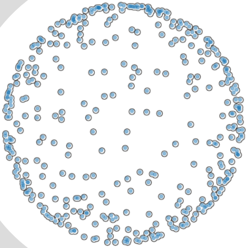
WiSe 2024/25

Machine Learning for Data Science

Lecture by G. Montavon

FREIE UNIVERSITÄT BERLIN



Lecture 6b

Correlation (cont.)

Outline

Deriving the CCA Eigenvalue Problem

- Recap: Lagrange multipliers
- Deriving the CCA eigenvalue problem
- CCA stability

CCA for High-Dimensional Data

- Solution in the span of the data
- Reformulation of CCA in terms of $N \times N$ matrices

Application 2

- Temporal CCA for neurovascular coupling

Part 1

Deriving the CCA Eigenvalue Problem

Recap: Lagrange Multipliers

General framework for finding solutions of constrained optimization problems of the type

$$\arg\max_{\theta} f(\theta) \qquad \text{subject to} \qquad g(\theta) = 0$$

- **Step 1:** Construct the 'Lagrangian':

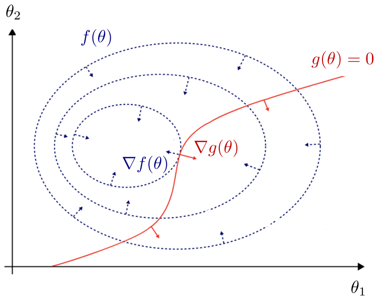
  $$\mathcal{L}(\theta, \lambda) = f(\theta) + \lambda \cdot g(\theta)$$

  where λ is called the Lagrange multiplier.

- **Step 2:** Solve the equation

  $$\nabla \mathcal{L}(\theta, \lambda) = 0$$

  which is a necessary condition for the solution.



Intuition for step 2: the equation includes the equation $\nabla f(\theta) = -\lambda \nabla g(\theta)$, i.e. the gradient of objective and constraint are aligned, but point in opposite directions (cf. 2d plot).

Solution of CCA

CCA Objective:

$$\arg\max_{\boldsymbol{w}} \overbrace{\boldsymbol{w}_x^\top C_{xy} \boldsymbol{w}_y}^{J(\boldsymbol{w})} \quad \text{s.t.} \quad \begin{aligned} \boldsymbol{w}_x^\top C_{xx} \boldsymbol{w}_x &= 1 \\ \boldsymbol{w}_y^\top C_{yy} \boldsymbol{w}_y &= 1 \end{aligned} \tag{1}$$

To find a solution, we apply the method of Lagrange Multipliers:

- **Step 1:** Writing the Lagrangian:

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\lambda}) = \boldsymbol{w}_x^\top C_{xy} \boldsymbol{w}_y + \frac{1}{2}\lambda_x \cdot (1 - \boldsymbol{w}_x^\top C_{xx} \boldsymbol{w}_x) + \frac{1}{2}\lambda_y \cdot (1 - \boldsymbol{w}_y^\top C_{yy} \boldsymbol{w}_y)$$

- **Step 2:** Set the gradient to zero

$$\begin{aligned} \nabla_{\boldsymbol{w}_x} \mathcal{L} = \mathbf{0} \quad &\Rightarrow \quad C_{xy} \boldsymbol{w}_y = \lambda_x C_{xx} \boldsymbol{w}_x \\ \nabla_{\boldsymbol{w}_y} \mathcal{L} = \mathbf{0} \quad &\Rightarrow \quad C_{yx} \boldsymbol{w}_x = \lambda_y C_{yy} \boldsymbol{w}_y \end{aligned}$$

A solution of the original problem must necessarily satisfy these two equations.

Solution of CCA

- **Step 3:** Multiplying the obtained equations by $\boldsymbol{w}_x^\top$ and $\boldsymbol{w}_y^\top$ respectively, we get:

$$\overbrace{\boldsymbol{w}_x^\top C_{xy} \boldsymbol{w}_y}^{J(\boldsymbol{w})} = \lambda_x \overbrace{\boldsymbol{w}_x^\top C_{xx} \boldsymbol{w}_x}^{1}$$
$$\underbrace{\boldsymbol{w}_y^\top C_{yx} \boldsymbol{w}_x}_{J(\boldsymbol{w})} = \lambda_y \underbrace{\boldsymbol{w}_y^\top C_{yy} \boldsymbol{w}_y}_{1}$$

  and find that $\lambda_x = \lambda_y \stackrel{\text{def}}{=} \lambda$.

- **Step 4:** Observe that the equations obtained in Step 2 can be written in block form as:

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} \tag{2}$$

  which is a *generalized* eigenvalue problem. I.e. the solution $(\boldsymbol{w}_x, \boldsymbol{w}_y)$ is one of the eigenvectors of this generalized eigenvalue problem.

  Note: The scale of the eigenvector is not determined by this equation and should be adjusted to satisfy the autocorrelation constraints, however, the original CCA objective (correlation in projected space) is invariant to the scale of $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$.

Solution of CCA

- **Remark 1:** To find which eigenvector it is, multiply both side of Equation (2) by $[\boldsymbol{w}_x, \boldsymbol{w}_y]^\top$:

$$\underbrace{\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}^\top \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}}_{2J(\boldsymbol{w})} = \lambda \underbrace{\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}^\top \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}}_{2}$$

  Observe that $J(\boldsymbol{w}) = \lambda$. Thus, to maximize the objective, we should maximize the eigenvalue λ, i.e. choose λ_1 (the highest eigenvalue).

- **Remark 2:** Let us interpret the eigenvalue λ_1 by observing that

$$\rho = \max_{\boldsymbol{w}} \mathrm{Corr}(\boldsymbol{w}_x^\top \boldsymbol{x}, \boldsymbol{w}_y^\top \boldsymbol{y}) = \max_{\boldsymbol{w}} \frac{\overbrace{\boldsymbol{w}_x^\top C_{xy} \boldsymbol{w}_y}^{J(\boldsymbol{w})}}{\underbrace{\sqrt{\boldsymbol{w}_x^\top C_{xx} \boldsymbol{w}_x}}_{1} \underbrace{\sqrt{\boldsymbol{w}_y^\top C_{yy} \boldsymbol{w}_y}}_{1}} = \lambda_1$$

  Therefore, λ_1 is simply the correlation of modalities when projected in CCA space.

CCA Stability

Generalized eigenvalue problem

$$\underbrace{\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}}_{A} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \underbrace{\begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}}_{B} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} \tag{3}$$

Equivalent eigenvalue problem (multiply Eq. (3) by B^{-1} on both sides):

$$\underbrace{\begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & C_{yy}^{-1} \end{bmatrix}}_{B^{-1}} \underbrace{\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}}_{A} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} \tag{4}$$

Remarks:

- Learned directions $[\boldsymbol{w}_x, \boldsymbol{w}_y]$ can be *instable* if B has near-zero eigenvalues.
- This can be addressed by adding a small diagonal term to B (i.e. $B \leftarrow B + \epsilon I$). This modification can be interpreted as augmenting the data with non-correlated Gaussian noise.

Part 2

CCA for High-Dimensional Data

CCA in High Dimensions

CCA generalized eigenvalue problem may be computationally expensive to solve if d is large, because it involves the eigendecomposition of matrices of size $d \times d$.

Idea:

- Force the solution to be a weighted combination of the data, and

$$\boldsymbol{w}_x = \sum_{i=1}^{N} (\boldsymbol{x}_i - \boldsymbol{\mu}_x)\alpha_{x,i} = \widetilde{X}\boldsymbol{\alpha}_x$$

$$\boldsymbol{w}_y = \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{\mu}_y)\alpha_{y,i} = \widetilde{Y}\boldsymbol{\alpha}_y$$

  with $\widetilde{X}$ and $\widetilde{Y}$ two matrices containing the *centered* data, and with $\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y \in \mathbb{R}^N$, and optimize the objective w.r.t. $\boldsymbol{\alpha}$.

Observation:

- The optimal choice of parameters $(\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y)$ always corresponds to parameters $(\boldsymbol{w}_x, \boldsymbol{w}_y)$ that are an optimal solution of the CCA problem ($\rightarrow$ homeworks).

CCA in High Dimensions

- CCA objective where we have injected $\boldsymbol{\alpha}$:

$$\arg\max_{\boldsymbol{\alpha}} \overbrace{\boldsymbol{\alpha}_x^\top \widetilde{X}^\top}^{\boldsymbol{w}_x^\top} C_{xy} \overbrace{\widetilde{Y}\boldsymbol{\alpha}_y}^{\boldsymbol{w}_y} \quad \text{s.t.} \quad \begin{array}{l} \overbrace{\boldsymbol{\alpha}_x^\top \widetilde{X}^\top}^{\boldsymbol{w}_x^\top} C_{xx} \overbrace{\widetilde{X}\boldsymbol{\alpha}_x}^{\boldsymbol{w}_x} = 1 \\ \underbrace{\boldsymbol{\alpha}_y^\top \widetilde{Y}^\top}_{\boldsymbol{w}_y^\top} C_{yy} \underbrace{\widetilde{Y}\boldsymbol{\alpha}_y}_{\boldsymbol{w}_y} = 1 \end{array} \tag{5}$$

- Observing that $C_{xy} = \frac{1}{N}\widetilde{X}\widetilde{Y}^\top$, $C_{xx} = \frac{1}{N}\widetilde{X}\widetilde{X}^\top$, $C_{yy} = \frac{1}{N}\widetilde{Y}\widetilde{Y}^\top$ the objective becomes:

$$\arg\max_{\boldsymbol{\alpha}} \boldsymbol{\alpha}_x^\top \overbrace{\tfrac{1}{N}\widetilde{X}^\top\widetilde{X}\widetilde{Y}^\top\widetilde{Y}}^{Q_{xy}} \boldsymbol{\alpha}_y \quad \text{s.t.} \quad \begin{array}{l} \boldsymbol{\alpha}_x^\top \overbrace{\tfrac{1}{N}\widetilde{X}^\top\widetilde{X}\widetilde{X}^\top\widetilde{X}}^{Q_{xx}} \boldsymbol{\alpha}_x = 1 \\ \boldsymbol{\alpha}_y^\top \underbrace{\tfrac{1}{N}\widetilde{Y}^\top\widetilde{Y}\widetilde{Y}^\top\widetilde{Y}}_{Q_{yy}} \boldsymbol{\alpha}_y = 1 \end{array} \tag{6}$$

where Q_{xy}, Q_{xx} and Q_{yy} are matrices of size $N \times N$, and where we never need to maintain matrices of size $d \times d$.

CCA in High Dimensions

- Recall that we have formulated the optimization problem as:

$$\arg\max_{\boldsymbol{\alpha}} \; \boldsymbol{\alpha}_x^\top Q_{xy} \boldsymbol{\alpha}_y \quad \text{s.t.} \quad \boldsymbol{\alpha}_x^\top Q_{xx} \boldsymbol{\alpha}_x = 1$$
$$\boldsymbol{\alpha}_y^\top Q_{yy} \boldsymbol{\alpha}_y = 1$$

- It has exactly the same structure as the original CCA formulation but with different matrices and optimization variables.
- Therefore, using the same framework of Lagrange multipliers, we can again find the solution $\boldsymbol{\alpha}$ as the leading eigenvector of a generalized eigenvalue problem (compare to Eq. (2)):

$$\underbrace{\begin{bmatrix} 0 & Q_{xy} \\ Q_{yx} & 0 \end{bmatrix}}_{A} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} = \lambda \underbrace{\begin{bmatrix} Q_{xx} & 0 \\ 0 & Q_{yy} \end{bmatrix}}_{B} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} \tag{7}$$

- The leading eigenvalue λ_1 can again be interpreted as the correlation in CCA space.
- For stability, it is useful to add a small diagonal term on the matrix B.

Part 3

Application: Neuro-Vascular Coupling

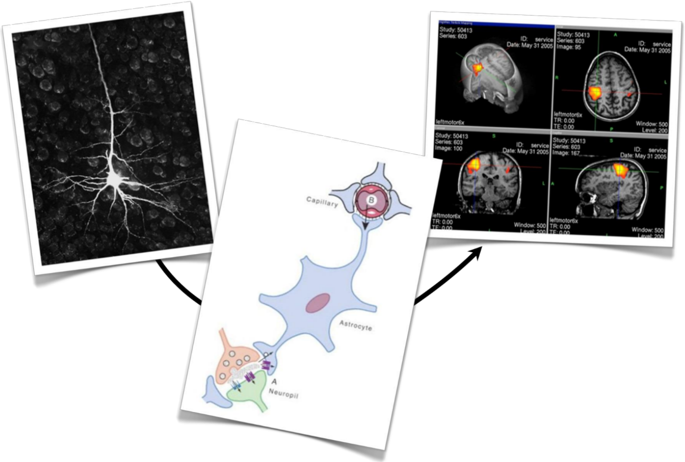
Adapted from K-R. Müller, Machine Learning 2 @ TU Berlin

Results from the paper Murayama Y et al. Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. Magn Reson Imaging 2010

Application: Neuro-Vascular Coupling

intercortical neural activity

FMRI/BOLD signal

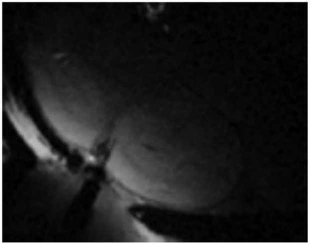


haemodynamic response

Experimental Setup

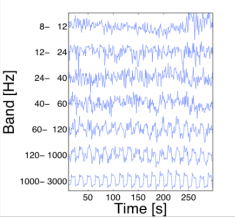
Simultaneous measurements of

- fMRI/ BOLD signal
- Intracortical neural activity

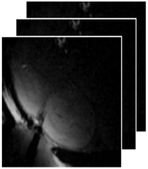


Analysis of Simultaneous Recordings

Spectrogram of neural activity



fMRI Time series



- X and Y have different dimensions
- High-dimensional data
- non-instantaneous couplings

From CCA to Temporal CCA

If variables are coupled with delays

- simultaneous samples will not be correlated
- standard CCA will not find the right solution

Solution:

- Shift one variable relative to the other
- Maximise correlation for (a sum over) all relative time lags

$$\boxed{\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_\tau w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)}$$

Relation between CCA and Temporal CCA

- Temporal CCA can be seen as a standard CCA on data expanded with multiple time lags (and with parameter w_x shared across these lags).

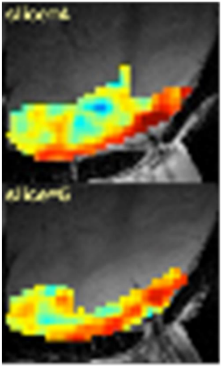
$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_\tau w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

$$\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_T} \end{bmatrix} \qquad \Downarrow \qquad \tilde{w}_x = \begin{bmatrix} w_x(\tau_1) \\ w_x(\tau_2) \\ \vdots \\ w_x(\tau_T) \end{bmatrix}$$

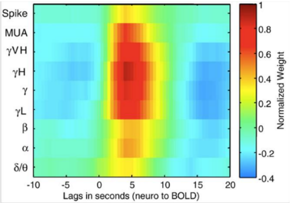
$$\underset{w_{\tilde{x}}, w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\tilde{w}_x^\top \tilde{X}, w_y^\top Y\right)$$

- High dimensionality can be addressed by representing w_x as a weighted combination of the data ($w_x = X\alpha_x$)

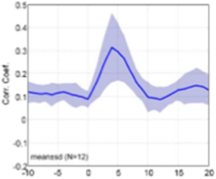
Application: Neuro-Vascular Coupling



Spatial Dependencies



Haemodynamic Response Function



Canonical Correlogram

Murayama et al., "Relationship between neural and haemodynamic signals during spontaneous activity studied with temporal kernel CCA", Magnetic Resonance Imaging, 2010

Summary

Summary

- The framework of Lagrange multipliers allows to convert the CCA objective into a generalized eigenvalue problem ($\Rightarrow$ no need for an iterative algorithm, and guarantee to find the true optimum).
- In high dimensions ($d \gg N$), the CCA computation can be made more efficient by solving an eigenvalue problem of size $N \times N$ instead of an eigenvalue problem of size $d \times d$.
- Multiple extensions of canonical correlation analysis for increasing usefulness in practical applications (e.g. CCA for high-dimensional data, temporal CCA and its application to modeling neurovascular coupling).