“Machine Learning estadístico para interfaces Cerebro-Computadora”

MODULO II - parte II

Dra. Victoria Peterson

vpeterson@santafe-conicet.gov.ar

Nov 2023

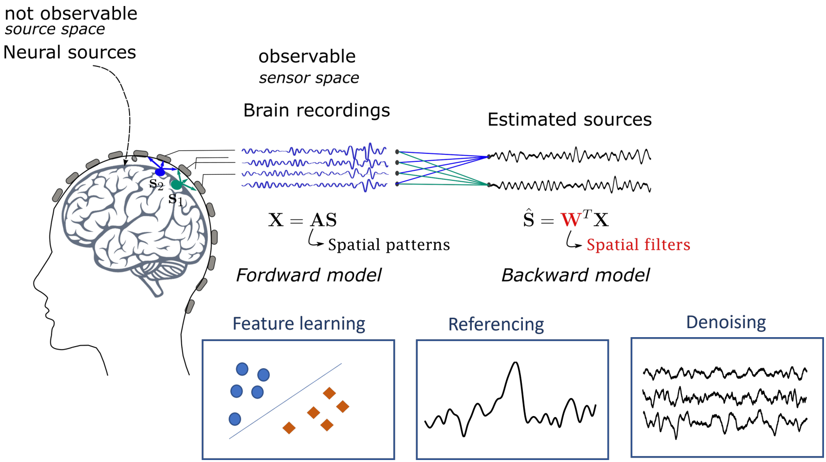
NiCALab

UNL

CONICET

I M A L

Filtrado Espacial: Recap



1. Filtrado Espacial en Denoising
    - Análisis de componentes principales (PCA)
    - Análisis de componentes Independientes (ICA)

1. Filtrado Espacial en Denoising
   - Análisis de componentes principales (PCA)
   - Análisis de componentes Independientes (ICA)

Motivación

not observable
source space

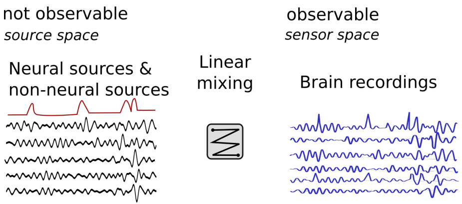
observable
sensor space

Neural sources

Linear mixing

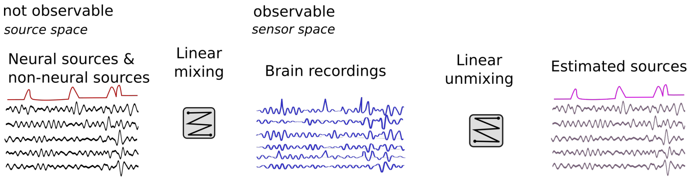
Brain recordings

Motivación



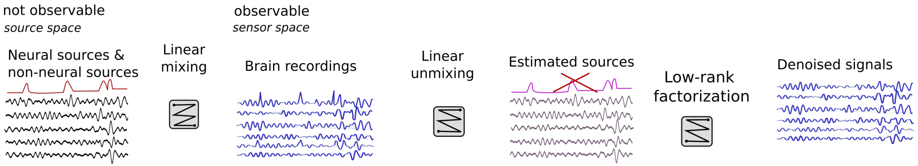
not observable
source space
Neural sources &
non-neural sources
Linear
mixing
observable
sensor space
Brain recordings

Motivación

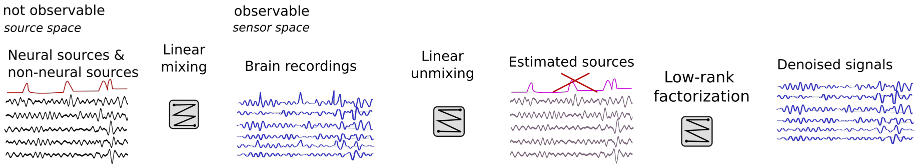


not observable
source space
Neural sources &
non-neural sources
Linear
mixing
observable
sensor space
Brain recordings
Linear
unmixing
Estimated sources

Motivación


not observable
source space
observable
sensor space
Neural sources &
non-neural sources
Linear
mixing
Brain recordings
Linear
unmixing
Estimated sources
Low-rank
factorization
Denoised signals

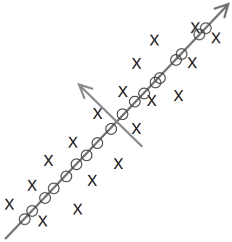
Motivación



¿Podremos encontrar una matriz de transformación W tal que los componentes de artefacto sean estimados?

1. Filtrado Espacial en Denoising
   - Análisis de componentes principales (PCA)
   - Análisis de componentes Independientes (ICA)

"Hallar el subespacio en el cuál la varianza de los datos sea maximizada"



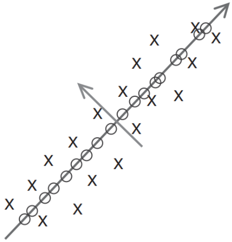
From: [Rao, 2013]

Formulación

$$\mathbf{w} = \arg\max \left\{ \mathbf{w}^T \Sigma \mathbf{w} \right\} \quad \text{s.t } \|\mathbf{w}\| = 1,$$

donde $\boldsymbol{\Sigma} \doteq \frac{1}{p}\tilde{X}\tilde{X}^T$, con $\tilde{X} \in \mathbb{R}^{N_c \times p}$ matriz de datos de EEG de media cero de dimensión p (datos crudos o época concatenadas).

"Hallar el subespacio en el cuál la varianza de los datos sea maximizada"



From: [Rao, 2013]

Formulación

$$\mathbf{w} = \arg\max \left\{ \mathbf{w}^T \Sigma \mathbf{w} \right\} \quad \text{s.t } \|\mathbf{w}\| = 1,$$

donde $\boldsymbol{\Sigma} \doteq \frac{1}{p}\tilde{X}\tilde{X}^T$, con $\tilde{X} \in \mathbb{R}^{N_c \times p}$ matriz de datos de EEG de media cero de dimensión p (datos crudos o época concatenadas).

Resolución via Lagrange Multipliers

Sean $x_1, \ldots x_p$ nuestro conjunto de variables originales, y sea z_i, $i = 1, \ldots, p$ una combinación lineal de esas variaciones:

$$z_i = \sum_{j=1}^{N_c} w_{i,j} x_i.$$

Sean $x_1, \ldots x_p$ nuestro conjunto de variables originales, y sea z_i, $i = 1, \ldots, p$ una combinación lineal de esas variaciones:

$$z_i = \sum_{j=1}^{N_c} w_{i,j} x_i.$$

$$var(z_i) = \mathbf{w}_i^T \Sigma \mathbf{w}_i,$$

Sean $x_1, \ldots x_p$ nuestro conjunto de variables originales, y sea z_i, $i = 1, \ldots, p$ una combinación lineal de esas variaciones:

$$z_i = \sum_{j=1}^{N_c} w_{i,j} x_i.$$

$$var(z_i) = \mathbf{w}_i^T \Sigma \mathbf{w}_i,$$

$\Rightarrow$ El autovector correspodiente al autovalor σ más grande de Σ será el **primer componente principal**.

Sean $x_1,\ldots x_p$ nuestro conjunto de variables originales, y sea z_i, $i = 1,\ldots,p$ una combinación lineal de esas variaciones:

$$z_i = \sum_{j=1}^{N_c} w_{i,j} x_i.$$

$$var(z_i) = \mathbf{w}_i^T \Sigma \mathbf{w}_i,$$

⇒ El autovector correspodiente al autovalor σ más grande de Σ será el **primer componente principal**.
⇒ $\mathbf{w}_i$ son los autovectores de Σ

Componentes principales, cantidad, aporte en la varianza

Cada componente principal aporta $\dfrac{\sigma_i}{\sum \sigma_i}$, a la varianza total, siendo σ_i el i-ésimo autovalor de Σ.

Componentes principales, cantidad, aporte en la varianza

Cada componente principal aporta $\dfrac{\sigma_i}{\sum \sigma_i}$, a la varianza total, siendo σ_i el i-ésimo autovalor de Σ. Se suele elegir las componentes que representan el 90 % de la varianza,

Componentes principales, cantidad, aporte en la varianza

Cada componente principal aporta $\frac{\sigma_i}{\sum \sigma_i}$, a la varianza total, siendo σ_i el i-ésimo autovalor de Σ. Se suele elegir las componentes que representan el 90 % de la varianza, o realizar una gráfica del espectro de los autovalores y seleccionar la cantidad de componentes mínima.

Componentes principales, cantidad, aporte en la varianza

Cada componente principal aporta $\frac{\sigma_i}{\sum \sigma_i}$, a la varianza total, siendo σ_i el i-ésimo autovalor de Σ. Se suele elegir las componentes que representan el 90 % de la varianza, o realizar una gráfica del espectro de los autovalores y seleccionar la cantidad de componentes mínima.

Corrección de la media

PCA supone datos estándarizados (media cero, desvío uno). Para corregir esto simplemente:

$$\boldsymbol{z} = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})$$

Componentes principales, cantidad, aporte en la varianza

Cada componente principal aporta $\frac{\sigma_i}{\sum \sigma_i}$, a la varianza total, siendo σ_i el i-ésimo autovalor de Σ. Se suele elegir las componentes que representan el 90 % de la varianza, o realizar una gráfica del espectro de los autovalores y seleccionar la cantidad de componentes mínima.

Corrección de la media

PCA supone datos estándarizados (media cero, desvío uno). Para corregir esto simplemente:

$$\boldsymbol{z} = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})$$

Aproximación de los datos

La transformación $\mathbf{W}$ es invertible. Definimos $\mathbf{A} = (\mathbf{W}^T)^{-1}$ Entonces $\mathbf{x} = \mathbf{A}\boldsymbol{z} + \boldsymbol{\mu}$

Aproximación de los datos

La transformación $\mathbf{A}$ es invertible. Entonces $\mathbf{x} = \mathbf{A}\boldsymbol{z} + \boldsymbol{\mu}$. Si k, $k < p$ componentes principales se utilizan:

$$\mathbf{x}_k = \mathbf{A}_k \boldsymbol{z}_k + \boldsymbol{\mu}.$$

Aproximación de los datos

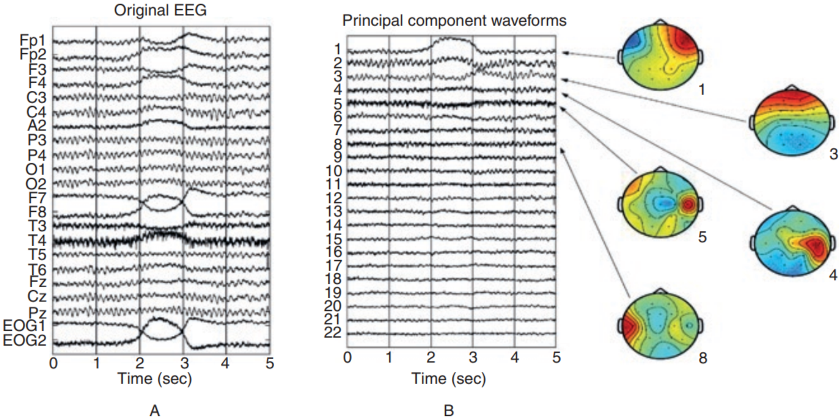
La transformación $\mathbf{A}$ es invertible. Entonces $\mathbf{x} = \mathbf{A}\boldsymbol{z} + \boldsymbol{\mu}$. Si k, $k < p$ componentes principales se utilizan:

$$\mathbf{x}_k = \mathbf{A}_k \boldsymbol{z}_k + \boldsymbol{\mu}.$$

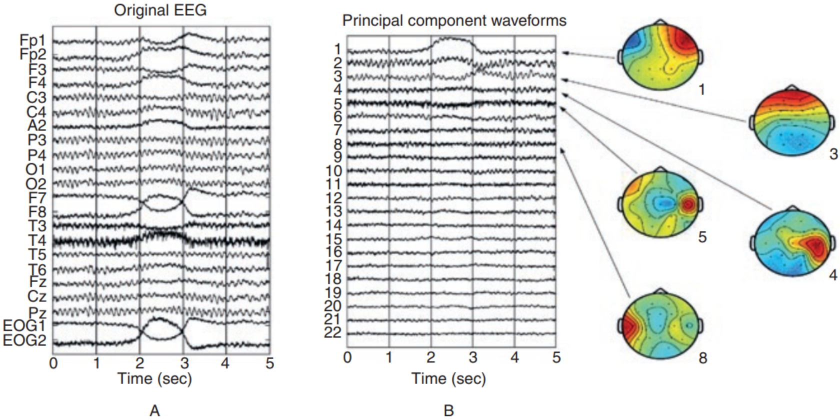
Whitening y PCA

La transformación de "blanqueo" o transformación esférica es una transformación lineal que mapea un vector de variables aleatorias de matriz de covarianza conocida en un nuevo conjunto de variables cuya varianza es la identidad $\mathbf{I}$, es decir están decorrelacionados y su varianza es 1.
Mediante PCA:

$$\tilde{\mathbf{x}} = \mathbf{D}^{1/2}\mathbf{A}^T(\mathbf{x} - \boldsymbol{\mu}), \quad \text{siendo } \mathbf{D} = diag(\sigma_i).$$



Fuente: [Rao, 2013]



Fuente: [Rao, 2013]

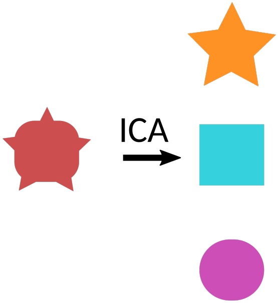
¿Será siempre buena idea maximizar la varianza en el EEG?

1. Filtrado Espacial en Denoising
   - Análisis de componentes principales (PCA)
   - Análisis de componentes Independientes (ICA)

Análisis de componentes Independientes (ICA)

[Oja and Hyvarinen, 2000]

"Maximizar la independencia entre componentes"



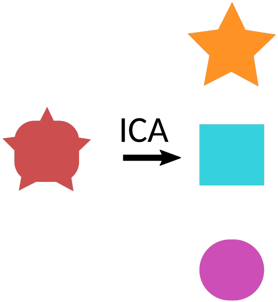
Problem formulation in brief

Hallar W tal que $\hat{\mathbf{s}} = W^T X$ sea *independiente*.

Análisis de componentes Independientes (ICA)

[Oja and Hyvarinen, 2000]

"Maximizar la independencia entre componentes"



Problem formulation in brief

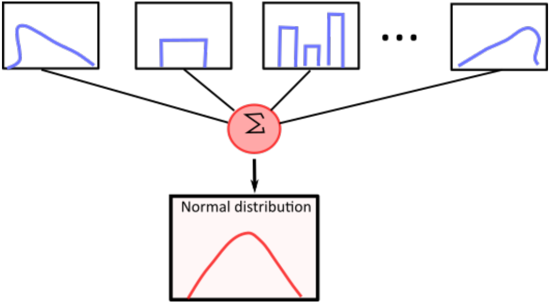
Hallar W tal que $\hat{\mathbf{s}} = W^T X$ sea *independiente*.

Varias formas de atacar el problema!!!

Análisis de componentes Independientes (ICA)

[Oja and Hyvarinen, 2000]

No-Gaussianidad e independencia



"No-Gaussianidad implica independencia"

Objetivo: maximizar la Negentropía

Una variable Gaussiana tiene mayor *entropía* entre todas las variables aleatorias de la misma varianza. Definimos la Negentropía de una v.a. $\mathbf{y}$ mediante

$$N(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}),$$

donde $H(\cdot)$ es la entropía y $\mathbf{y}_{gauss}$ es una variable Gaussiana de misma covarianza que $\mathbf{y}$

Objetivo: maximizar la Negentropía

Una variable Gaussiana tiene mayor *entropía* entre todas las variables aleatorias de la misma varianza. Definimos la Negentropía de una v.a. $\mathbf{y}$ mediante

$$N(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}),$$

donde $H(\cdot)$ es la entropía y $\mathbf{y}_{gauss}$ es una variable Gaussiana de misma covarianza que $\mathbf{y}$ $\Rightarrow$ Maximizar $N(\mathbf{y})$ implica encontrar funciones no-Gaussianas

[Oja and Hyvarinen, 2000]

No-Gaussianidad e independencia

Objetivo: maximizar la Negentropía

Una variable Gaussiana tiene mayor *entropía* entre todas las variables aleatorias de la misma varianza. Definimos la Negentropía de una v.a. $\mathbf{y}$ mediante

$$N(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}),$$

donde $H(\cdot)$ es la entropía y $\mathbf{y}_{gauss}$ es una variable Gaussiana de misma covarianza que $\mathbf{y}$ $\Rightarrow$ Maximizar $N(\mathbf{y})$ implica encontrar funciones no-Gaussianas

$\rightarrow$ difícil de computar $\Rightarrow$ formulaciones para aproximar la negentropía son utilizadas

Whitening

Variables independientes son siempre variables no correlacionadas

Whitening

Variables independientes son siempre variables no correlacionadas
$\Rightarrow$ el primer paso es hacer whitening de los datos

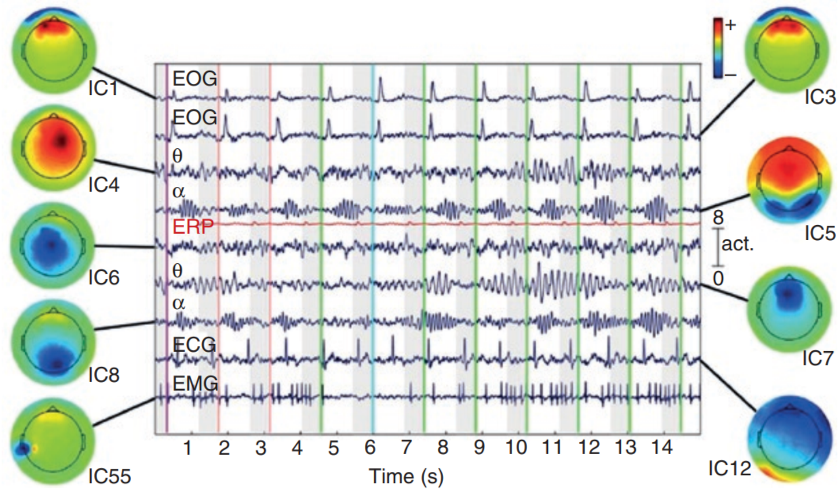
Whitening

Variables independientes son siempre variables no correlacionadas
$\Rightarrow$ el primer paso es hacer whitening de los datos
$\Rightarrow$ PCA puede ser utilizado para blanquear los datos y reducir la dimensionalidad.

Solvers

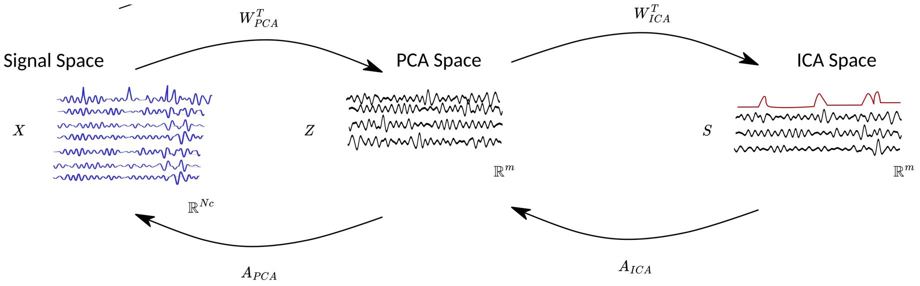
Differentes solvers existen:

- FastICA: maximiza la no-Gaussianidad [Hyvarinen, 1999].
- Infomax: maximiza la información mutua [Bell and Sejnowski, 1995, Lee et al., 1999].
- PICCARD: maximiza la verosimilitud [?].

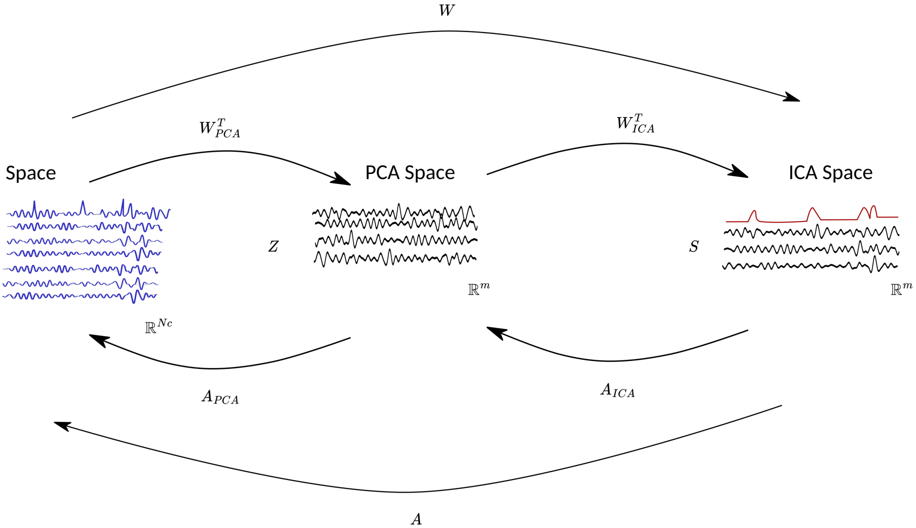


Fuente: [Rao, 2013]

PCA + ICA denoising pipeline



PCA + ICA denoising pipeline



Bibliografía utilizada I

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3):626–634.

Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441.

Oja, E. and Hyvarinen, A. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

Rao, R. P. N. (2013). *Brain-computer interfacing: an introduction*. Cambridge University Press.

Webb, A. R. (2003). *Statistical pattern recognition*. John Wiley & Sons.