# Canonical Correlation Analysis

[HORVATH - KOKOSZKA CH]

General Idea:

DATA $(X_1, X_2, \ldots, X_N)$ $X_i \in \mathbb{R}^d$,
$(Y_1, Y_2, \ldots, Y_N)$ $Y_i \in \mathbb{R}$

Find linear functional $a, b$ of $X_i$, $Y_j$ respectively

$$\underline{A}_1 = (a \cdot X_1, \; a \cdot X_2, \; a \cdot X_3, \ldots, \; a \cdot X_N)$$

$$\underline{B}_1 \; (b \cdot Y_1, \; b \cdot Y_2, \ldots, \; b \cdot Y_N)$$

s.t. $\underline{A}$ & $\underline{B}$ maximally correlated

$a, b$ weight functions

<u>Step 2</u>: Find $a_2, b_2$ s.t. $a_1 \perp a_2$

$b_1 \perp b_2$

and

$A_2 = (a_2 \cdot X_1, \ldots, a_2 \cdot X_n)$

is max correlated to

$B_2 = (b_2 \cdot Y_1, \ldots, b_2 \cdot Y_n)$

- CCA in finite Dimensions. ($X_i \in \mathbb{R}^d$, $Y_j \in \mathbb{R}^l$)

$$\begin{cases} C_{11} = \operatorname{Cov}(X) \in \mathbb{R}^{d \times d} \\ C_{12} = \operatorname{Cov}(X, Y) \in \mathbb{R}^{d \times l} \\ C_{22} = \operatorname{Cov}(Y, Y) \in \mathbb{R}^{l \times l} \end{cases}$$

<u>Objective</u>: Want to find $a, b \in \mathbb{R}^d, \mathbb{R}^l$

s.t.

$$a, b = \operatorname{argmax} \underbrace{\langle a, C_{12} b \rangle}_{\operatorname{Cov}(a \cdot X,\ b \cdot Y)}$$

subject to $\langle a, C_{11} a \rangle = 1$

$\langle b, C_{22} b \rangle = 1$

Want to solve the following optimisation problem

$$\mathcal{L} = \langle a, C_{12} b\rangle - \frac{\rho_1}{2}\left(\langle a, C_{11} a\rangle - 1\right) - \frac{\rho_2}{2}\left(\langle b, C_{22} b\rangle - 1\right)$$

$C_{21} = C_{12}^*$

Derivative wrt a:

$$C_{12} b - \rho_1 C_{11} a = 0$$

b:

$$\underline{C_{21}}\, a - \rho_2 C_{22} b = 0$$

Solve for the $\rho_i$:

$$\langle a, C_{12} b\rangle = \rho_1 \quad \text{and} \quad \langle b, \underbrace{C_{21}}_{C_{12}^*} a\rangle = \rho_2$$

$$\Rightarrow \quad \rho_1 = \rho_2 =: \rho$$

Then

$$a = \frac{C_{11}^{-1} C_{12} b}{\rho}$$

$$\frac{C_{21}(C_{11}^{-1}C_{12}b)}{\rho} = \rho C_{22}b$$

$$\rho^2 b = C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} b$$

i.e. $b$ is sol$^n$ to eigenvalue problem for $\lambda = \rho^2$

Write $b = C_{22}^{-\frac{1}{2}} f$ then

$$\rho^2 f = C_{22}^{-\frac{1}{2}} C_{12}^{*} C_{11}^{-1} C_{12} C_{22}^{-\frac{1}{2}} f$$

$$= \underbrace{\left(C_{22}^{-\frac{1}{2}} C_{12}^{*} C_{11}^{-\frac{1}{2}}\right)}_{R^*} \underbrace{\left(C_{11}^{-\frac{1}{2}} C_{12} C_{22}^{-\frac{1}{2}}\right)}_{R} f$$

$\therefore$ $R^*R$ has non-negative eigenvalues

Then the weights are $b = C_{22}^{-\frac{1}{2}} f$

Similarly writing $a = C_{11}^{-\frac{1}{2}} e$ we have a similar eigenvalue problem for $e$:

$$\lambda e = RR^* e$$

Then $(a_1, b_1), (a_2, b_2)$ etc can obtained from the ~~basis~~ eigenbasis $e_1, e_2, e_3, \ldots$
$f_1, f_2, f_3, \ldots$

In Function Space: ( $X$, & $Y$ are mean $0$ ]

$$c_{11}(s,t) = \mathbb{E}[X(t)X(s)]$$

$$c_{22}(s,t) = \mathbb{E}[Y(t)Y(s)]$$

$$c_{12}(s,t) = \mathbb{E}[X(t)Y(s)].$$

$$(C_{11}x)(t) = \int c_{11}(s,t)\,x(s)\,ds$$

& $C_{12}$, $C_{22}$ defined analogously.

Idea: Replace covariance matrices with covariance operators

Q1: What is $C^{\frac{1}{2}}$?

$F = C^{\frac{1}{2}}$ if $FF = C$

Recall since $C$ is a covariance operator:

$$C = \sum_{i=1}^{\infty} \lambda_i \, e_i \otimes e_i$$

In particular

$$C^{\frac{1}{2}} = \sum_{i=1}^{\infty} \lambda_i^{\frac{1}{2}} \, e_i \otimes e_i .$$

e.g.

$$C^{\frac{1}{2}} C^{\frac{1}{2}} = \left( \sum_{i=1}^{\infty} \lambda_i^{\frac{1}{2}} \, e_i \otimes e_i \right) \left( \sum_{j} \lambda_j^{\frac{1}{2}} \, e_j \otimes e_j \right)$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_i^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} \underbrace{\langle e_i, e_j \rangle}_{\substack{= 1 \text{ iff } i = j \\ 0 \text{ otherwise}}} \, e_i \otimes e_j$$

$$= \sum_{i=1}^{\infty} \lambda_i \, e_i \otimes e_i = C$$

Assumptions: $C$ has tree eigenvalues $\to 0$

Formally:

$$C^{-1} = \sum_{i=1}^{\infty} \lambda_i^{-1} e_i \otimes e_i \quad \text{on } H$$

But take $\bullet\ x = \sum \underline{\lambda_i^{\frac{1}{2}} e_i}$

then $\|x\|^2 = \sum \lambda_i < \infty$

$\therefore x \in L^2$

But

$$C^{-\frac{1}{2}} x = \sum e_i$$

$$\|C^{-1/2} x\|^2 = \sum_{i=1}^{\infty} 1 = \infty$$

$\therefore C^{-1/2} x \notin H$. $\quad$ $C^{-1}$ is not invertible on $H$

However: if $x \in \text{Range}\, C^{\frac{1}{2}} = \text{Im}\, C^{\frac{1}{2}}$

then $C^{-\frac{1}{2}} x \in H$.

i.e. $\text{Domain}(C^{-\frac{1}{2}}) = \text{Im}\, C^{\frac{1}{2}}$.

For this to make sense

$$\boxed{\underbrace{\text{Im}\left[C_{12}\right]}_{C_{12}(H_2)} \subseteq \text{Im}\, C_{11}^{\frac{1}{2}}}$$

smoothing operator

$$\underbrace{\text{Im}\left[C_{21}\right]}_{C_{21}(H_1)} \subseteq \text{Im}\, C_{22}^{\frac{1}{2}}$$

Let's express

$$X = \sum_{i=1}^{\infty} x_i e_i \,, \qquad x_i = \langle X, e_i \rangle$$

$$Y = \sum_{j=1} y_i f_i \,, \quad y_j = \langle Y, f_j \rangle$$

$e_i$ : eigenfunctions of covariance operator $C_{11}$

$f_i$ : " — " $C_{22}$.

$$\begin{cases} C_{11} e_i = \lambda_i e_i \\ C_{22} f_i = \gamma_i f_i \end{cases}$$

Define the correlation coefficients :

$$\sigma_{ji} = \frac{\mathbb{E}[x_i y_j]}{\sqrt{\mathbb{E} x_i^2 \, \mathbb{E} y_j^2}} = \frac{\mathbb{E}[x_i y_j]}{\lambda_i^{1/2} \gamma_j^{1/2}}$$

Prop : Condition ⊛⊛ holds if

$$\sum_{i,j=1}^{\infty} \sigma_{ji}^2 < \infty$$

If condition holds then can define

$$R = C_{11}^{-\frac{1}{2}} C_{12} C_{22}^{-\frac{1}{2}} : \operatorname{Im}(C_{22}^{\frac{1}{2}}) \to H_1$$

Note: $R$ and $R^*$ are Hilbert-Schmidt operators.

Define $\underline{M} = R^* R : \operatorname{Im}(C_{22}^{\frac{1}{2}}) \to \operatorname{Im}(C_{22}^{\frac{1}{2}})$

is ① symmetric
② positive definite.
③ Hilbert-Schmidt.

By the HS theorem:

$$My = \sum_{k=1}^{\infty} \rho_k^2 \langle y, \varphi_k \rangle \varphi_k.$$

Then $\varphi_1, \varphi_2, \varphi_3$ are the eigenbasis of $M$

**Recall:** $b_i = C_{22}^{-\frac{1}{2}} \varphi_i$

$$a_i = C_{11}^{-\frac{1}{2}} e_i$$

To get canonical components I need

$$\boxed{\begin{aligned} f_i &\in \operatorname{Im} C_{22}^{\frac{1}{2}} \\ e_j &\in \operatorname{Im} C_{11}^{\frac{1}{2}} . \end{aligned}}$$

<u>Proposition</u>: Under the conditions of the previous proposition

$$\boxed{\begin{gathered} \sum_{i,j=1}^{\infty} \lambda_i^{-\frac{1}{2}} \sigma_{ji}^{2} < \infty \quad \text{and} \\ \sum_{i,j=1}^{\infty} \gamma_i^{-1} \sigma_{ji}^{2} < \infty \end{gathered}}$$

then we can define the weight functions

- $a_k = C_{11}^{-\frac{1}{2}} e_k \in H_1$
- $b_k = C_{22}^{-\frac{1}{2}} f_k \in H_2$

where $\{e_i\}$ are the eigenfunctions of $RR^*$

$\{f_j\}$ " " $R^*R$.

---

In particular:

- $\langle a_k, C_{11} a_k \rangle = \langle b_k, C_{22} b_k \rangle = 1$
- $\langle a_k, C_{12} b_k \rangle = \rho_k$

which is the $k^{th}$ eigenvalue of

$$\rho_k e_k = RR^* e_k$$

or $\rho_k f_k = R^*R f_k$.

and this correlation is maximal.

Finally:

$\forall\ j \neq k$,

$$\boxed{\langle a_j, X\rangle,\ \langle b_j, Y\rangle \text{ are uncorrelated.}}$$

---

In practice we observe pairs of curves
$(X_1, Y_1), \ldots, (X_N, Y_N)$

Original Problem:

Find $a, b$ s.t.

$\langle a, C_{12} b\rangle$ is maximised s.t.

- $\langle a, C_{11} a\rangle = 1$
- $\langle b, C_{22} b\rangle = 1$

equiv to maximising

$$\frac{\langle a, C_{12} b\rangle^2}{\underbrace{\langle a, (C_{11}+\varepsilon_1 I) a\rangle}_{\geq \varepsilon_1 \|a\|^2} \underbrace{\langle b, (C_{22}+\varepsilon_2 I) b\rangle}_{\geq \varepsilon_2 \|b\|^2}}$$

<u>Idea</u>: Introduce regularized covariance operators

$$C_{11,\varepsilon_1} = C_{11} + \varepsilon I$$

$$C_{22,\varepsilon_2} = C_{22} + \varepsilon I.$$

$$\|C_{11,\varepsilon_1}^{-\frac{1}{2}} x\| \leq \frac{1}{\sqrt{\varepsilon_1}} \|x\|,$$

- $\operatorname{Im} C_{11,\varepsilon}^{\frac{1}{2}} = H_1$

The condition that $\operatorname{Im} C_{12} \subseteq \operatorname{Im} C_{11,\varepsilon}^{\frac{1}{2}}$ holds trivially.

$$R_\varepsilon = C_{11,\varepsilon}^{-\frac{1}{2}} C_{12} C_{22,\varepsilon}^{-\frac{1}{2}} : H_2 \to H_1$$

$$M_\varepsilon = R_\varepsilon^* R_\varepsilon$$

---

## Two Sample Inference for Functional Data.

$$X_1, X_2, \dots, X_{n_1} \sim D_1$$

$$Y_1, Y_2, \dots, Y_m \sim D_2$$

$$H_0 : \boxed{D_1 = D_2 \, ?}$$

$$E[\varepsilon_i] = E[\varepsilon_i^*] = 0$$

$$X_i(t) = \mu(t) + \varepsilon_i(t) \qquad 1 \le i \le N$$

$$Y_j(t) = \mu^*(t) + \varepsilon_j^*(t) \qquad 1 \le j \le M$$

$$H_0 : \text{ is } \mu = \mu^*$$

Sample Means:

$$\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i \in L^2$$

$$\overline{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j \in L^2$$

$$U_{N,m} = \left(\frac{Nm}{N+m}\right)\left\| \overline{X}_N - \overline{Y}_m \right\|_{L^2}^2$$