

Aprendizaje no supervisado

VC09: Análisis de Componentes Independientes

Félix José Fuentes Hurtado

felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia

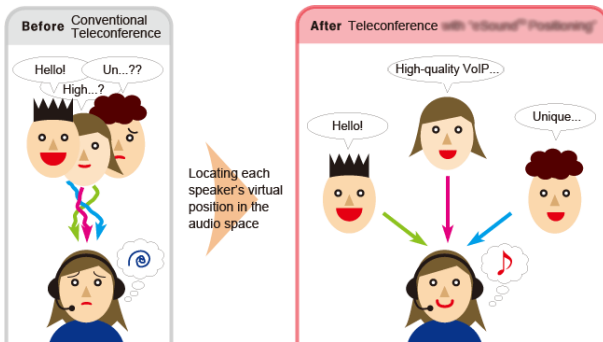
ICA vs. PCA

- ▶ Mismo resultado (datos transformados)
- ▶ Distinta hipótesis
- ▶ Reducción de dimensionalidad (dimensiones que mejor explican los datos)
- ▶ Separación de las fuentes/variables originales

Personas que hablan simultáneamente

Una persona escuchando (ej., micrófonos) en diferentes puntos de la sala captaría una mezcla diferente de mensajes

¿se pueden separar las diferentes voces (fuentes) para descifrar los distintos mensajes?



Basic:

<https://www.youtube.com/watch?v=wI1rddNbXDo>

Intermediate:

<https://www.youtube.com/watch?v=pSwR05d266I>

Final:

<https://www.youtube.com/watch?v=e4woe8GRjEI>

Definición del problema

- ▶ Existe un grupo de u **fuentes independientes**
- ▶ Un vector original es $\mathbf{s} \in \mathbb{R}^u$, un punto de cada fuente en un instante concreto
- ▶ Existe un grupo de v **receptores**
- ▶ El vector observado es $\mathbf{x} \in \mathbb{R}^v$
- ▶ Se asume la existencia de una **matriz de mezcla** A tal que:

$$\mathbf{x} = A\mathbf{s}$$

o una **matriz de separación**, W ($W = A^{-1}$):

$$\mathbf{s} = W\mathbf{x}$$

Problema de optimización

Encontrar el vector w_j que permite recomponer la j -ésima fuente:

$$s_j = w_j^t x$$

para todo $j \in \{1, \dots, u\}$

Problema de optimización

Encontrar el vector w_j que permite recomponer la j -ésima fuente:

$$s_j = w_j^t x$$

para todo $j \in \{1, \dots, u\}$

Solución

Buscaremos el estimador máximo verosímil de W

Propiedad de la suposición de independencia

La probabilidad conjunta es igual al producto de las distribuciones marginales de las fuentes:

$$p(\mathbf{s}) = \prod_{j=1}^u p_s(s_j)$$

Expresado según lo observado, \mathbf{x} :

$$p_{\mathbf{x}}(\mathbf{x}) = \prod_{j=1}^u p_s(\mathbf{w}_j^t \mathbf{x}) \cdot |W|$$

el determinante de W hace que la distribución de probabilidad integre a 1

Distribución de probabilidad de p_s de una fuente

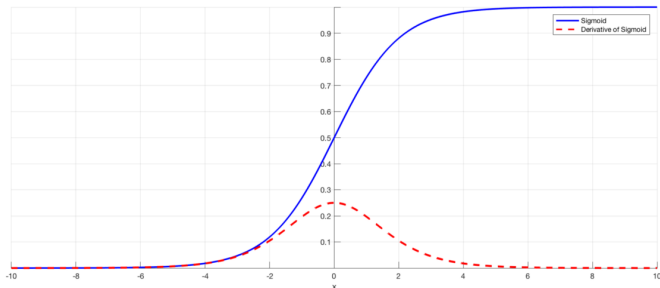
Usaremos la función sigmoide como distribución acumulativa:

$$g(s_j) = 1/(1 + e^{-s_j})$$

** Una distr. acumulativa es una función monótona que crece suavemente de 0 a 1

Su derivada es la función de densidad:

$$p_s(s_j) = g'(s_j)$$



Verosimilitud:

$$L(W; \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = \prod_{i=1}^n \left(\prod_{j=1}^u g'(\mathbf{w}_j^t \mathbf{x}_i) \cdot |W| \right)$$

Logaritmo de la verosimilitud:

$$\log L(W; \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = \sum_{i=1}^n \left(\sum_{j=1}^u \log g'(\mathbf{w}_j^t \mathbf{x}_i) + \log |W| \right)$$

Estimador máximo-verosímil de W

Algoritmo de ascenso de gradiente estocástico

- ▶ Obtener una matriz W inicial
- ▶ Iterativamente y para cada caso observado \mathbf{x}_i , tomar un paso en la dirección de máximo ascenso dada por el gradiente*:

$$W \leftarrow W + \alpha \left(\begin{bmatrix} 1 - 2g(\mathbf{w}_1^t \mathbf{x}_i) \\ 1 - 2g(\mathbf{w}_2^t \mathbf{x}_i) \\ \vdots \\ 1 - 2g(\mathbf{w}_n^t \mathbf{x}_i) \end{bmatrix} \mathbf{x}_i^t + (W^t)^{-1} \right)$$

donde α (ratio de aprendizaje) determina el paso de actualización de W

- ▶ Converge a un máximo local
(W no cambia sustancialmente entre iterat. consecutivas)

* generalización de la derivada a múltiples dimensiones

Dada una estimación (máximo verosímil) de la matriz W , los valores de las fuentes se obtienen a partir de una observación \mathbf{x}_i :

$$\mathbf{s}_i = W\mathbf{x}_i$$

- ▶ Suposición de independencia entre observaciones no realista
- ▶ Aun así, si el conjunto de datos suficientemente grande, el rendimiento del algoritmo no se ve comprometido

Si la correlación entre muestras es evidente, se realiza un recorrido aleatorio por las observaciones x_i en el ascenso del gradiente estocástico (acelerar la convergencia)

- ▶ El algoritmo converge a un óptimo local
- ▶ Dependiendo del ratio de aprendizaje, α , el algoritmo puede escapar de los óptimos locales con cierta facilidad

Hacer una exploración aleatoria del conjunto de datos en cada iteración ayuda a prevenir este problema

Aprendizaje no supervisado

VC09: Análisis de Componentes Independientes

Félix José Fuentes Hurtado

felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia