UNIVERSIDAD DEL QUINDÍO
Res.MEN 014915 - 02 AGO 2022
RENOVACIÓN ACREDITACIÓN

Método de machine Learning: Factorización No Negativa de Matrices

- Elkin Yesid Gómez Buriticá
- Daniel Felipe villarraga Gonsales

UNIQUINDÍO
en conexión territorial
www.uniquindio.edu.co

1. La Factorización de Matrices no Negativas (NMF)

NMF se define como la descomposición de una matriz de entrada, $\mathbf{V}(\mathbb{R})$, en otras dos matrices de menor tamaño, $\mathbf{W}(\mathbb{R})$ y $\mathbf{H}(\mathbb{R})$, cuyo producto se aproxima a la primera; es decir,

$$\mathbf{V}_{n\times m} \approx \mathbf{W}_{n\times k}\mathbf{H}_{k\times m} \tag{1}$$

Esta operación equivale a representar cada elemento mediante la *combinación lineal* de un conjunto de k factores en la proporción indicada por otros tantos coeficientes. Esto es:

$$(\mathbf{V})_{ij} \approx (\mathbf{WH})_{ij} = \sum_{q=1}^{k} w_{iq}h_{qj} \tag{2}$$

De manera que las matrices $\mathbf{W}$ y $\mathbf{H}$ almacenan, todos los factores y coeficientes necesarios para reconstruir todos los elementos de $\mathbf{V}$.

2. El algoritmo de NMF

El algoritmo consiste en inicializar las matrices **W** y **H** con valores aleatorios mayores o iguales a cero para modificarlas iterativamente hasta que su producto se aproxime a **V** .

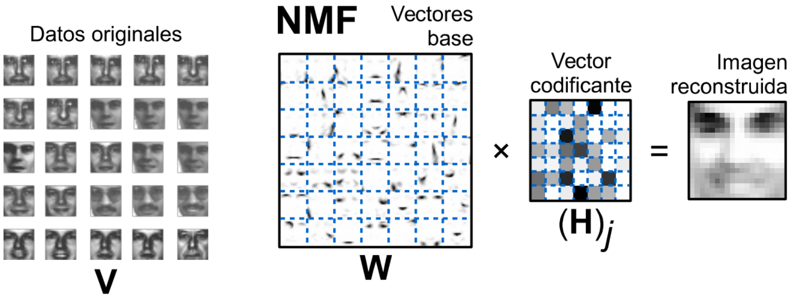


Figura 1: Ejemplo de factorización NMF.

Se basa en la minimización de la siguiente función objetivo:

$$F = \sum_{ij} [v_{ij} \log\left((\mathbf{WH})_{\mathbf{ij}}\right) - (\mathbf{WH})_{\mathbf{ij}}] \tag{3}$$

Esta función se obtiene al representar **WH** como una copia de **V** a la que se añaden pequeñas fluctuaciones, conocidas como ruido de Poisson o de dispar. Las reglas que se derivan son las siguientes:

$$\boxed{h_{pj} \leftarrow h_{pj} \sum_{i} \left(w_{ip} \frac{v_{ij}}{(\mathbf{WH})_{ij}} \right)} \quad \boxed{w_{ip} \leftarrow w_{ip} \sum_{j} \left(\frac{v_{ij}}{(\mathbf{WH})_{ij}} h_{pj} \right)}$$
$$\boxed{w_{ip} \leftarrow \frac{w_{ip}}{\sum_r w_{rp}}} \tag{4}$$

UNIQUINDÍO
en conexión territorial

la función objetivo a minimizar es la Distancia euclídea entre **V** y **WH**

$$||\mathbf{V} - \mathbf{WH}||^2 = \sum_{ij}[v_{ij} - (\mathbf{WH})_{ij}]^2 \tag{5}$$

de la que se derivan las siguientes reglas:

$$\boxed{h_{pj} \leftarrow h_{pj}\frac{(\mathbf{W}^T * \mathbf{V})_{pj}}{(\mathbf{W}^T * \mathbf{WH})_{pj}}} \quad \boxed{w_{ip} \leftarrow w_{ip}\frac{(\mathbf{V} * \mathbf{H}^T)_{ip}}{(\mathbf{WH} * \mathbf{H}^T)_{ip}}} \tag{6}$$

2.3 NMF - Divergencia

Esta variante debe su nombre al empleo de la siguiente función objetivo:

$$D(\mathbf{V}||\mathbf{WH}) = \sum_{ij} \left[v_{ij} \ \log \left(\frac{v_{ij}}{(\mathbf{WH})_{ij}} \right) - v_{ij} + (\mathbf{WH})_{ij} \right] \tag{7}$$

Se trata de una adaptación de la Divergencia de Kullback-Leible.
Las reglas de actualización que se obtienen son las siguientes:

$$\boxed{h_{pj} \leftarrow h_{pj} \frac{\sum_i \left(w_{ip} \frac{v_{ij}}{(\mathbf{WH})_{ij}} \right)}{\sum_i w_{ip}}} \qquad \boxed{w_{ip} \leftarrow w_{ip} \frac{\sum_j \left(\frac{v_{ij}}{(\mathbf{WH})_{ij}} h_{pj} \right)}{\sum_j h_{pj}}} \tag{8}$$

UNIQUINDÍO
en conexión territorial

2.4 Método del gradiente descendente

Se parte de un cierto punto x_0 y, en cada iteración, se calcula un nuevo punto que esté situado en la dirección marcada por el vector gradiente negativo del punto anterior. Es decir, actualizar cada punto mediante la siguiente regla aditiva:

$$x_{t+1} \quad \leftarrow \quad x_t - \eta_t \nabla F(x_t) \tag{9}$$

la coordenada del punto correspondiente a la matriz $\mathbf{H}$ se actualizaría de la siguiente manera:

$$h_{pj} \quad \leftarrow \quad h_{pj} - \eta_{pj} \frac{\partial}{\partial h_{pj}} D(\mathbf{V}||\mathbf{WH}) \tag{10}$$

Por tanto, se deriva respecto de **H**:
Derivada de $-v_{ij}\log((\mathbf{WH})_{\mathbf{ij}})$

$$\frac{\partial}{\partial h_{pj}}(-v_{ij}\log((\mathbf{WH})_{\mathbf{ij}}) = -v_{ij}\frac{1}{(\mathbf{WH})_{\mathbf{ij}}}\frac{\partial(\mathbf{WH})_{\mathbf{ij}}}{\partial h_{pj}} = -v_{ij}\frac{1}{(\mathbf{WH})_{\mathbf{ij}}}w_{ip} \tag{11}$$

Derivada de $(\mathbf{WH})_{\mathbf{ij}}$,

$$\frac{\partial}{\partial h_{pj}}(\mathbf{WH})_{\mathbf{ij}} = \frac{\partial}{\partial h_{pj}}\sum_q w_{iq}h_{qj} = w_{ip} \tag{12}$$

2.4 Método del gradiente descendente

Por tanto se obtiene.

$$\frac{\partial}{\partial h_{pj}} D(\mathbf{V}||\mathbf{WH}) = \sum_i w_{ip} - \sum_i \frac{w_{ip} v_{ij}}{\sum_{q=1}^{k} w_{iq} h_{qj}} \tag{13}$$

y se aplica la regla aditiva, intercambiando los signos, para obtener:

$$h_{pj} \quad \leftarrow \quad h_{pj} + \eta_{pj} \left[\sum_i \left(w_{ip} \frac{v_{ij}}{(\mathbf{WH})_{ij}} \right) - \sum_i w_{ip} \right] \tag{14}$$

siendo el coeficiente η_{pj} un valor pequeño para garantizar que la función $D(\mathbf{V}||\mathbf{WH})$ decrece.

2.4 Método del gradiente descendente

Con la matriz **W**: se sigue un proceso similar, obteniendo la siguiente regla:

$$w_{ip} \quad \leftarrow \quad w_{ip} + \eta_{ip} \left[\sum_i \left(\frac{v_{ij} h_{pj}}{(\mathbf{WH})_{ij}} \right) - \sum_i h_{pj} \right] \tag{15}$$

La desventaja de este método es su lentitud para converger, especialmente cerca del mínimo local, donde se produce un cierto "zigzagueo". Es por ello que se suelen emplear, en su lugar, las denominadas Reglas multiplicativas,

$$\eta_{pj} = \frac{h_{pj}}{\sum_i w_{ip}} \qquad \text{y} \qquad \eta_{ip} = \frac{w_{ip}}{\sum_j h_{pj}} \tag{16}$$

2.4 Método del gradiente descendente

Partimos de la forma aditiva:

$$h_{pj} \leftarrow h_{pj} + \eta_{pj}(A - B), \quad \text{donde } A = \sum_i w_{ip} \frac{v_{ij}}{(\mathbf{WH})_{\mathbf{ij}}}, \; B = \sum_i w_{ip} \tag{17}$$

si tomamos $\eta_{pj} = \frac{h_{pj}}{B}$, entonces

$$h_{pj} \leftarrow h_{pj} + \frac{h_{pj}}{B}(A - B) = h_{pj}\left(1 + \frac{A - B}{B}\right) = h_{pj}\left(\frac{A}{B}\right) \tag{18}$$

asi obteniendo que

$$h_{pj} \leftarrow h_{pj} \frac{\sum_i w_{ip} \frac{v_{ij}}{(\mathbf{WH})_{ij}}}{\sum_i w_{ip}} \quad \mathbf{y} \quad w_{ip} \leftarrow w_{ip} \frac{\sum_i h_{pj} \frac{v_{ij}}{(\mathbf{WH})_{ij}}}{\sum_i h_{pj}} \tag{19}$$

Incluyendo una nueva matriz, $\mathbf{S} \in \mathbf{M_{k\times k}}(\mathbb{R})$, , pasa a descomponerse como $V \approx WSH$, definiéndose esta nueva matriz, **S**, de la siguiente manera:

$$\mathbf{S} = (1-\theta)\mathbf{I} + \frac{\theta}{k}\mathbf{1} \qquad (20)$$

En este caso cuando $\theta = 0$ **S** se vuelve la identidad y cuando θ se acerca a 1 los valores de **S** se vuelven la media de cada fila y cada columna siendo este el estado de menor *raleza* o maor suavidad.

las reglas quedan de la siguiente manera:

$$\boxed{h_{pj} \leftarrow h_{pj} \frac{\sum_i \left((\mathbf{WS})_{ip} \frac{v_{ij}}{((\mathbf{WS})\mathbf{H})_{ij}} \right)}{\sum_i (\mathbf{WS})_{ip}}} \quad \boxed{w_{ip} \leftarrow w_{ip} \frac{\sum_j \left(\frac{v_{ij}}{(\mathbf{W}(\mathbf{SH}))_{ij}} (\mathbf{SH})_{pj} \right)}{\sum_j (\mathbf{SH})_{pj}}} \qquad (21)$$

$$\boxed{w_{ip} \leftarrow \frac{w_{ip}}{\sum_r w_{rp}}}$$

UNIQUINDÍO
en conexión territorial

Referencias

[1] D. D. Lee y H. S. Seung: Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 10/1999.

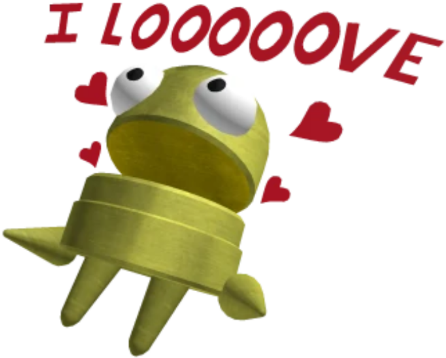
[2] D. D. Lee y H. S. Seung: Algorithms for Non-negative Matrix Factorization. En T. K. Leen, T. G. Dietterich y V. Tresp (eds.): Adv. Neural Inf. Process. Syst. 13 (NIPS 2000), págs. 556–562. The MIT Press, Cambridge, MA, 5/2001.

[3] G. H. Golub y C. F. Van Loan: Matrix Computations. Johns Hopkins University Press, Baltimore, 4a ed., 12/2012.

[4] M. W. Berry, S. T. Dumais y T. A. Letsche: Computational Methods for Intelligent Information Access. En Proc. ACM/IEEE Conf. Supercomput., pág. 20, New York, NY, 12/1995. San Diego Supercomputer Center, ACM.

[5] S. Kullback y R. A. Leibler: On Information and Sufficiency. Ann. Math. Stat., 22(1):79– 86, 3/1951.

[6] A. Pascual-Montano, J. M. Carazo, K. Kochi y cols.: Nonsmooth nonnegative matrix factorization (nsNMF). IEEE Trans. Pattern Anal. Mach. Intell., 28(3):403–15, 3/2006



Gracias :)

UNIVERSIDAD
DEL QUINDÍO ®

ALTA
CALIDAD
Res.MEN 014915 - 02 AGO 2022

CERTIFICADO
POR SU COMPROMISO
RESPONSABILIDAD SOCIAL
FENALCO SOLIDARIO