Canonical Correlation Analysis (CCA)

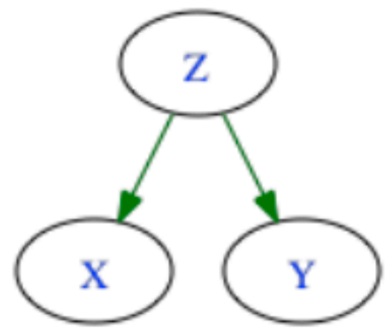
Lecture 2 - ML2 @ TUB Machine learning group

Mina Jamshidi @ TUB Machine learning group

2/May/2024

CCA

Latent variable Z is measured in multivariate datasets X and Y



$$X \in \mathbb{R}^M, Y \in \mathbb{R}^N$$

Which representation of X and Y reflects Z best?

CCA: That representation that maximises the correlation between X and Y.

CCA

Given two (or more) multivariate variables

$$X \in \mathbb{R}^M, Y \in \mathbb{R}^N$$

CCA finds projections

$$w_x \in \mathbb{R}^M, w_y \in \mathbb{R}^N$$

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{E\left[\boldsymbol{x}\boldsymbol{y}^T\right]}{\sqrt{E\left[\boldsymbol{x}\boldsymbol{x}^T\right] E\left[\boldsymbol{y}\boldsymbol{y}^T\right]}}$$

$$= \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\left[\boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{Y}^T \boldsymbol{w}_y\right]}{\sqrt{\left[\boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}_x\right] \left[\boldsymbol{w}_y^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}_y\right]}}$$

CCA

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{E\left[\boldsymbol{x}\boldsymbol{y}^T\right]}{\sqrt{E\left[\boldsymbol{x}\boldsymbol{x}^T\right] E\left[\boldsymbol{y}\boldsymbol{y}^T\right]}}$$

$$= \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\left[\boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{Y}^T \boldsymbol{w}_y\right]}{\sqrt{\left[\boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}_x\right] \left[\boldsymbol{w}_y^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}_y\right]}}$$

$$\underset{w_x, w_y}{\operatorname{argmax}} \left(w_x^\top X Y^\top w_y\right) \quad \text{s.t.} \quad \begin{array}{c} w_x^\top X X^\top w_x = 1 \\ w_y^\top Y Y^\top w_y = 1 \end{array}$$

[Hotelling, Biometrika, 1936]

CCA

Assuming centered data $$\sum_i x_i = \sum_i y_i = 0$$

We can compute empirical
cross-covariance matrices
and auto-covariance matrices

$$C_{xy} = \frac{1}{N} XY^\top$$
$$C_{xx} = \frac{1}{N} XX^\top$$

CCA

CCA objective

$$\underset{w_x, w_y}{\operatorname{argmax}} \left(w_x^\top X Y^\top w_y \right) \quad \text{s.t.} \quad \begin{array}{c} w_x^\top X X^\top w_x = 1 \\ w_y^\top Y Y^\top w_y = 1 \end{array}$$

Lagrangian

$$\mathcal{L} = w_x^\top C_{xy} w_y - \frac{1}{2}\alpha(w_x^\top C_{xx} w_x - 1) - \frac{1}{2}\beta(w_y^\top C_{yy} w_y - 1)$$

Partial Derivatives

$$\frac{\partial \mathcal{L}}{\partial w_x^\top} = C_{xy} w_y - \alpha C_{xx} w_x \qquad \frac{\partial \mathcal{L}}{\partial w_y^\top} = C_{yx} w_x - \beta C_{yy} w_y$$

CCA

Given $\alpha = \beta$

the partial derivatives become

$$C_{xy} w_y = \alpha \; C_{xx} w_x$$
$$C_{yx} w_x = \alpha \; C_{yy} w_y$$

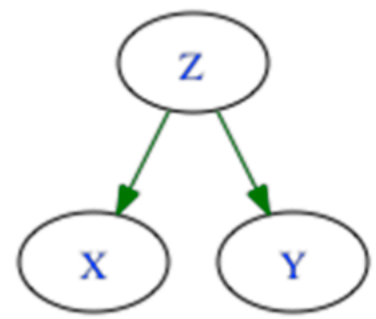
We can now reformulate these equations in block matrix form

$$\left[\begin{array}{cc} 0 & C_{xy} \\ C_{yx} & 0 \end{array} \right] \left[\begin{array}{c} w_x \\ w_y \end{array} \right] = \alpha \left[\begin{array}{cc} C_{xx} & 0 \\ 0 & C_{yy} \end{array} \right] \left[\begin{array}{c} w_x \\ w_y \end{array} \right]$$

which is just a *generalised eigenvalue equation*

Example

- Latent Variable **Z: Car Types**
- Measurements
- **X**: Displacement, Horsepower, Weight
- **Y**: Acceleration, Miles/Gallon



Dimensions

Car models

Displacement	Horsepower	Weight
307	130	3504
350	165	3693
318	150	3436

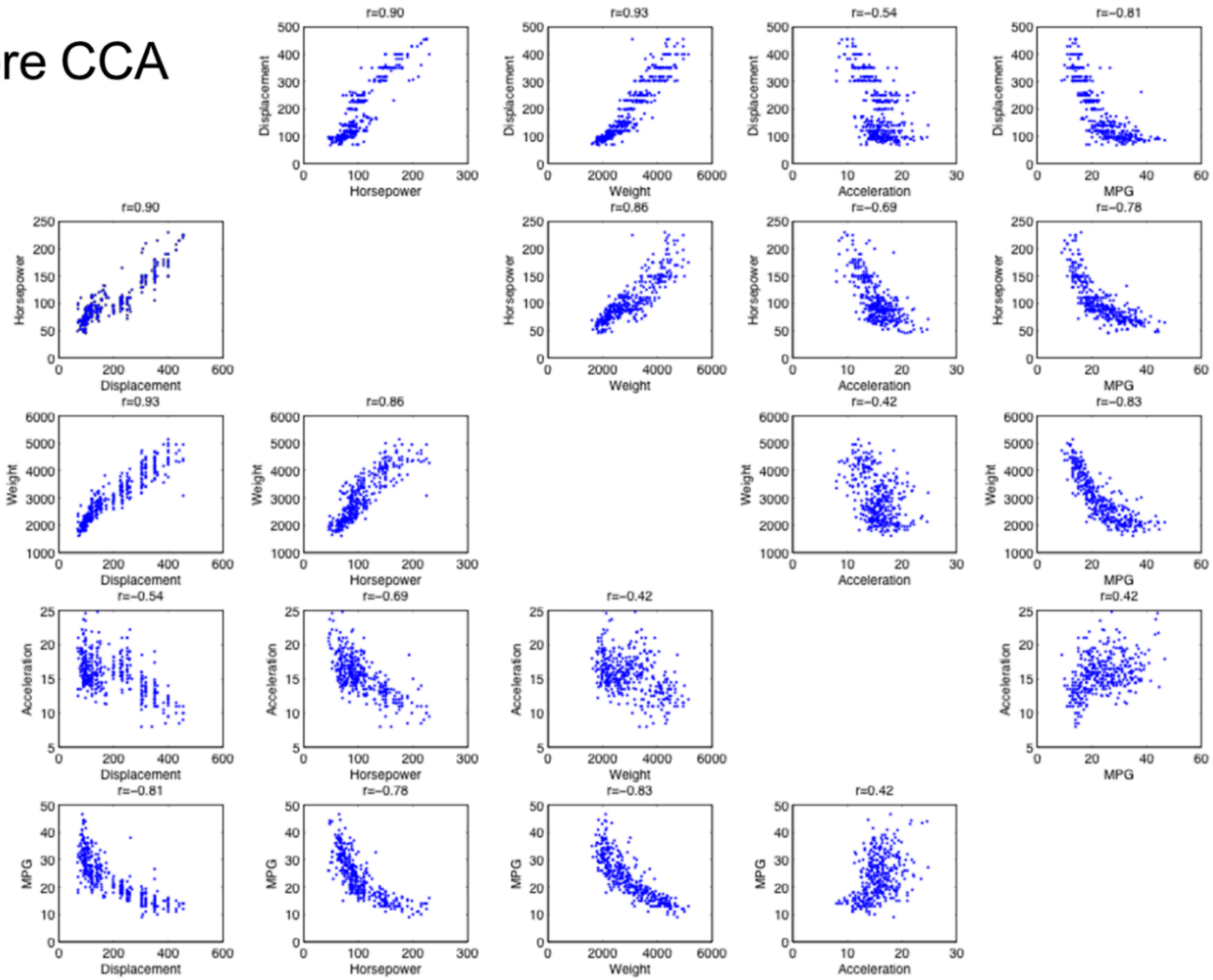
= X

Y =

Acceleration	Miles/Gallon
12	18
11	15
11	18

Example

Before CCA



Example

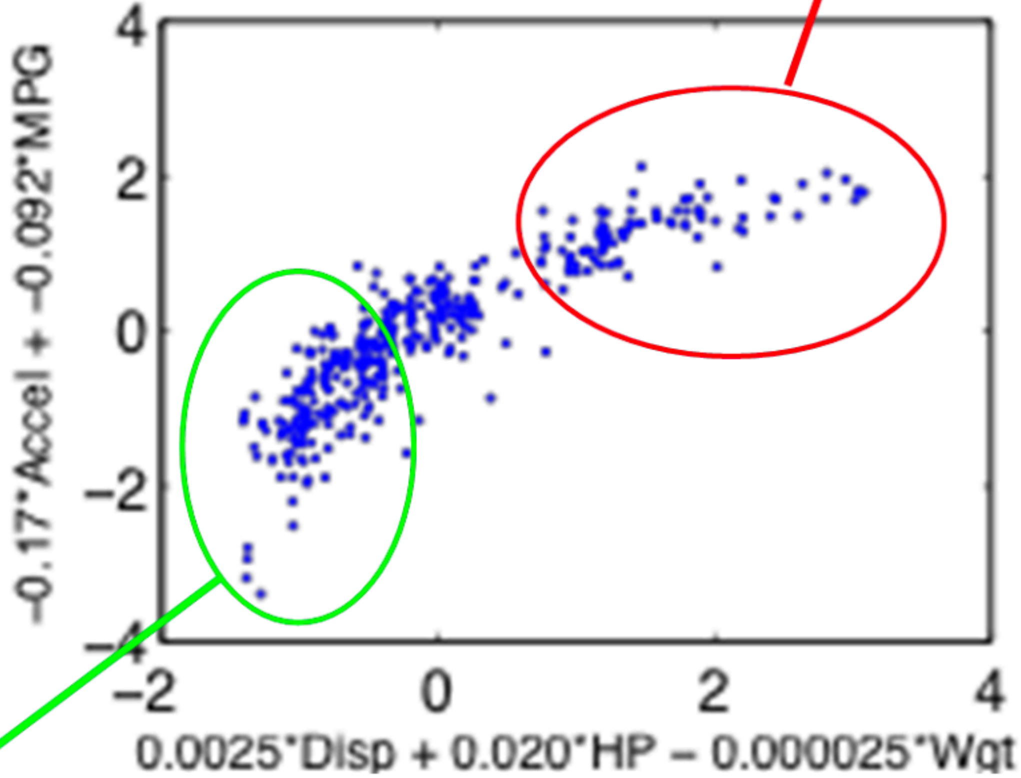
Latent Variable
Z: Car Types
Measurements
X: Displacement, Horsepower, Weight
Y: Acceleration, Miles/Gallon

$$w_x = \begin{bmatrix} 0.0025 \\ 0.0202 \\ -0.000025 \end{bmatrix}$$

$$w_y = \begin{bmatrix} -0.17 \\ -0.092 \end{bmatrix}$$

Yw_y

Sport Cars
(big engine, strong, thirsty)

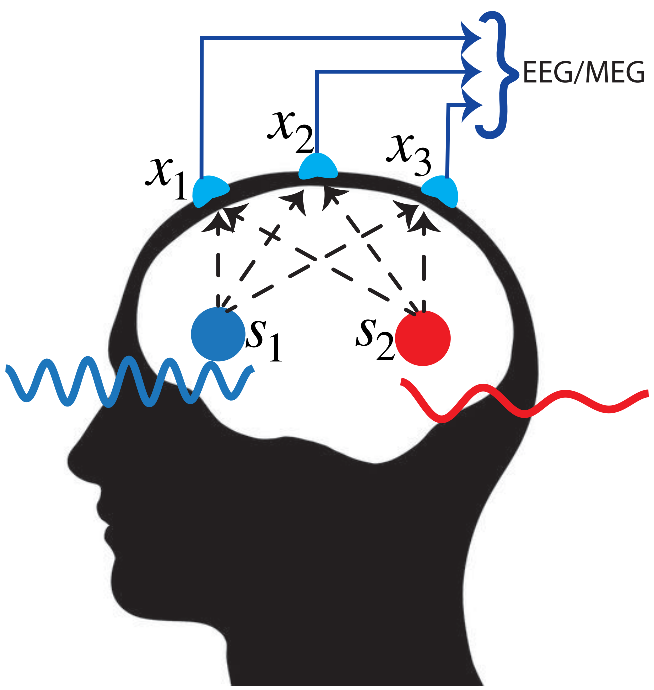


Xw_x

Commercial Cars
(small engine, low consumption)

Linear Generative Model in EEG/MEG

- $\mathbf{X}(t) \in \mathbb{R}^{C \times T}$ recorded signal with C channels
- $\mathbf{S}(t) \in \mathbb{R}^{M \times T}$ is a matrix containing C signals

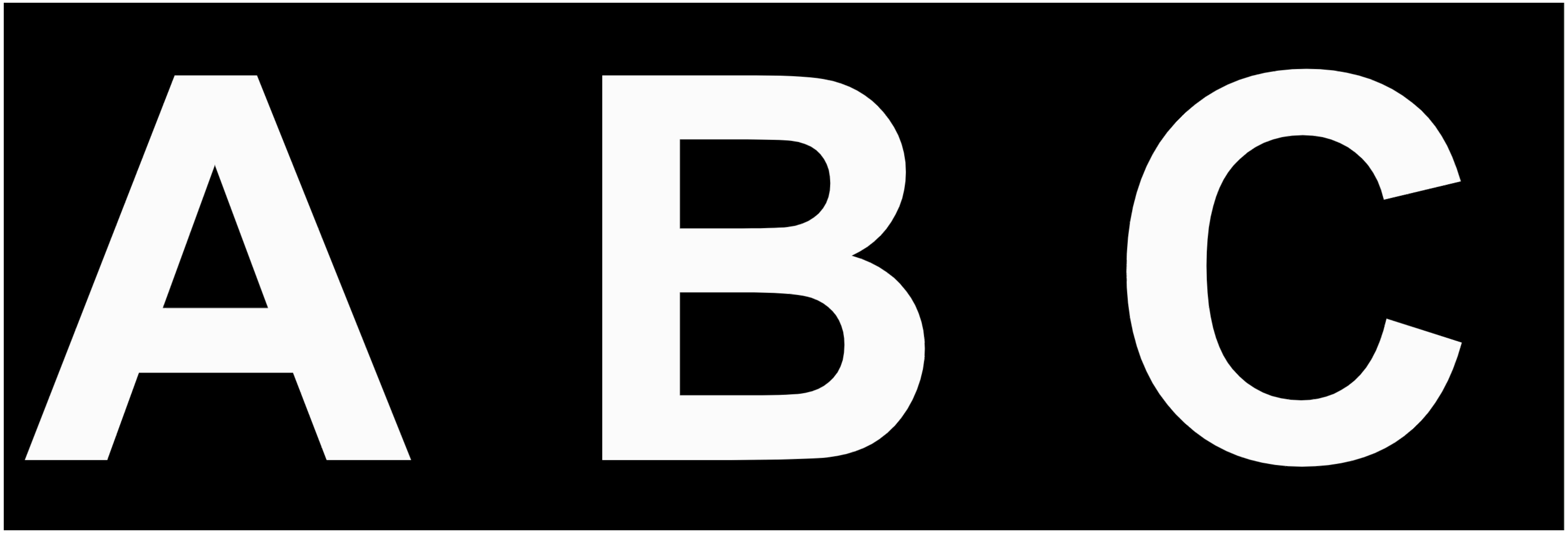


$$\mathbf{X}(t) = \mathbf{A}\mathbf{S}(t)$$ Forward model

$$\hat{\mathbf{s}}(t) = \mathbf{w}^T \mathbf{X}(t)$$ Backward model

source separation problem: only **X**(t) is known

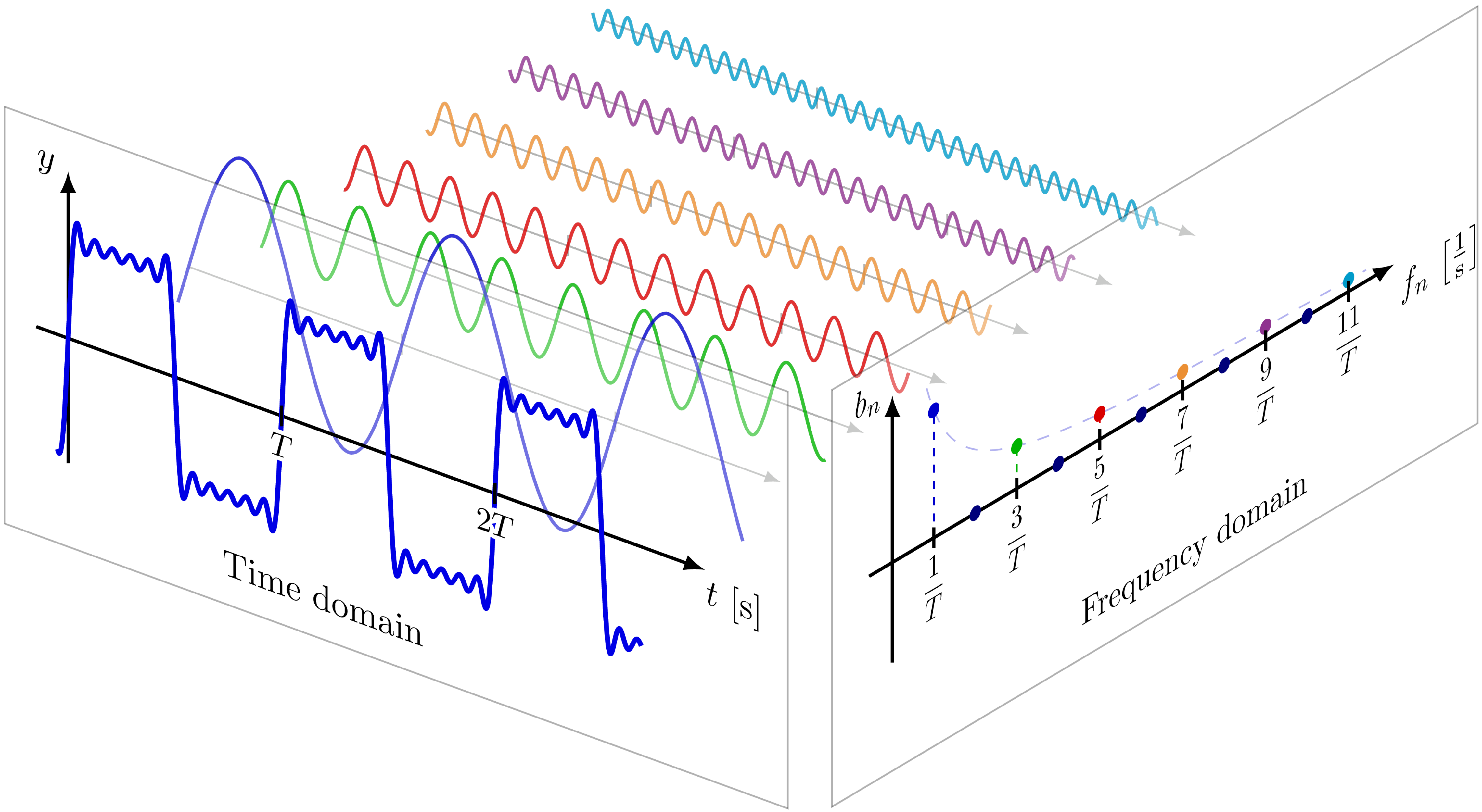
SSVEP BCI



- Flickering objects evoke brain signals at the flicker frequency

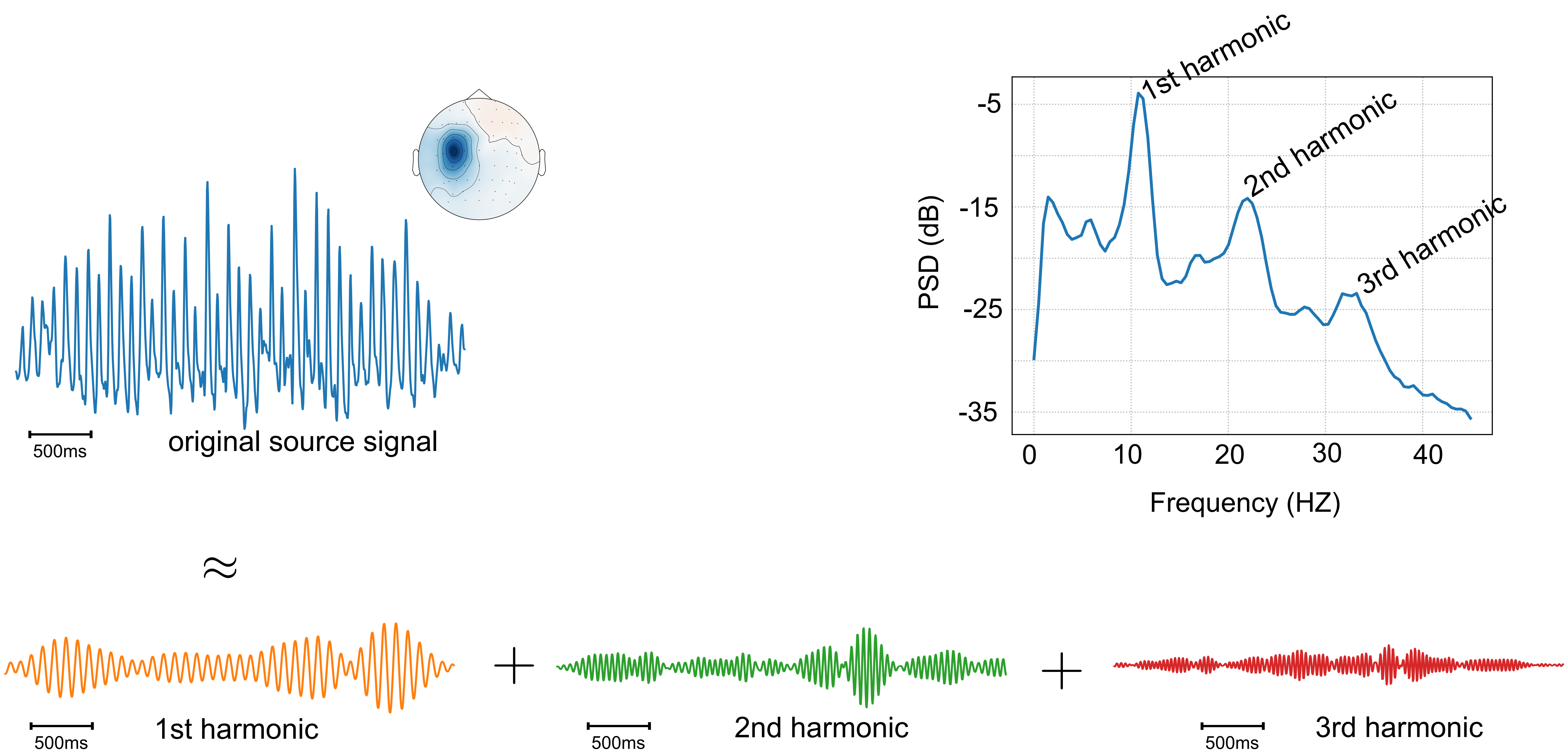
(c) video recorded from https://omids.github.io/quickssvep/

Fourier decomposition of a signal



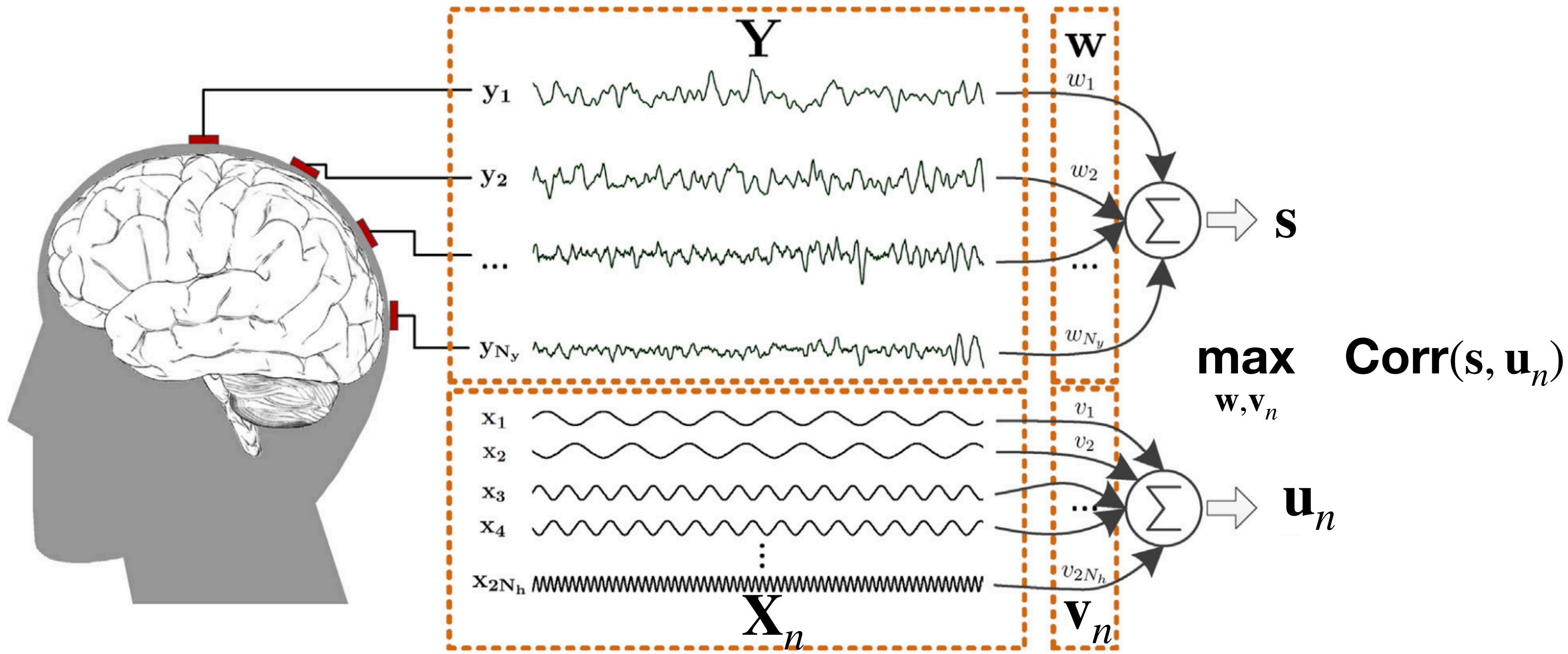
(c) figure from https://dibsmethodsmeetings.github.io/fourier-transforms/

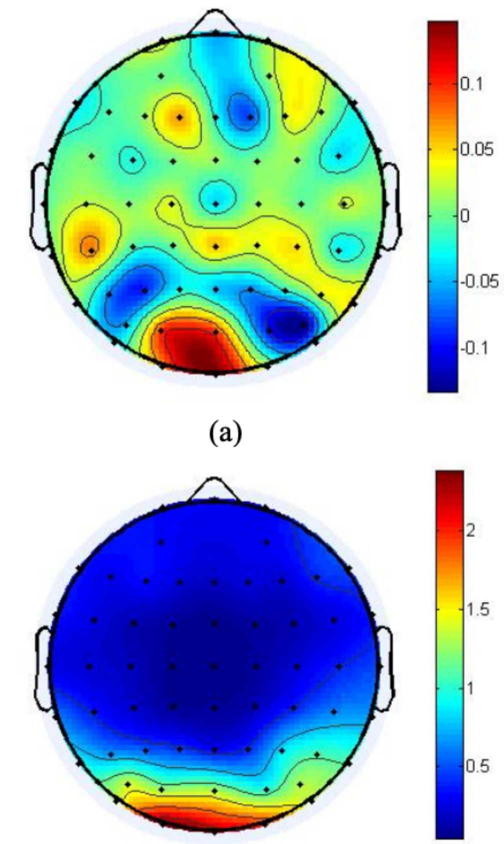
Non-sinusoidal signals in the brain



(c) figure from Jamshidi et al 2022

SSVEP BCI + CCA





(a)
(b)

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_C \end{bmatrix} \in \mathbb{R}^{C \times T}$$

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_c \end{bmatrix} \in \mathbb{R}^{C}$$

$$\mathbf{X}_n = \begin{bmatrix} \sin(2\pi f_n t) \\ \cos(2\pi f_n t) \\ \vdots \\ sin(2\pi N_h f_n t) \\ cos(2\pi N_h f_n t) \end{bmatrix}$$

$$\mathbf{v}_n = \begin{bmatrix} v_{n,1} \\ \vdots \\ v_{n,2N_h} \end{bmatrix} \in \mathbb{R}^{2N_h}$$

$$n_{flicker} = \underset{n}{\mathbf{argmax}} \quad \mathbf{Corr}(\mathbf{w}^T\mathbf{Y}, \mathbf{v}_n^T\mathbf{X}_n)$$

(c) left figure modified from Pryzala & Materka 2014

(c) right figure from Bin et al 2008

CCA is linear. What about non-linear case?

Kernel Trick

- What is a kernel function?

$$k : S \times S \to \mathbb{R}$$

- Given a nonlinear mapping ϕ to a high dimensional space:

$$k(x_i, x_j) = \left\langle \phi(x_i), \phi(x_j) \right\rangle$$

- Kernel matrix **K**:

$$\mathbf{K}_{ij} = k(x_i, x_j)$$

- An algorithm formulated in terms of an inner product
- replacing the inner product by a kernel function k.
- kernel PCA, kernel SVM

Kernel CCA

The solution of CCA in kernel space is obtained by solving the generalised eigenvalue problem

$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

The solutions in the input space can be recovered by

$$w_x = X\alpha_x$$
$$w_y = Y\alpha_y$$

Read the details: Akaho 2007

Summary

- CCA: a method for finding linear subspaces where the representations of two datasets is maximized
    - Source separation problem
    - SSVEP BCI + CCA
- Kernel CCA: the non-linear extension of CCA

Kernel CCA

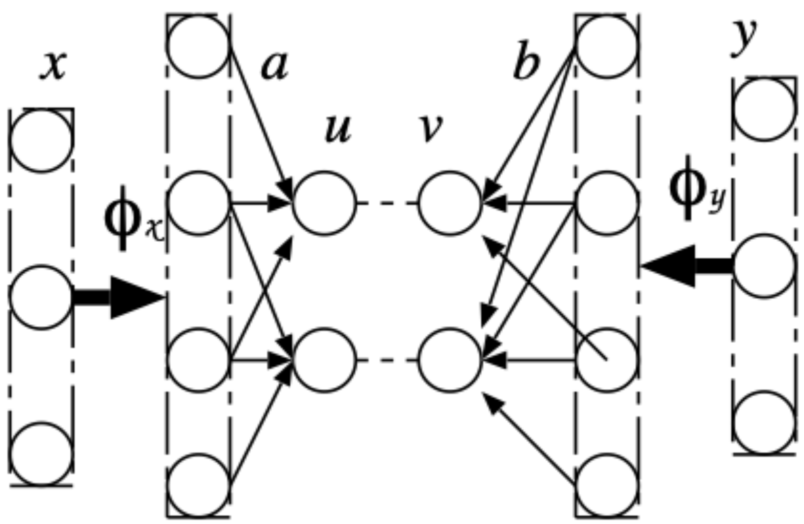


Figure 2: Kernel CCA

First, $\mathbf{x}$ and $\mathbf{y}$ are transformed into the Hilbert space, $\phi_x(\mathbf{x}) \in H_x$ and $\phi_y(\mathbf{y}) \in H_y$. By taking inner products with a parameter in the Hilbert spaces, $a \in H_x$, and $b \in H_y$, we find a feature

$$u = \langle a, \phi_x(\mathbf{x}) \rangle, \tag{5}$$

$$v = \langle b, \phi_y(\mathbf{y}) \rangle, \tag{6}$$

which maximizes the correlation coefficients.

$$\begin{aligned} \mathcal{L}_0 \quad &= \quad \mathrm{E}[(u - \mathrm{E}[u])(v - \mathrm{E}[v])] \\ &\quad -\frac{\lambda_1}{2}\mathrm{E}[(u - \mathrm{E}[u])^2] \\ &\quad -\frac{\lambda_2}{2}\mathrm{E}[(v - \mathrm{E}[v])^2]. \end{aligned} \tag{7}$$

$$\mathcal{L} = \mathcal{L}_0 + \frac{\eta}{2}(\|a\|^2 + \|b\|^2), \tag{8}$$

Kernel CCA

Now, from the condition that the derivative of $\mathcal{L}$ by a is equal to 0, we get

$$a = \sum_i \alpha_i \phi_x(\mathbf{x}_i), \tag{11}$$

where α_i is a schalar, then as a result, we have

$$u = \sum_i \alpha_i \langle \phi_x(\mathbf{x}_i), \phi_x(\mathbf{x}) \rangle. \tag{12}$$

Then, we obtain $\mathcal{L}$ by

$$\begin{aligned} \mathcal{L} &= \alpha^{\mathrm{T}} M \beta \\ &\quad - \frac{\lambda_1}{2} \alpha^{\mathrm{T}} L \alpha - \frac{\lambda_2}{2} \beta^{\mathrm{T}} N \beta \end{aligned} \tag{15}$$

where

$$M = \frac{1}{N} K_x^{\mathrm{T}} J K_y, \tag{16}$$

$$L = \frac{1}{N} K_x^{\mathrm{T}} J K_x + \eta_1 K_x, \tag{17}$$

$$N = \frac{1}{N} K_y^{\mathrm{T}} J K_y + \eta_2 K_y, \tag{18}$$

$$J = I - \frac{1}{N} \mathbf{1}\mathbf{1}^{\mathrm{T}}, \tag{19}$$

$$\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}, \tag{20}$$

and $\eta_1 = \eta/\lambda_1$, $\eta_2 = \eta/\lambda_2$.

If $\eta > 0$ is satisfied, L and N are positive definite almost surely, and we can show $\lambda_1 = \lambda_2 = \lambda$ from the constraint, then as a result we have a generalized eigenvalue problem for α, β

$$M\beta = \lambda L \alpha, \tag{21}$$

$$M^{\mathrm{T}} \alpha = \lambda N \beta, \tag{22}$$