Non-negative matrix factorization (NMF)

Daniel Arruda Ponte

Especificação

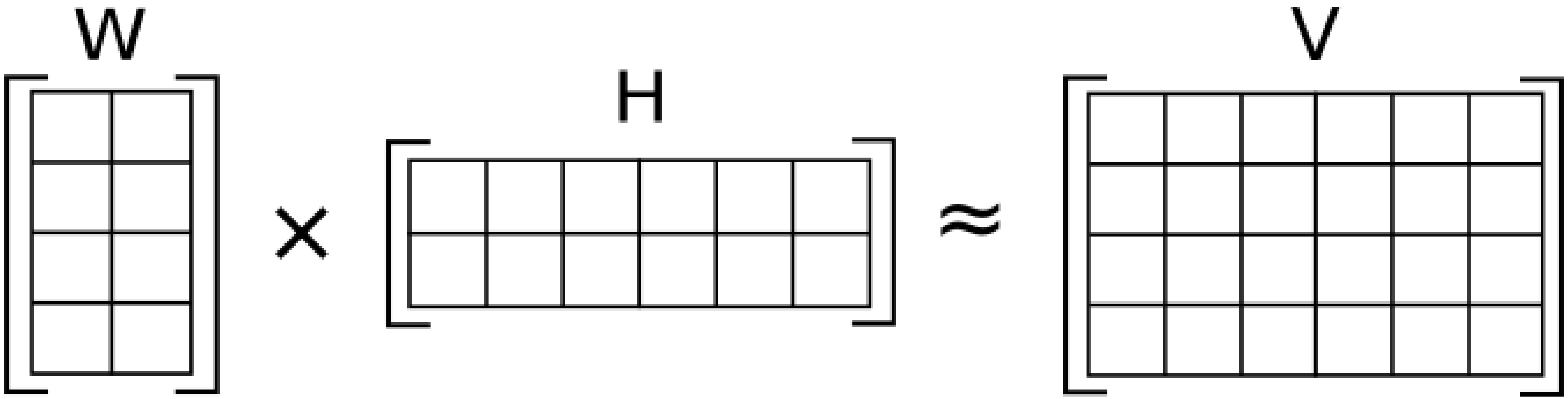
Matrizes não-negativas

$$V_{ij} \geq 0, \forall i, j$$

$$V \in \mathbb{R}_{+}^{m \times n}$$

$$V \geq 0$$

Fatoração



Problema aproximado

$$[WH] = \underset{W\geq 0, H\geq 0}{\arg\min}\, f(W,H) = \frac{1}{2}||V - WH||_F^2$$

Algoritmo

Gradiente Descendente Alternado

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \nabla_x f(x^{(k)})$$

- $k \in \mathbb{N}$ o número da iteração
- $x^{(k)}$ é a variável na iteração k
- $x^{(k+1)}$ é a variável na iteração $k+1$
- $\alpha^{(k)} \geq 0$ é o tamanho do passo na iteração k
- $\nabla_x f(x^{(k)})$ é o gradiente da função f com respeito à x

Gradiente Descendente Alternado

$$W^{(k+1)} = W^{(k)} - \alpha_W^{(k)} \nabla_W f(W^{(k)}, H^{(k)})$$
$$H^{(k+1)} = H^{(k)} - \alpha_H^{(k)} \nabla_H f(W^{(k+1)}, H^{(k)})$$

$$f(W^{(k+1)}, H^{(k+1)}) \leq f(W^{(k)}, H^{(k)})$$

Propriedades base

$$
\begin{aligned}
\nabla_{\mathbf{X}} \operatorname{tr}(\mathbf{A}\mathbf{X}) &= \mathbf{A}^T \\
\nabla_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^T\mathbf{A}) &= \mathbf{A} \\
\nabla_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) &= (\mathbf{A} + \mathbf{A}^T)\mathbf{X} \\
\nabla_{\mathbf{X}} \operatorname{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^T) &= \mathbf{X}(\mathbf{A}^T + \mathbf{A})
\end{aligned}
$$

Propriedades utilizadas

$$||X||_F^2 = tr(X^T X) \tag{1}$$

$$tr(X) = tr(X^T) \tag{2}$$

$$\nabla_x tr(AXB) = A^T B^T \tag{3}$$

$$\nabla_x tr(X^T AX) = (A + A^T)X \tag{4}$$

$$\nabla_x tr(B^T X^T XB) = 2XBB^T \tag{5}$$

Desenvolvendo a função objetivo

$$
\begin{aligned}
||V - WH||_F^2 &\overset{(1)}{=} tr\left\{(V - WH)^T(V - WH)\right\} \\
&= tr\left\{V^TV - V^TWH - (WH)^TV + (WH)^TWH\right\} \\
&\overset{(2)}{=} tr\left\{V^TV - 2W^TH^TV + H^TW^TWH\right\} \\
&\overset{(1)}{=} ||V||_F^2 - 2tr(W^TH^TV) + tr(H^TW^TWH)
\end{aligned}
$$

Desenvolvendo a função objetivo

$$f(W,H) = \frac{1}{2}||V - WH||_F^2 = ||V||_F^2 - tr(W^TH^TV) + \frac{1}{2}tr(H^TW^TWH)$$

Cálculo do gradiente

$$\nabla_W f \overset{(3,5)}{=} (WH - V)H^T$$

$$\nabla_H f \overset{(3,4)}{=} W^T(WH - V)$$

$$||V||_F^2 - 2tr(W^T H^T V) + tr(H^T W^T W H)$$

$$\nabla_x tr(AXB) = A^T B^T \tag{3}$$

$$\nabla_x tr(X^T AX) = (A + A^T)X \tag{4}$$

$$\nabla_x tr(B^T X^T XB) = 2XBB^T \tag{5}$$

Fórmula de atualização

$$W^{(k+1)} = W^{(k)} - \alpha_W^{(k)}(WH - V)H^T$$
$$H^{(k+1)} = H^{(k)} - \alpha_H^{(k)}W^T(WH - V)$$

Tamanho do passo

Lee e Seung 2001

$$[\alpha_W]_{ij} = \frac{W_{ij}}{[WHH^T]_{ij}}, [\alpha_H]_{ij} = \frac{H_{ij}}{[W^TWH]_{ij}}$$

$$
\begin{aligned}
W_{ij} &= W_{ij} - [\alpha_W]_{ij}[(WH - V)H^T]_{ij} \\
&= W_{ij} - \frac{W_{ij}}{[WHH^T]_{ij}}[(WH - V)H^T]_{ij} \\
&= W_{ij} - \frac{[W(WH - V)H^T]_{ij}}{[WHH^T]_{ij}} \\
&= W_{ij}\left(1 - \frac{[(WH - V)H^T]_{ij}}{[WHH^T]_{ij}}\right) \\
&= W_{ij}\left(\frac{[WHH^T]_{ij}}{[WHH^T]_{ij}} - \frac{[(WH - V)H^T]_{ij}}{[WHH^T]_{ij}}\right) \\
&= W_{ij}\frac{[VH^T]_{ij}}{[WHH^T]_{ij}}
\end{aligned}
$$

$$
\begin{aligned}
H_{ij} &= H_{ij} - [\alpha_H]_{ij}[W^T(WH - V)]_{ij} \\
&= H_{ij} - \frac{H_{ij}}{[W^TWH]_{ij}}[W^T(WH - V)]_{ij} \\
&= H_{ij}\left(1 - \frac{[W^T(WH - V)]_{ij}}{[W^TWH]_{ij}}\right) \\
&= H_{ij}\left(\frac{[W^TWH]_{ij}}{[W^TWH]_{ij}} - \frac{[W^T(WH - V)]_{ij}}{[W^TWH]_{ij}}\right) \\
&= H_{ij}\frac{[W^TV]_{ij}}{[W^TWH]_{ij}}
\end{aligned}
$$

Fórmula de atualização

- A cada iteração a função objetivo não aumenta seu valor
- O algoritmo não garante a convergência para um mínimo global
- Para manter a não-negatividade, o tamanho do passo realiza operações elemento a elemento com valores não-negativos

$$W^{(k+1)} = W^{(k)} \circ \frac{V {H^{(k)}}^T}{W H^{(k)} {H^{(k)}}^T}$$

$$H^{(k+1)} = H^{(k)} \circ \frac{{W^{(k)}}^T V}{{W^{(k)}}^T W^{(k)} H^{(k)}}$$

Convergência

Teorema

Lee e Seung 2001

$$\frac{1}{2}||V - WH||_F \qquad f(W^{(k+1)}, H^{(k+1)}) \leq f(W^{(k)}, H^{(k)})$$

$$W^{(k+1)} = W^{(k)} \circ \frac{VH^{(k)^T}}{WH^{(k)}H^{(k)^T}}$$

$$H^{(k+1)} = H^{(k)} \circ \frac{W^{(k)^T}V}{W^{(k)^T}W^{(k)}H^{(k)}}$$

Função objetivo

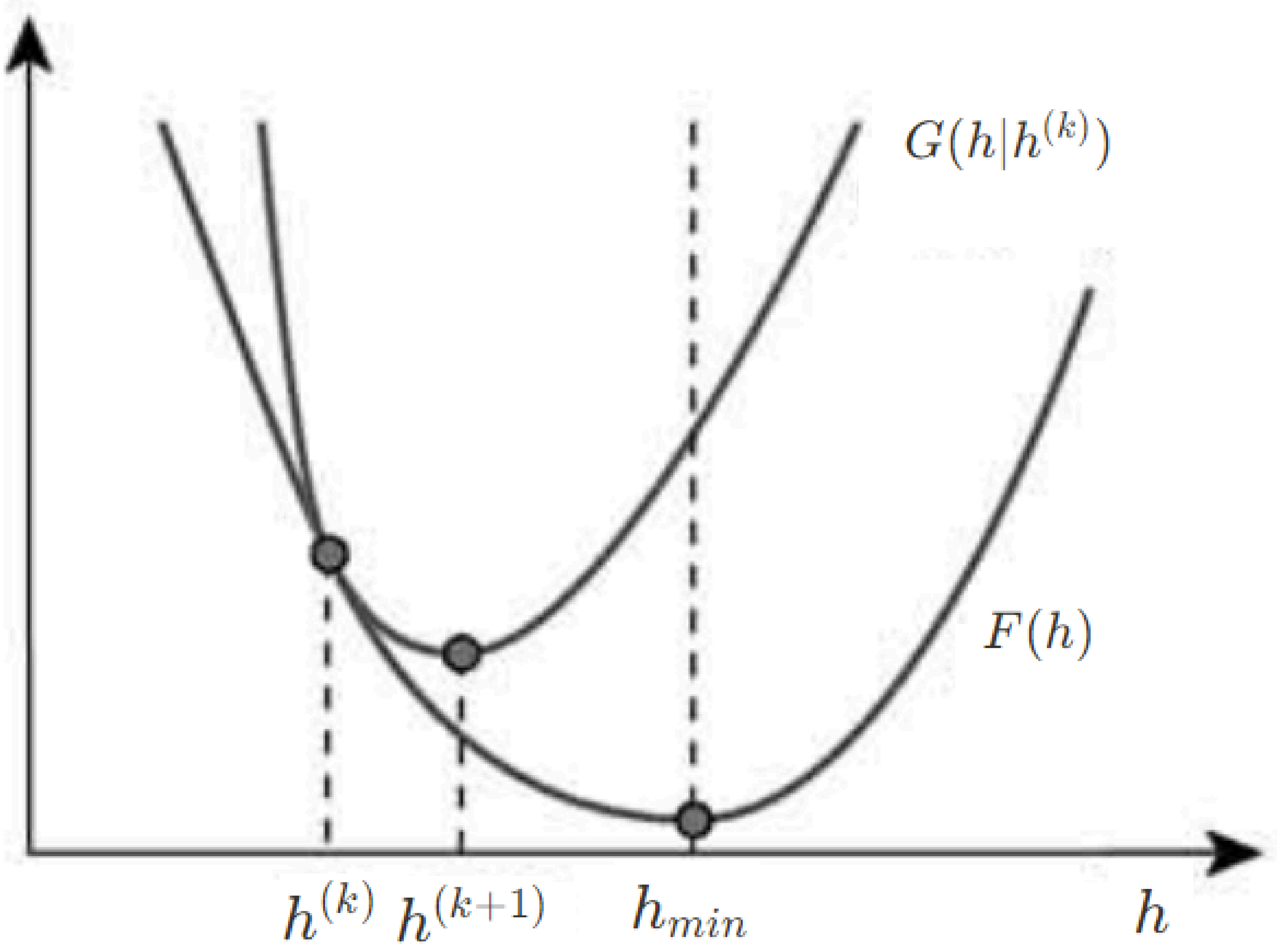
$$f(h) = \frac{1}{2}||Wh - v||_2^2$$

$$F(h) = f(h^{(k)}) + \nabla f(h^{(k)})^T (h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T \nabla^2 f(h^{(k)})(h - h^{(k)})$$

Técnica de otimização

- Majorization Minimization
- Minimização de uma função auxiliar (limite superior)

$$G(h|h^{(k)}) \geq F(h), G(h|h) = F(h)$$



Lema 1

Se ***G*** é uma função auxiliar, então o valor de ***F*** não cresce quando a seguinte atualização for feita:

$$h^{(k+1)} = \arg\min_{h} G(h|h^{(k)})$$

$$F(h^{(k)}) = G(h^{(k)}|h^{(k)}) \geq G(h^{(k+1)}|h^{(k)}) \geq F(h^{(k+1)})$$

Definição de *G*

$$\begin{aligned}\frac{1}{2}||Wh - v||_2^2 &= \frac{1}{2}(Wh - v)^T(Wh - v)\\ &= \frac{1}{2}(h^TW^T - v^T)(Wh - v)\\ &= \frac{1}{2}(h^TW^TWh - h^TW^Tv - v^TWh - v^Tv)\\ &= \frac{1}{2}(h^TW^TWh - 2h^TW^Tv + v^Tv)\end{aligned}$$

$$\nabla_h f(h) = W^TWh - W^Tv \qquad \nabla_h^2 f(h) = W^TW$$

Definição de *G*

$$G(h|h^{(k)}) = f(h^{(k)}) + \nabla f(h^{(k)})^T(h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T K(h^{(k)})(h - h^{(k)})$$

$$K(h^{(k)}) \geq \nabla^2 f(h^{(k)})$$

$$K_{ab}(h^{(k)}) = Diag\left(\frac{[W^T W h^{(k)}]_a}{[h^{(k)}]_a}\right)_{ab}$$

Definição de *G*

$$\begin{aligned}
Diag\left(\frac{[Ax]_i}{[x]_i}\right) &= Diag\left(\frac{[y]_i}{[x]_i}\right)\\
&= \frac{Diag(y)}{Diag(x)}\\
&= D_x^{-1} D_y\\
&= \begin{bmatrix} \frac{y_1}{x_1} & 0 & ... & 0 \\ 0 & \frac{y_1}{x_1} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \frac{y_d}{x_d} \end{bmatrix}
\end{aligned}$$

Condição de função auxiliar

$$F(h) = f(h^{(k)}) + \nabla f(h^{(k)})^T (h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T W^T W (h - h^{(k)})$$

$$G(h|h^{(k)}) = f(h^{(k)}) + \ldots + \frac{1}{2}(h - h^{(k)})^T Diag\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right)(h - h^{(k)})$$

$$G(h|h^{(k)}) \geq F(h)$$

$$(h - h^{(k)})^T \left(Diag\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right) - W^T W\right)(h - h^{(k)}) \geq 0$$

Condição de função auxiliar

$$L_{ab}(h^{(k)}) = h_a^{(k)} \left(K(h^{(k)}) - W^T W \right)_{ab} h_b^{(k)}$$

$$\begin{aligned}
v^T L v &= \sum_{ab} v_a L_{ab} v_b \\
&= \sum_{ab} v_a h_a^{(k)} \left(Diag\left(\frac{[W^T W h^{(k)}]_a}{[h^{(k)}]_a} \right)_{ab} - W^T W_{ab} \right) h_b^{(k)} v_b \\
&= \sum_{ab} h_a^{(k)} (W^T W)_{ab} h_b^{(k)} v_a^2 - v_a h_a^{(k)} (W^T W)_{ab} h_b^{(k)} v_b \\
&= \sum_{ab} (W^T W)_{ab} h_a^{(k)} h_b^{(k)} \left[\frac{1}{2} v_a^2 + \frac{1}{2} v_b^2 - v_a v_b \right] \\
&= \frac{1}{2} \sum_{ab} (W^T W)_{ab} h_a^{(k)} h_b^{(k)} (v_a - v_b)^2 \\
&\geq 0
\end{aligned}$$

Tamanho do passo

$$\frac{\partial G(h|h^{(k)})}{\partial h} = \nabla f(h^{(k)}) + K(h - h^{(k)})$$

$$\frac{\partial G(h|h^{(k)})}{\partial h} = 0$$

$$h = h^{(k)} - K^{-1}\nabla f(h^{(k)})$$

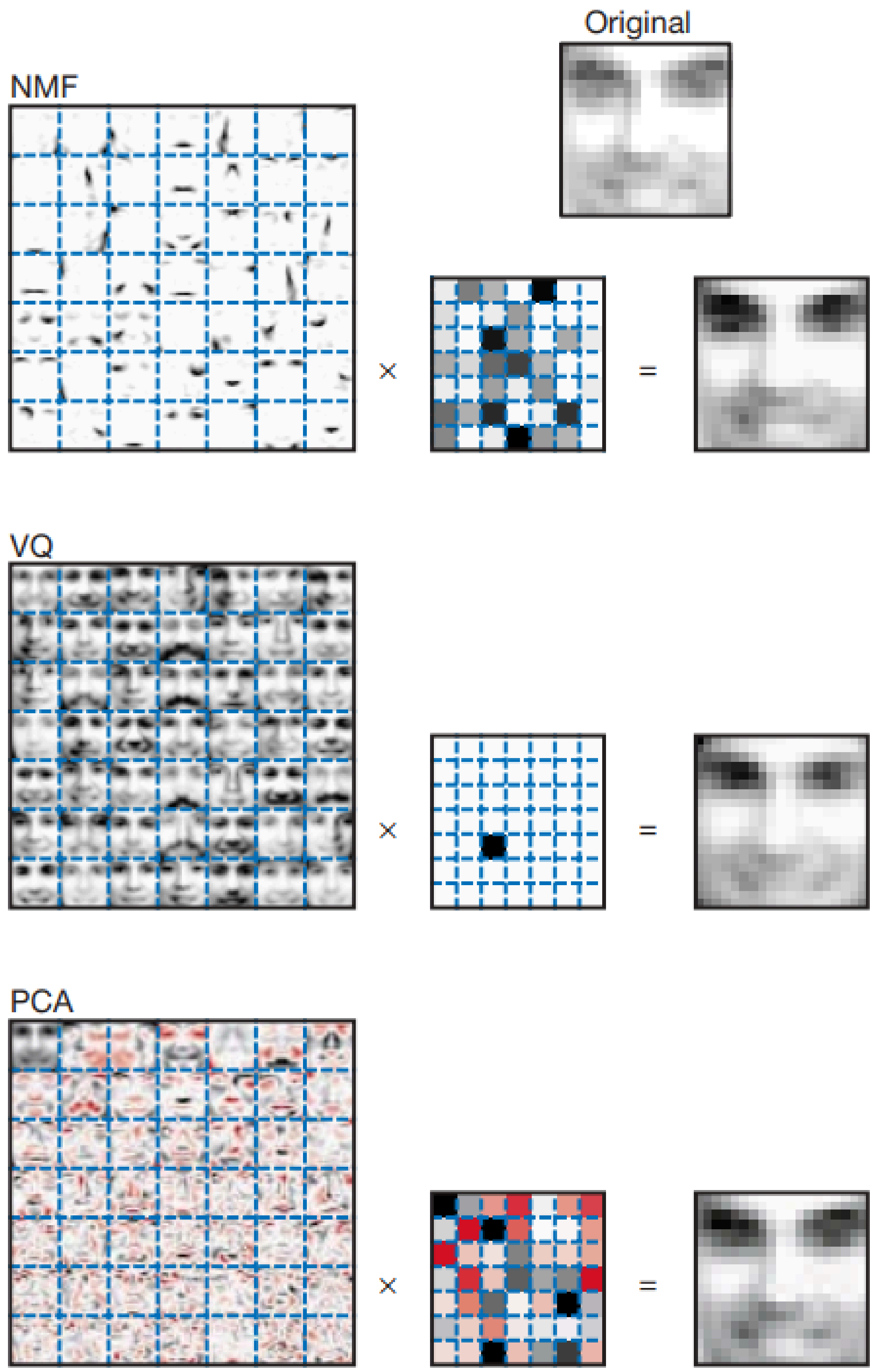
Tamanho do passo

$$
\begin{aligned}
h &= h^{(k)} - K^{-1}\nabla f(h^{(k)}) \\
&= h^{(k)} - Diag\left(\frac{[h^{(k)}]_i}{[W^TWh^{(k)}]_i}\right)W^T(Wh^{(k)} - v) \\
&= h^{(k)} - h^{(k)} \otimes \frac{[W^T(Wh^{(k)} - v)]_{ij}}{[W^TWh^{(k)}]_{ij}} \\
&= \frac{h^{(k)}W^TWh^{(k)}}{W^TWh^{(k)}} - \frac{h^{(k)}(W^TWh^{(k)} + W^Tv)}{W^TWh^{(k)}} \\
&= h^{(k)} \otimes \frac{[W^Tv]_{ij}}{[W^TWh^{(k)}]_{ij}}
\end{aligned}
$$

Aplicações

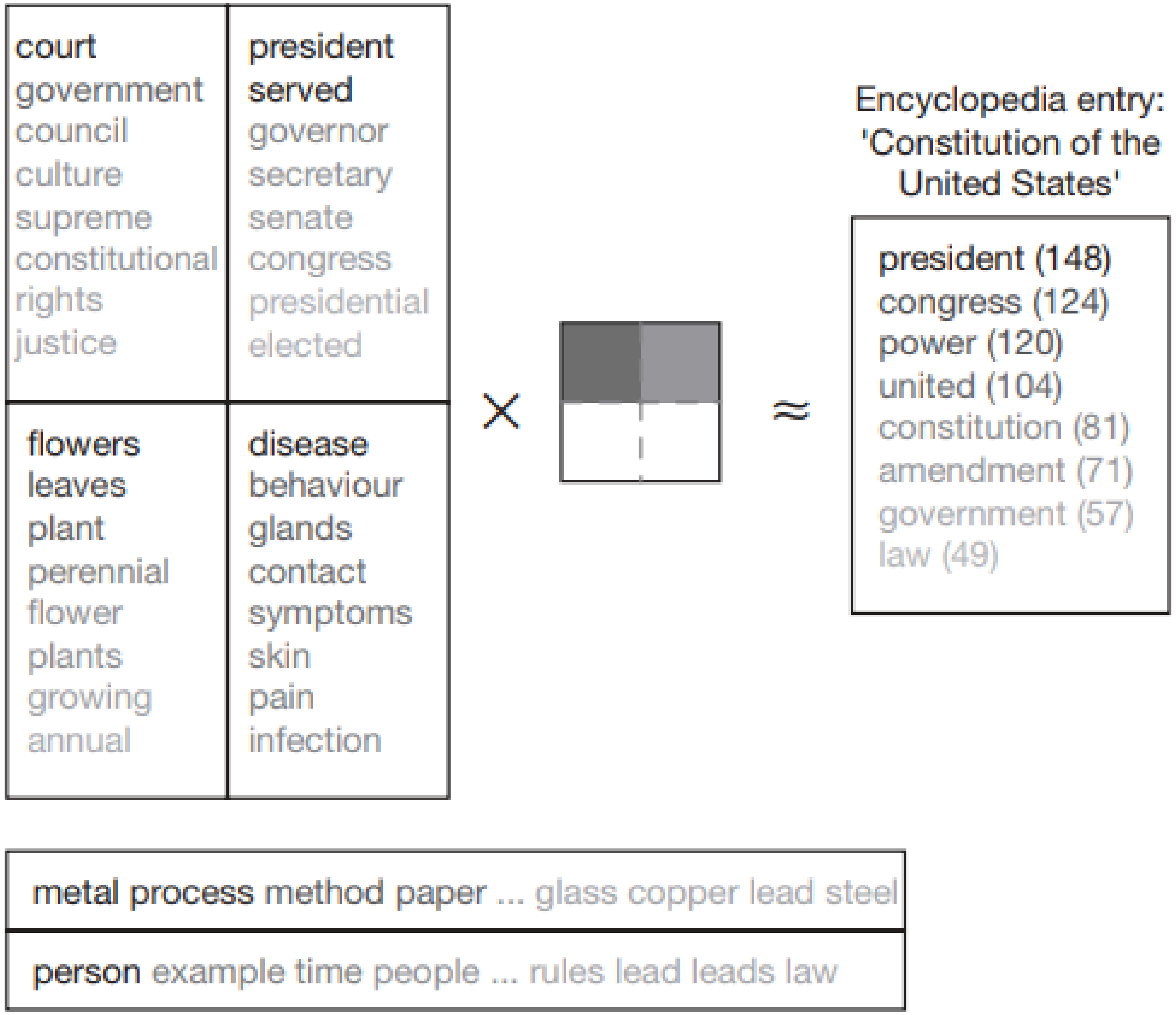
Imagem

- Imagens de rostos
- 2429 imagens
- 19 X 19 pixels
- Aproximação de posto 49
- Preto e branco



Texto

- Artigos do Grolier
- 30991 artigos
- Vocabulário de 15276 palavras
- Aproximação de posto 200
- Vetorização por frequência



Referências

ANG, A. Convergence analysis of nmf algorithm. Mathematique et recherche operationnelle, UMONS, Belgium, 2018. First draft: June 6, 2017; Last update: February 27, 2018. Disponível em:⟨Ar 1.pdf⟩.

ANG, A. Non-negative matrix factorization and multiplicative update. Mathematique et recherche operationnelle, UMONS, Belgium, 2019. First draft: June 6, 2017; Last update: February 25, 2019. Disponível em: ⟨Ar 1.pdf⟩.

BURRED, J. J. Detailed derivation of multiplicative update rules for nmf. Paris, France, 2014.

CONTRIBUTORS, W. Non-negative Matrix Factorization. 2024. ⟨Ar matrix factorization⟩. Accessed: 2024-07-13.

DUCHI, J. Properties of the trace and matrix derivatives. Available electronically at Ar˜jduchi/projects/matrix prop.pdf, 2007.

LEE, D.; SEUNG, H. S. Algorithms for non-negative matrix factorization. Advances in neural information processing systems, v. 13, 2000.

LEE, D. D.; SEUNG, H. S. Learning the parts of objects by non-negative matrix factorization. nature, Nature Publishing Group UK London, v. 401, n. 6755, p. 788–791, 1999.