $\boldsymbol{X}_{\cdot j}$s have been normalized (centering and reducing)
- By contrast, the h^js (PLS components):
  - are an orthogonal basis

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante
- The $\boldsymbol{X}_{\cdot j}$s have been normalized (centering and reducing)
- By contrast, the h^js (PLS components):
  - are an orthogonal basis
  - are not unitary vectors! (unlike the $\boldsymbol{X}_{\cdot j}$s)

Comments on PLS

Covariance and not correlation

- Covariance is not scale invariante
- The $\boldsymbol{X}_{\cdot j}$s have been normalized (centering and reducing)
- By contrast, the h^js (PLS components):
  - are an orthogonal basis
  - are not unitary vectors! (unlike the $\boldsymbol{X}_{\cdot j}$s)
- Can induce some bias in the vizualization

Canonical Correlation Analysis (CCA)

Supervised situation

Notation

- n individuals
- Each individual i described by d (quantitative) variables $X^1, \ldots, X^d$
- $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$
- Label:

  Each individual i described by Y_i
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$

Supervised situation

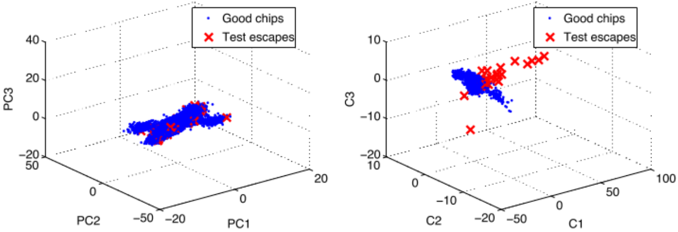
Notation

- n individuals
- Each individual i described by d (quantitative) variables $X^1, \ldots, X^d$
- $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
- $\boldsymbol{X}$: $n \times d$ "full" design matrix
- $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$ (n measurements for variable X^j)
- $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$
- Label:

  Each individual i described by Y_i
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$

Unlike PCA, the CCA framework is a supervised one!

CCA algorithm



3-D Principal Component Space

3-D Canonical Space

CCA-1 algorithm: Step 1

$$\widehat{\mathrm{Corr}}\left(\boldsymbol{X}\theta, \boldsymbol{Y}\right) = \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|}$$

(columns of $\boldsymbol{X}$ centered and reduced)
(vector $\boldsymbol{Y}$ is centered)

CCA-1 algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\theta \in \mathbb{R}^d} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first CCA component

Remark:

- Focuses on maximizing correlation!
- h^1 maximizes the correlation with $\boldsymbol{Y}$ (explicability)

CCA-1 algorithm: Step 2

CCA-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}\,(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j = 1, \ldots, t-1}} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth CCA component

CCA-1 algorithm: Step 2

CCA-1 algorithm: Step t ($2 \leq t \leq r = \mathrm{rank}(\boldsymbol{X})$)

- Find

$$\theta^t \in Arg \max_{\substack{\langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ j=1,\ldots,t-1}} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\}$$

- $h^t = \widetilde{\boldsymbol{X}}_{\cdot t} = \boldsymbol{X}\theta^t \in \mathbb{R}^n$: tth CCA component

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, (\boldsymbol{X}^\top \boldsymbol{X})\, \theta^j \rangle = 0$ for all $2 \leq j \leq d$
- Any new component must be $\perp$ previous ones
- $(h^1, \ldots, h^r)$: orthonormal family
- Each individual i described by:

$$(h_i^1, h_i^2) \in \mathbb{R}^2 \text{ instead of } \boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$$

(with two components)

Revisiting CCA algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\theta \in \mathbb{R}^d} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$

Calculating the solution

Revisiting CCA algorithm: Step 1

- Find

$$\theta^1 \in Arg \max_{\theta \in \mathbb{R}^d} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$

Calculating the solution

$$\begin{aligned} & \theta^1 \in Arg \max_{\theta \in \mathbb{R}^d} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|} \right\} \\ \Leftrightarrow \quad & \theta^1 \in Arg \max_{\|\boldsymbol{X}\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle \right\} \\ \Leftrightarrow \quad & \theta^1 \in Arg \max_{\|\boldsymbol{X}\theta\|=1} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle \right\} \\ \Leftrightarrow \quad & \theta^1 = \frac{\boldsymbol{X}^\top \boldsymbol{Y}}{\|\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\|} \end{aligned}$$

$\longrightarrow$ The first CCA component is $h^1 = \boldsymbol{X}\theta^1 = \dfrac{\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}}{\|\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\|}$

Revisiting CCA algorithm: Step 2 ($t = 2$)

Orthogonality constraint $\left(V_1 = \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right)\right)$

$$\left\langle \boldsymbol{X}\theta, h^1 \right\rangle = 0 \quad \Leftrightarrow \quad \left\langle \theta, \boldsymbol{X}^\top h^1 \right\rangle = 0$$

$$\Leftrightarrow \quad \theta \in \mathrm{Vect}\left(\boldsymbol{X}^\top h^1\right)^\perp = V_1^\perp$$

Constrained maximization

Under this orthogonality constraint, solving

$$\max_{\|\boldsymbol{X}\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle \right\} \quad \Leftrightarrow \quad \max_{\|\boldsymbol{X}\theta\|=1} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\}$$

where $\Pi_{V_1}(\cdot)$ denotes the orthogonal projection onto V_1

Reminder:

$$u - \Pi_{V_1}(u) \in V_1^\perp, \qquad \forall u \in \mathbb{R}^d$$

Revisiting CCA algorithm: Step 2 ($t = 2$)

Solution at $t = 2$

$$\theta^2 = Arg \max_{\substack{\|\boldsymbol{X}\theta\| = 1 \\ \langle \theta, \boldsymbol{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^2 = Arg \max_{\substack{\|\boldsymbol{X}\theta\| = 1 \\ \langle \theta, \boldsymbol{X}^\top h^1 \rangle = 0}} \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^2 = \frac{\boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})}{\left\| \boldsymbol{X} \left(\boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) \right) \right\|}$$

Reminder:

$$\Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|} \right\rangle \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|}$$

CCA-1 algorithm: Summary

Pseudo-code

Input: $\boldsymbol{X}$, $\boldsymbol{Y}$

- **Initialization** $t = 0$:
$$R^0 = \boldsymbol{X}^\top \boldsymbol{Y}$$
- **Step** $t = 1$:
  - $\theta^1 = \frac{R^0}{\|\boldsymbol{X} R^0\|}$ and $h^1 = \boldsymbol{X}\theta^1$
  - $\Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|} \right\rangle \frac{\boldsymbol{X}^\top h^1}{\|\boldsymbol{X}^\top h^1\|}$
  - $R^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$
- **For** $t = 2$ **to** $\operatorname{rank}(\boldsymbol{X}) = r$:
  - $\theta^t = \frac{R^{t-1}}{\|\boldsymbol{X} R^{t-1}\|}$ and $h^t = \boldsymbol{X}\theta^t$
  - $\Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top \boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|} \right\rangle \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|}$
  - $R^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$

Output: $\left(h^1, h^2, \ldots, h^r\right)$ such that $\qquad (r = \operatorname{rank}(\boldsymbol{X}))$

$$\widetilde{\boldsymbol{X}} = \left[\, h^1; h^2; \ldots; h^r \,\right] = \boldsymbol{X} \cdot \left[\, \theta^1; \theta^2; \ldots; \theta^r \,\right]$$

Extension to two design matrices

Unsupervised situation

Notation

- n individuals
- First design matrix $\boldsymbol{X}$:
  - $X_i = (X_i^1, \ldots, X_i^d)^\top \in \mathbb{R}^d$: features for individual i
  - $\boldsymbol{X}$: $n \times d$ "full" design matrix
  - $\boldsymbol{X}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{X}$
  - $\boldsymbol{X}_{i\cdot} \in \mathbb{R}^d$: ith row vector of $\boldsymbol{X}$
- Second design matrix $\boldsymbol{W}$:
  - $W_i = (W_i^1, \ldots, W_i^d)^\top \in \mathbb{R}^p$: features for individual i
  - $\boldsymbol{W}$: $n \times p$ "full" design matrix
  - $\boldsymbol{W}_{\cdot,j} \in \mathbb{R}^n$: jth column vector of $\boldsymbol{W}$
  - $\boldsymbol{W}_{i\cdot} \in \mathbb{R}^p$: ith row vector of $\boldsymbol{W}$

Examples for $\boldsymbol{W}$ and $\boldsymbol{X}$

- Results from two different experiments
  (Two psychologic tests, bio. exper. at two time-stamps)
- Wine critic ratings and wine biological features

CCA algorithm: Step 1

$$\widehat{\mathrm{Corr}\,(\boldsymbol{X}\theta, \boldsymbol{W}\gamma)} = \frac{\langle \boldsymbol{X}\theta, \boldsymbol{W}\gamma \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{W}\gamma\|}$$

(columns of $\boldsymbol{X}$, $\boldsymbol{W}$ centered and reduced)

CCA algorithm: Step 1

- Find

$$\theta^1, \gamma^1 \in Arg \max_{\theta \in \mathbb{R}^d, \gamma \in \mathbb{R}^p} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{W}\gamma \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{W}\gamma\|} \right\}$$

- $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$: first CCA component for $\boldsymbol{X}$
- $g^1 = \widetilde{\boldsymbol{W}}_{\cdot 1} = \boldsymbol{W}\gamma^1 \in \mathbb{R}^n$: first CCA component for $\boldsymbol{W}$

CCA algorithm: Step 2

CCA algorithm: Step t $(2 \leq t \leq \min(\mathrm{rank}(\boldsymbol{X}), \mathrm{rank}(\boldsymbol{W})))$

- Find

$$\theta^t, \gamma^t \in Arg \max_{\substack{\langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ \langle \boldsymbol{W}\gamma, g^j \rangle = 0 \\ j = 1, \ldots, t-1}} \left\{ \frac{\langle \boldsymbol{X}\theta, \boldsymbol{W}\gamma \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{W}\gamma\|} \right\}$$

- $r = \min(\mathrm{rank}(\boldsymbol{X}), \mathrm{rank}(\boldsymbol{W}))$

Interpretation

- $\langle \boldsymbol{X}\theta, h^j \rangle = \langle \theta, (\boldsymbol{X}^\top \boldsymbol{X}) \theta^j \rangle = 0$ for all $2 \leq j \leq t-1$
- $(h^1, \ldots, h^r)$ and $(g^1, \ldots, g^r)$: orthonormal families since the above problem is equivalent to

$$\max_{\substack{\|\boldsymbol{X}\theta\| = 1,\ \|\boldsymbol{W}\gamma\| = 1 \\ \langle \boldsymbol{X}\theta, h^j \rangle = 0 \\ \langle \boldsymbol{W}\gamma, g^j \rangle = 0 \\ j = 1, \ldots, t-1}} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{W}\gamma \rangle \right\}$$

CCA implementation

Summary

- Impelmentations similar to those of PLS and CCA-1 do exist for efficiently solving the CCA optimization problem
- CCA is used in a lot of applications for combining different types of information ("tables") used for describing "individuals"
- Extensions of CCA do exist for multiple tables (see Generalized CCA. . .)