Non-negative matrix factorization (NMF)

- Directly applying SVD to the term-document matrix enables us to compute distances between documents (and by extension queries) in a latent space, but it comes with a few undesirable properties
  - The right singular vectors (which do the projecting) have the ability to apply negative coefficients. Is this useful? Could we express the same thing using positive coefficients applied to antonyms? For example: one might argue that these two bases convey the same meaning:

$$\mathbf{u} = (0.5)swim + (0.7)sun + (-1.0)rain + (-0.2)dark$$

$$\mathbf{u} = (0.5)swim + (1.4)sun + (0.01)rain + (0.01)dark$$

  - There is no sparsity constraint on $\hat{\mathbf{U}}$, which makes each dimension of the resultant latent feature less interpretable w.r.t. the original feature dimensions.

- Non-Negative Matrix Factorization addresses these issues by imposing positivity and sparsity constraints on the matrix factors.

NMF cost functions

Factorization: $\mathbf{X} = \mathbf{WS}$ where $\mathbf{X} \in \mathbb{R}^{M \times N}$ $\mathbf{W} \in \mathbb{R}^{M \times K}$ $\mathbf{S} \in \mathbb{R}^{K \times N}$ $K < N$

NMF loss function: $$\mathbf{W}, \mathbf{S} = \min_{\mathbf{W},\mathbf{S} \geq 0} \mathbf{L}(\mathbf{X}, \mathbf{W}, \mathbf{S}) + \alpha\rho\left(\|\mathbf{W}\|_1 + \|\mathbf{S}\|_1\right) + \alpha(1-\rho)\left(\|\mathbf{W}\|_F^2 + \|\mathbf{S}\|_F^2\right)$$

where $\alpha =$ regularization constant

$\rho =$ sparsity constant (a.k.a. L1-ratio)

$\mathbf{L} =$ generic loss function

Frobenius loss: $\mathbf{L}_F = \|\mathbf{X} - \mathbf{WS}\|_F^2$

KL divergence: $$\mathbf{L}_{KL} = \sum_{i,j} X_{i,j}\left(\log \frac{X_{i,j}}{(WS)_{i,j}} - 1\right) + (WS)_{i,j}$$