ICA

ICA vs. PCA

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

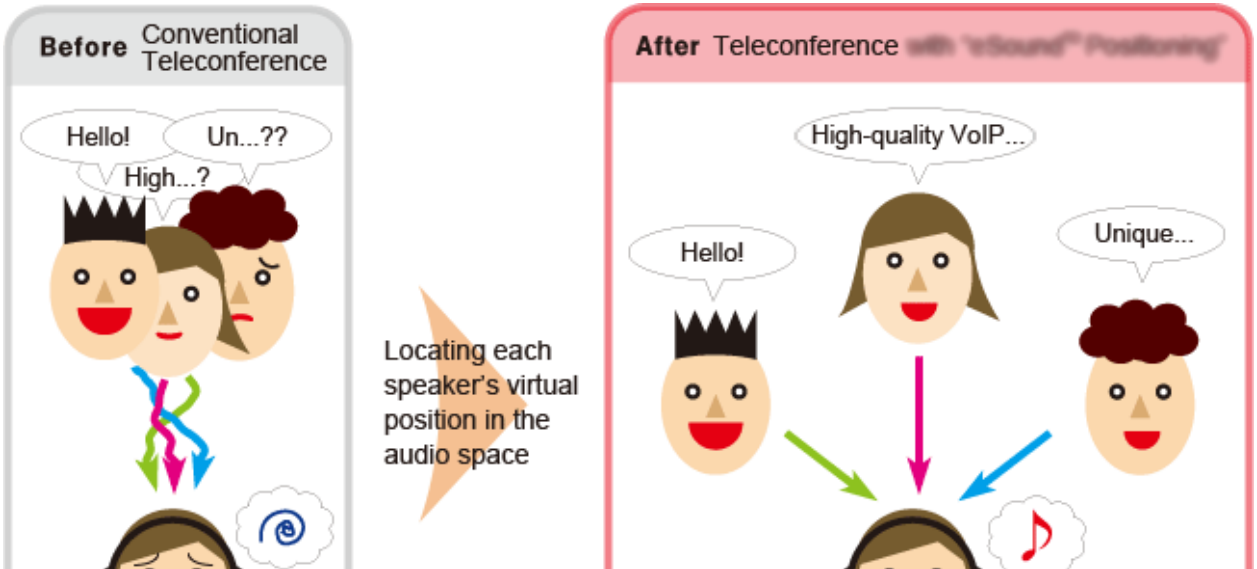
ICA

Ejemplo clásico: Fiesta-cóctel

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?



ICA

Basic:

https://www.youtube.com/watch?v=wIlrddNbXDo

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

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

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

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

ICA

Problema de optimización

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

$$s_j = \boldsymbol{w}_j^t \boldsymbol{x}$$

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

ICA

Problema de optimización

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

$$s_j = \boldsymbol{w}_j^t \boldsymbol{x}$$

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

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:

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

Expresado según lo observado, $\boldsymbol{x}$:

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

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

Distribución de probabilidad de p_s de una fuente

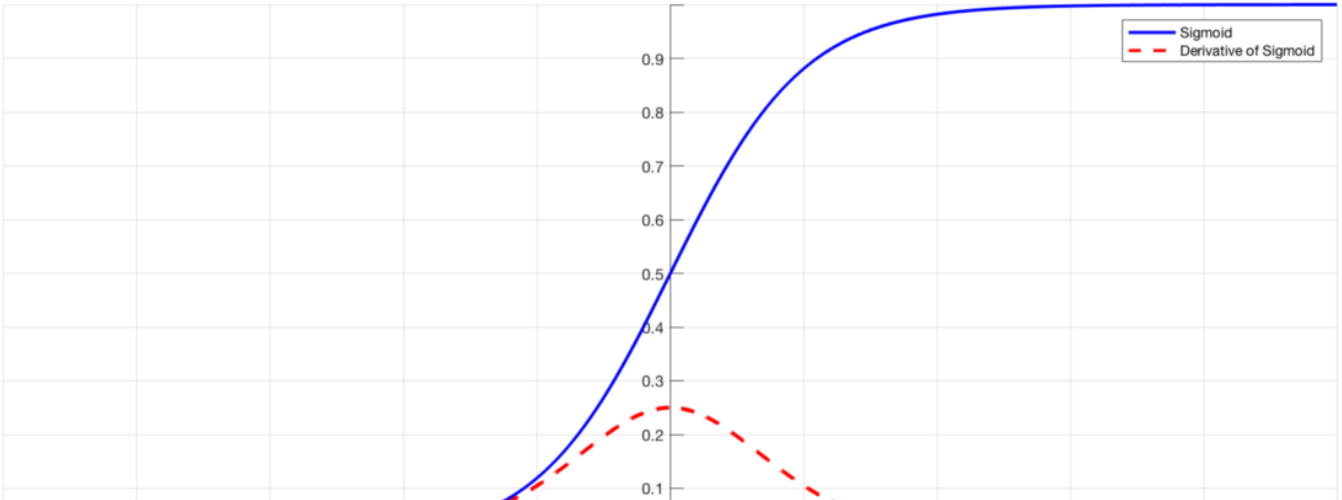
Usaremos la función sigmoide como distribución cumulativa:

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

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

Su derivada es la función de densidad:

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



ICA

Verosimilitud:

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

Logaritmo de la verosimilitud:

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

Estimador máximo-verosímil de W

Algoritmo de ascenso de gradiente estocástico

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

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

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

- Converge a un máximo local

  (W no cambia sustancialmente entre iterat. consecutivas)

ICA

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

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

ICA

Pros y contras

- 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 $\boldsymbol{x}_i$ 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