UC | Chile

UC | Chile

Algebra Lineal Aplicada para Ciencia de Datos

Clase sincrónica 4. Factorización No Negativa de Matrices

1. Motivación
2. Formulación del problema
3. Criterio de norma Frobenius
4. Criterio de divergencia KL

UC | Chile

Motivación

Técnicas de reducción de dimensionalidad lineal

Objetivo: Identificar la estructura detrás de una conjunto de datos y extraer informacion siginificativa

Ejemplos de técnicas LDR son:

- Análisis de componentes principales (PCA)
- Análisis de componentes principales robusto (RPCA)
- Completación de matrices de rango bajo
- Análisis de componentes sparse
- **Factorización no negativa de matrices**

Técnicas de reducción de dimensionalidad lineal

Estas técnicas pueden ser aplicadas a un amplio rango de aplicaciones como:

- Sistemas recomendadores
- Reducción de orden en modelos e identificacion de sistemas
- Clustering
- Análisis de imágenes
- Separació ciega de fuentes.

Factorización no negativa de matrices

La **factorización no negativa de matrices** (NMF) requiere que los factores de la aproximación de rango bajo sean no negativos en cada componente. **Intepretabilidad!**
Aplicaciones de NMF:

- Extraer partes de caras
- Identificar temas en documentos
- Aprender modelos de Markov
- Separar fuentes de audio de una mezcla
- Detectar comunidades en redes
- Analizar imágenes medicas
- Descomponer microarrays en expresión de genes

Motivación

Suponemos que tenemos datos representados por vectores $y_1, \ldots, y_n \in \mathbb{R}^m$

En la práctica, para analizar estos datos calculamos primero el **promedio**

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Esto nos permite definir la **matriz de desviaciones**

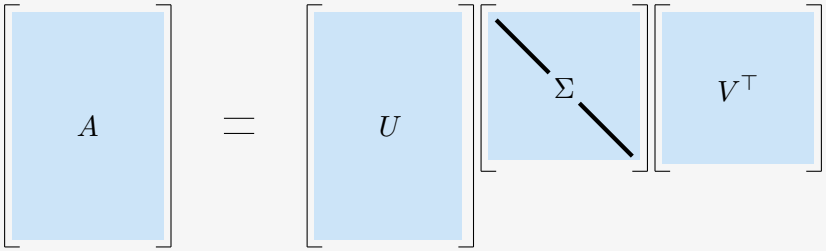
$$X = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \quad \text{con} \quad x_i := y_i - \bar{y} \quad \text{para} \quad i \in \{1, \ldots, n\}$$

Motivación

La **descomposición en valores singulares** o **SVD** nos permite representar esta matriz como

$$X = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$$

donde r es el **rango** de la matriz X, $u_1, \ldots, u_r \in \mathbb{R}^m$ y $v_1, \ldots, v_r \in \mathbb{R}^n$ son los **vectores singulares**, y $\sigma_1 \geq \ldots \geq \sigma_r > 0$ son los **valores singulares**



Motivación

Ya que

$$u_i v_i^\top = \begin{bmatrix} v_{i,1}u_i, & \ldots, & v_{i,n}u_i \end{bmatrix} = \begin{bmatrix} | & | & & | \\ v_{i,1}u_i, & v_{i,2}u_i, & \cdots, & v_{i,n}u_i \\ | & | & & | \end{bmatrix}$$

la representación anterior es equivalente a

$$X = \begin{bmatrix} \sum_{i=1}^{r} \sigma_i v_{i,1} u_i, & \ldots, & \sum_{i=1}^{r} \sigma_r v_{n,1} u_i \end{bmatrix}$$

Motivación

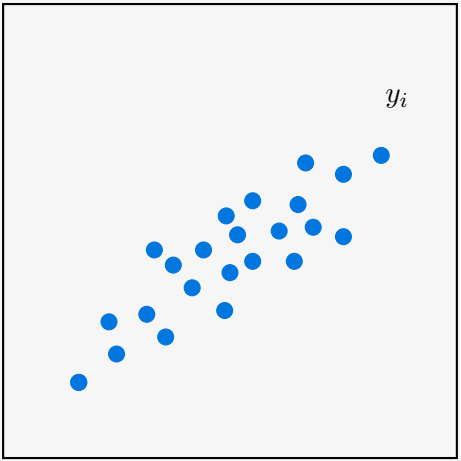
En particular, concluimos que

$$x_j = \sum_{i=1}^{r} \sigma_i v_{i,j} u_i \quad \Rightarrow \quad y_j = \bar{y} + \sum_{i=1}^{r} \sigma_i v_{i,j} u_i$$

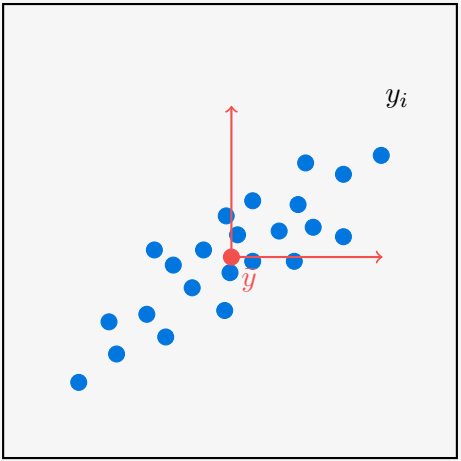
El **teorema de Eckart-Young** nos dice que esta es la forma más eficiente de aproximar la matriz X

Por lo tanto, la **colección ortonormal** $u_1, \ldots, u_r$ está **adaptada** a las desviaciones de los datos

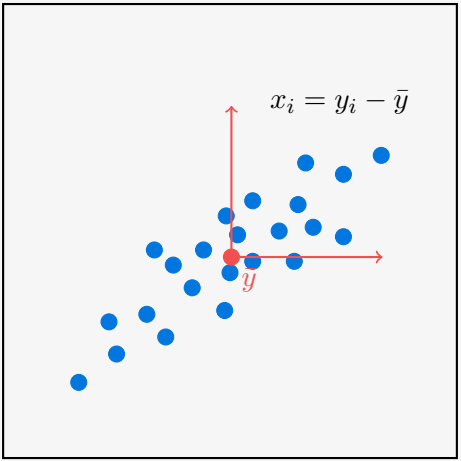
Motivación



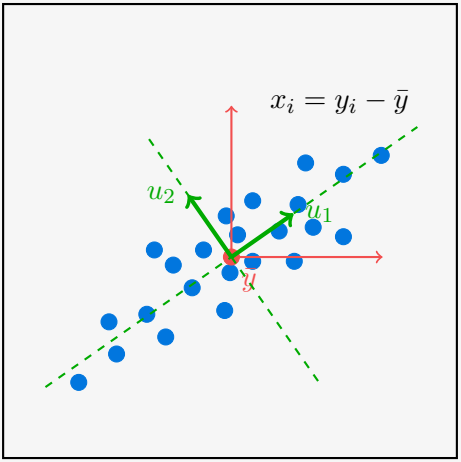
Motivación



Motivación



Motivación



Motivación

Sin embargo, hay algunos datos que tienen la propiedad de ser **no negativos**

Esto quiere decir que las componentes de los vectores $y_1, \ldots, y_n$ son mayores o iguales a cero

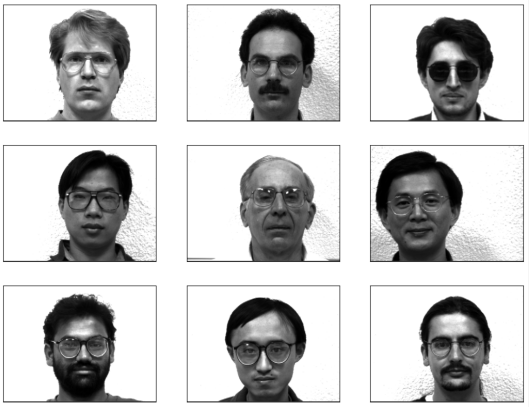
En este caso escribimos $y_1, \ldots, y_n \geq 0$

Este es el caso de las **imágenes** o **vídeos**

Motivación

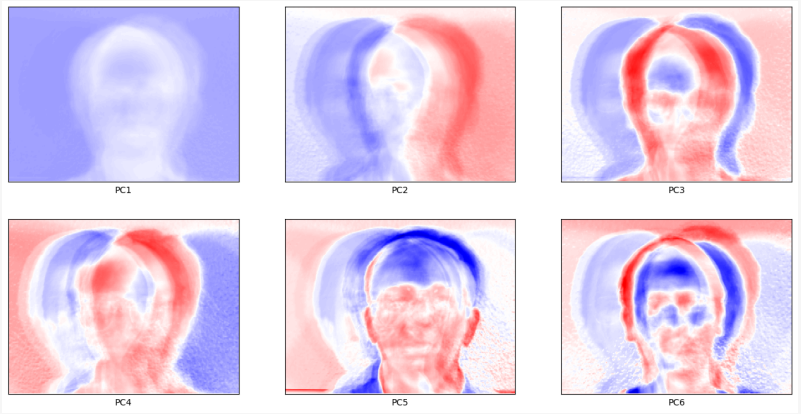
Para estos datos, puede no ser natural sustraer el promedio

Motivación



Motivación

Motivación



Motivación

Para estos datos, puede no ser natural sustraer el promedio

¿Qué hacemos en este caso?

UC | Chile

Formulación del problema

Formulación del problema

Suponemos que $Y \in \mathbb{R}^{m \times n}$ es **no negativa** y escribimos $Y \geq 0$

Nos damos un valor r **arbitrario**

Buscamos matrices $W \in \mathbb{R}^{m \times r}$ y $H \in \mathbb{R}^{r \times n}$ tales que

$$Y \approx WH \quad \text{y} \quad W, H \geq 0$$

Esto se llama **factorización no negativa** o **NMF**

Formulación del problema

Si escribimos

$$W = \begin{bmatrix} w_1 & \dots & w_r \end{bmatrix}$$

entonces $Y \approx WH$ quiere decir que

$$y_i \approx \sum_{j=1}^{r} H_{j,i} w_j$$

Aproximamos cada dato como una combinación de vectores no negativos

Formulación del problema

¿Cuál podría ser el análogo del promedio?

Podemos definir el vector

$$y_{\min} = \begin{bmatrix} \min\{y_{1,1}, \ldots, y_{1,n}\} \\ \vdots \\ \min\{y_{m,1}, \ldots, y_{m,n}\} \end{bmatrix}$$

En el contexto de imagenes, se conoce como **imagen de mínima proyección**

Formulación del problema

Podemos ensamblar la **matriz de desviaciones**

$$X = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \quad \text{con} \quad x_i = y_i - y_{\min} \quad \text{para} \quad i \in \{1, \dots, n\}$$

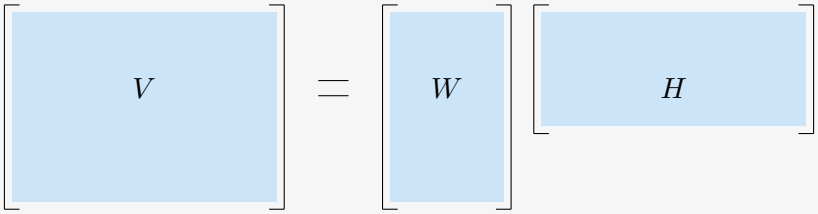
para luego buscar $W \in \mathbb{R}^{m \times r}$ y $H \in \mathbb{R}^{r \times n}$ tales que

$$X \approx WH \quad \text{y} \quad W, H \geq 0.$$

Formulación del problema

Para una matrix $V \in \mathbb{R}^{m \times n}$ donde cada elemento $v_{i,j} \geq 0$, la NMF descompone V en dos matrices $W \in \mathbb{R}^{m \times r}$ y $H \in \mathbb{R}^{r \times n}$, donde $w_{i,k} \geq 0$, y $h_{k,j} \geq 0$, y $r < \min\{m, n\}$, tales que:

$$V = W H$$

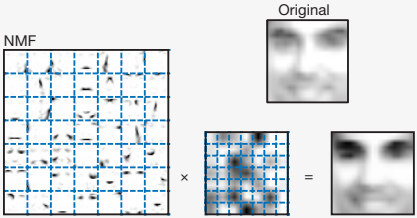


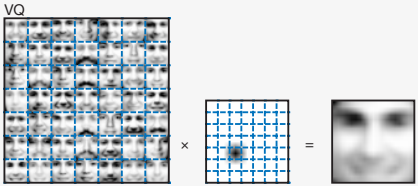
Este problema, en general no puede ser resuelto de forma analitica y se aproxima numéricamente

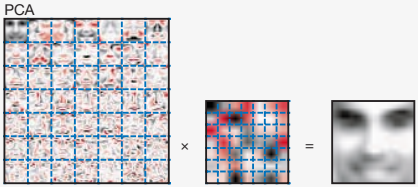
Referencias NMF

- Daniel D. Lee and H. Sebastian Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401,788-791.
- David Donoho and Victoria Stodden (2003). When does non-negative matrix factorization give a correct decomposition into parts? *Advances in neural information processing systems*, 16.
- Nicolas Gillis (2020). Nonnegative matrix factorization. Society for Industrial and Applied Mathematics.

Ejemplo de NMF

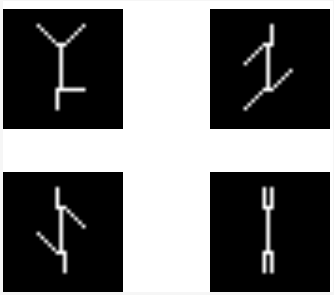


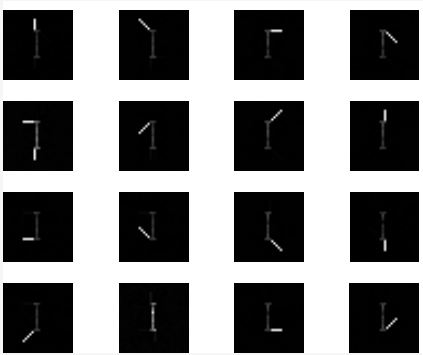




Ejemplo de NMF del paper de Lee & Seung. NMF aprende una representación por partes de rostros. El conjunto de datos es $m = 2429$ imagenes de rostros cada uno de $n = 19 \times 19$ pixeles, los que constituyen una matriz V de $n \times m$. Cada método aprende un conjunto de $r = 49$ imágenes base.

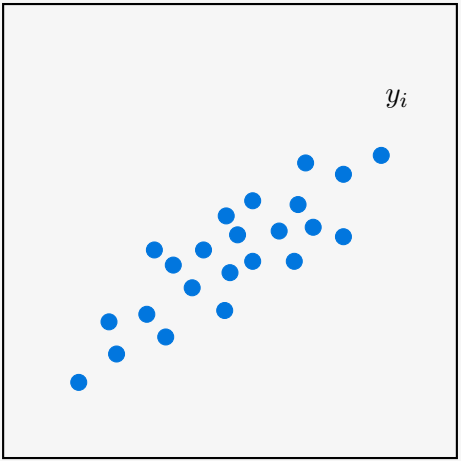
Ejemplo de NMF



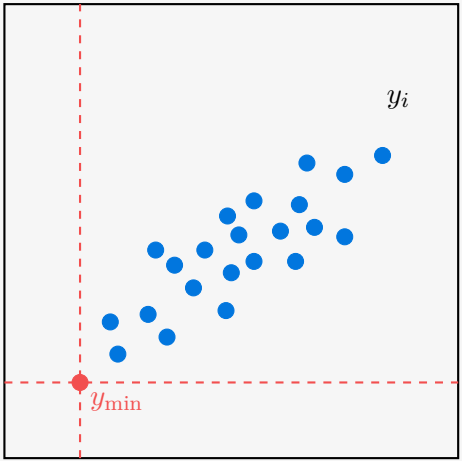


Ejemplo de NMF del paper de Donoho & Stodden. NMF aprende una representación por partes de las imágenes de “swimmers”. El conjunto de datos es $n = 256$ imagenes de “swimmers”. El método aprende un conjunto de $r = 16$ imágenes base.

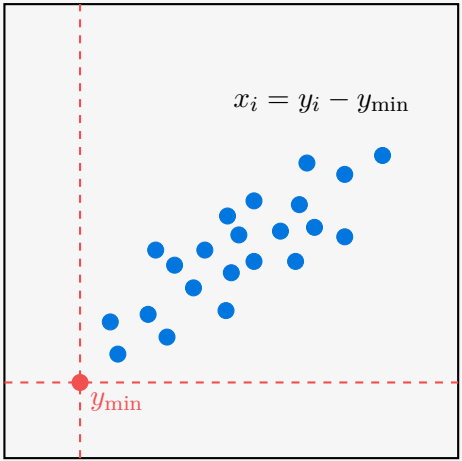
Formulación del problema



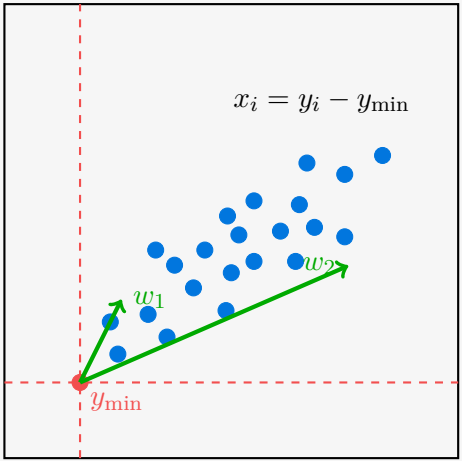
Formulación del problema



Formulación del problema



Formulación del problema



Formulación del problema

En general, **no necesariamente existen matrices** $W \in \mathbb{R}^{m \times r}$ y $H \in \mathbb{R}^{r \times n}$ tales que

$$V = WH$$

Por lo tanto, en la práctica se busca una factorización **aproximada**

Para encontrar W, H necesitamos definir un **criterio**

Siguiendo a Lee & Seung, estudiamos el criterio de **norma Frobenius** y **divergencia KL**

UC | Chile

Criterio de norma Frobenius

Criterio de norma Frobenius

Recordemos que para una matriz A de $m \times n$ se tiene

$$\|A\|_F^2 = \sum\nolimits_{i,j} A_{i,j}^2$$

Lee & Seung proponen como una primera alternativa encontrar W, H que resuelven

$$\underset{W,H}{\text{minimizar}} \quad \|V - WH\|_F \quad \text{sujeto a} \quad W, H \geq 0$$

Criterio de norma Frobenius

Este problema **no tiene** solución explícita

En general, la solución óptima no tiene por qué alcanzar el valor cero

En otras palabras, si $W^\star, H^\star$ son las soluciones óptimas, no necesariamente se tiene

$$V = W^\star H^\star$$

¿Cómo resolvemos este problema?

Criterio de norma Frobenius

Lee & Seung proponen el siguiente **método iterativo**

A partir de $W^{(0)}, H^{(0)}$ dados, generamos la sucesión $W^{(1)}, H^{(1)}, \ldots$ dada por

$$W_{i,j}^{(k+1)} = \frac{(VH^{(k)\top})_{i,j}}{(W^{(k)}H^{(k)}H^{(k)\top})_{i,j}} W_{i,j}^{(k)}$$

$$H_{i,j}^{(k+1)} = \frac{(W^{(k)\top}V)_{i,j}}{(W^{(k)\top}W^{(k)}H^{(k)})_{i,j}} H_{i,j}^{(k)}$$

para $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$

Criterio de norma Frobenius

Este método converge a distintas matrices **dependiendo de la inicialización** $W^{(0)}, H^{(0)}$

Por lo tanto, no podemos hablar de **la** factorización, sino que de **una** factorización aproximada

Lin recomienda usar una versión modificada de este algoritmo

UC | Chile

Criterio de divergencia KL

Criterio de divergencia KL

La **divergencia KL** entre dos matrices **no negativas** A, B de **igual tamaño** $m \times n$ es

$$D_{\mathsf{KL}}(A\|B) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(A_{i,j}\log\left(\frac{A_{i,j}}{B_{i,j}}\right) - A_{i,j} + B_{i,j}\right)$$

Lee & Seung proponen como una segunda alternativa encontrar W, H que resuelven

$$\underset{W,H}{\text{minimizar}} \quad D_{\mathsf{KL}}(V\|WH) \quad \text{sujeto a} \quad W, H \geq 0$$

Criterio de divergencia KL

Este problema **no tiene** solución explícita

En general, la solución óptima no tiene por qué alcanzar el valor cero

En otras palabras, si $W^\star, H^\star$ son las soluciones óptimas, no necesariamente se tiene

$$V = W^\star H^\star$$

¿Cómo resolvemos este problema?

Criterio de divergencia KL

Lee & Seung proponen el siguiente **método iterativo**

A partir de $W^{(0)}, H^{(0)}$ dados, generamos la sucesión $W^{(1)}, H^{(1)}, \ldots$ para la que definimos las variables auxiliares

$$\widehat{V}^{(k)} := W^{(k)} H^{(k)} \quad \text{y} \quad R_{i,j}^{(k)} = \frac{V_{i,j}}{\widehat{V}_{i,j}^{(k)}}$$

Criterio de divergencia KL

Luego, calculamos

$$W_{i,j}^{(k+1)} = \frac{\sum_{\ell=1}^{n} H_{j,\ell}^{(k)} R_{i,\ell}^{(k)}}{\sum_{\ell'=1}^{n} H_{j,\ell'}^{(k)}} W_{i,j}^{(k)}$$

$$H_{i,j}^{(k+1)} = \frac{\sum_{\ell=1}^{n} W_{\ell,i}^{(k)} R_{\ell,j}^{(k)}}{\sum_{\ell'=1}^{m} W_{\ell',i}^{(k)}} H_{i,j}^{(k)}$$

para $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$

Criterio de divergencia KL

Este método converge a distintas matrices **dependiendo de la inicialización** $W^{(0)}, H^{(0)}$

Por lo tanto, no podemos hablar de **la** factorización, sino que de **una** factorización aproximada

Se recomienda este método por sobre el criterio de norma Frobenius

UC | Chile