Derivation of multiplicative updates for iONMF with KL divergence

Suppose we have a non-negative matrix $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n \times p}$ and we want to find two non-negative matrices $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times r}$ and $\mathbf{H} \in \mathbb{R}_{\geq 0}^{r \times p}$ whose product best approximates the original matrix $\mathbf{X}$.

The NMF problem can be solved by minimizing the distance between $\mathbf{X}$ and $\mathbf{WH}$. Two commonly used objective functions are Euclidean distance and Kullback-Leibler (KL) divergence.

The objective function based on Euclidean distance (Lee et al. 2001) is:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{H}}{\text{minimize}} \quad & \|\mathbf{X} - \mathbf{WH}\|_2^2 \\ \text{subject to} \quad & \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0}. \end{aligned}$$

The objective function based on KL divergence (Brunet et al. 2004) is:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{H}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} \log \frac{X_{ij}}{(\mathbf{WH})_{ij}} - X_{ij} + (\mathbf{WH})_{ij} \\ \text{subject to} \quad & \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0}. \end{aligned}$$

An orthogonality regularization term can be added to the objective function (Stražar et al. 2016) to reduce feature redundancy between modules (rows) in matrix $\mathbf{H}$. Greater modularity can be achieved by adding more weight to orthogonality. The original iONMF model uses the following objective function based on Euclidean distance:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{H}}{\text{minimize}} \quad & \|\mathbf{X} - \mathbf{WH}\|_2^2 + \alpha \|\mathbf{HH}^T - \mathbf{I}\|_2^2 \\ \text{subject to} \quad & \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0}. \end{aligned}$$

There is no unique solution to the NMF problem and the objective function is non-convex. A computationally efficient algorithm is to initialize $\mathbf{W}$ and $\mathbf{H}$ with non-negative values and then use multiplicative update rules to update $\mathbf{W}$ and $\mathbf{H}$ alternatively until convergence. The multiplicative update rules are based on gradient descent but ensure that the non-negativity constraint is satisfied if all the matrices are initially non-negative.

According to (Stražar et al. 2016), The multiplicative update rules are:

$$\mathbf{H} \leftarrow \mathbf{H} \circ \sqrt{\frac{\mathbf{W}^T\mathbf{X} + 2\alpha\mathbf{H}}{\mathbf{W}^T\mathbf{W}\mathbf{H} + 2\alpha\mathbf{H}\mathbf{H}^T\mathbf{H}}} \tag{1}$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \sqrt{\frac{\mathbf{X}\mathbf{H}^T}{\mathbf{W}\mathbf{H}\mathbf{H}^T}} \tag{2}$$

We also implemented an objective function by replacing the Euclidean distance in the iONMF model with KL divergence:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{H}}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{p} X_{ij}\log\frac{X_{ij}}{(\mathbf{W}\mathbf{H})_{ij}} - X_{ij} + (\mathbf{W}\mathbf{H})_{ij} + \alpha\|\mathbf{H}\mathbf{H}^T - \mathbf{I}\|_2^2 \\ \text{subject to} \quad & \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0}. \end{aligned}$$

After applying the Lagrange multiplier for non-negative constraint for **W** and **H**, the objective function can be rewritten as:

$$\begin{aligned} L(\mathbf{W}, \mathbf{H}, \lambda_H, \lambda_W) = & \mathbf{1}_n^T \left[\mathbf{X} \circ \log(\frac{\mathbf{X}}{\mathbf{W}\mathbf{H}}) - \mathbf{X} + \mathbf{W}\mathbf{H} \right] \mathbf{1}_p \\ & + \alpha\text{tr}((\mathbf{H}\mathbf{H}^T - \mathbf{I})^T(\mathbf{H}\mathbf{H}^T - \mathbf{I})) \\ & - \mathbf{1}_n^T(\lambda_W \circ \mathbf{W})\mathbf{1}_r \\ & - \mathbf{1}_r^T(\lambda_H \circ \mathbf{H})\mathbf{1}_p \end{aligned}$$

where $\mathbf{1}_{m,n} \in \mathbb{R}^{m\times n}$ is a $m \times n$ matrix of all ones and $\mathbf{1}_n \in \mathbf{R}^n$ is a vector of all ones.

Fixing **H**, the derivative of the Lagrangian with respect to **W** is:

$$\frac{\partial L}{\partial \mathbf{W}} = -\frac{\mathbf{X}}{\mathbf{W}\mathbf{H}}\mathbf{H}^T + \mathbf{1}_{n,p}\mathbf{H}^T - \lambda_W$$

The Karush-Kuhn-Tucker (KKT) conditions must be satisfied:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{W}} = -\frac{\mathbf{X}}{\mathbf{W}\mathbf{H}}\mathbf{H}^T + \mathbf{1}_{n,p}\mathbf{H}^T - \lambda_W = \mathbf{0} \\ \mathbf{W} \geq \mathbf{0} \\ \lambda_W \geq \mathbf{0} \\ \mathbf{W} \circ \lambda_W = \mathbf{0} \end{aligned}$$

We can rewrite $\mathbf{W} \circ \lambda_W = \mathbf{0}$ to a fixed point equation:

$$\mathbf{W}^2 \circ (\mathbf{1}_{n,p}\mathbf{H}^T - (\mathbf{X}/\mathbf{W}\mathbf{H})\mathbf{H}^T) = \mathbf{0}$$

Then we obtain the multiplicative update rule:

$$\mathbf{W} \leftarrow \mathbf{W} \circ \sqrt{\frac{(\mathbf{X}/\mathbf{W}\mathbf{H})\mathbf{H}^T}{\mathbf{1}_{n,p}\mathbf{H}^T}} \tag{3}$$

Fixing $\mathbf{W}$, the derivative of the Lagrangian with respect to $\mathbf{H}$ is:

$$\frac{\partial L}{\partial \mathbf{H}} = -\mathbf{W}^T \frac{\mathbf{X}}{\mathbf{WH}} + \mathbf{W}^T \mathbf{1}_{n,p} - \lambda_H + 4\alpha(\mathbf{H}\mathbf{H}^T\mathbf{H} - \mathbf{H})$$

Similarly, we can derive the multiplicative update rules for $\mathbf{H}$.

$$\mathbf{H} \leftarrow \mathbf{H} \circ \sqrt{\frac{\mathbf{W}^T(\mathbf{X}/\mathbf{WH}) + 4\alpha\mathbf{H}}{\mathbf{W}^T\mathbf{1}_{n,p} + 4\alpha\mathbf{H}\mathbf{H}^T\mathbf{H}}} \tag{4}$$

References

1. Lee DD, Seung HS. Algorithms for Non-negative Matrix Factorization. In: Leen TK, Dietterich TG, Tresp V, editors. Advances in Neural Information Processing Systems 13 [Internet]. MIT Press; 2001. p. 556–562.

2. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. PNAS. 2004;101:4164–9.

3. Stražar M, Žitnik M, Zupan B, Ule J, Curk T. Orthogonal matrix factorization enables integrative analysis of multiple RNA binding proteins. Bioinformatics. 2016;32:1527–35.