Deriving the Block Coordinate Descent Rules for (single-task) NMF with sparsity regularization

Objective

Given an input matrix $X \in \mathbb{R}_{\geq 0}^{n \times m}$ and $k \ll n, m$, the objective is to find $U \in \mathbb{R}_{\geq 0}^{n \times k}$, $V \in \mathbb{R}_{\geq 0}^{m \times k}$ that minimizes:

$$O = \left\| X - UV^\top \right\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{m} \|V[i,:]\|_1 \tag{1}$$

where $V[i,:]$ is the i row of factor matrix V. The regularization term involving λ tries to enforce sparsity in each row of V, ultimately so that only one latent dimension "lights up" for each row of V. Higher λ will enforce stricter sparsity.

Breaking down to task-level and column-level subproblems

The objective is equivalent to minimizing:

$$O = \left\| X - \sum_{j=1}^{k} u_j v_j^\top \right\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{m} \sum_{j=1}^{k} |V[i,j]| \tag{2}$$

$$= \left\| X - \sum_{j=1}^{k} u_j v_j^\top \right\|_{\mathrm{F}}^2 + \lambda \sum_{j=1}^{k} \sum_{i=1}^{m} |V[i,j]| \tag{3}$$

$$= \left\| X - \sum_{j=1}^{k} u_j v_j^\top \right\|_{\mathrm{F}}^2 + \lambda \sum_{j=1}^{k} \|v_j\|_1 \tag{4}$$

Where $u_j \in \mathbb{R}_{\geq 0}^n$ is the jth column vector of U, i.e. $U[:,j]$, and $v_j \in \mathbb{R}_{\geq 0}^m$ is the jth column vector of V, i.e. $V[:,j]$. Now we 'pull out' terms involving the jth column:

$$O = \left\| X - u_j v_j^\top - \sum_{l \neq j} u_l v_l^\top \right\|_{\mathrm{F}}^2 + \lambda \|v_j\|_1 + \lambda \sum_{l \neq j} \|v_l\|_1 \tag{5}$$

Now we'll substitute with $R_j = X - \sum_{l \neq j} u_l v_l^\top$:

$$O = \left\| R_j - u_j v_j^\top \right\|_{\mathrm{F}}^2 + \lambda \|v_j\|_1 + \lambda \sum_{l \neq j} \|v_l\|_1 \tag{6}$$

We can now attempt to optimize u_j and v_j, fixing all other parameters to be constant.

Optimize v_j

To find v_j that minimizes the objective, we find the derivative of the objective with respect to v_j and set it to 0, then solve. First we expand the objective into matrix multiplications:

$$O = \left\| R_j - u_j v_j^\top \right\|_{\mathrm{F}}^2 + \lambda \|v_j\|_1 + C = \mathrm{Tr}\left[\left(R_j - u_j v_j^\top \right)^\top \left(R_j - u_j v_j^\top \right) \right] + \lambda \|v_j\|_1 + C \tag{7}$$

Here C subsumes all elements of the objective that does not involve v_j, since they will be zeroed out when the derivative is taken with respect to v_j. Now we keep expanding:

$$O = \mathrm{Tr}\left[R_j^\top R_j - 2 R_j^\top u_j v_j^\top + \left(u_j v_j^\top \right)^\top \left(u_j v_j^\top \right) \right] + \lambda \|v_j\|_1 + C \tag{8}$$

$$= \mathrm{Tr}\left(R_j^\top R_j \right) - 2\,\mathrm{Tr}\left(R_j^\top u_j v_j^\top \right) + \mathrm{Tr}\left(v_j u_j^\top u_j v_j^\top \right) + \lambda \|v_j\|_1 + C \tag{9}$$

$$= \mathrm{Tr}\left(R_j^\top R_j \right) - 2 \left(R_j^\top u_j \right)^\top v_j + \left(u_j^\top u_j \right) \left(v_j^\top v_j \right) + \lambda \|v_j\|_1 + C \tag{10}$$

$$= \mathrm{Tr}\left(R_j^\top R_j \right) - 2 \left(R_j^\top u_j \right)^\top v_j + \left(u_j^\top u_j \right) \left(v_j^\top v_j \right) + \lambda \mathbf{1}_m^\top v_j + C \tag{11}$$

where $\mathbf{1}_m$ is a vector of size m, filled with 1's. What allows us to expand $\|v_j\|_1$ from (10) to $\mathbf{1}_m^\top v_j$ (11) is the fact that we're enforcing v_j to be non-negative at initialization and at each iteration.

Now the fun part:

$$\frac{\partial O}{\partial v_j} = 0 = 0 - 2R_j^\top u_j + 2v_j u_j^\top u_j + \lambda \mathbf{1}_m + 0 \tag{12}$$

$$= -R_j^\top u_j + u_j^\top u_j v_j + \frac{\lambda}{2}\mathbf{1}_m \tag{13}$$

$$v_j = \frac{R_j^\top u_j - \frac{\lambda}{2}\mathbf{1}_m}{\|u_j\|_2^2} \tag{14}$$

With the non-negativity constraint $v_j \geq 0$, we want $R_j^\top u_j - \frac{\lambda}{2}\mathbf{1}_m \geq 0$, because if $R_j^\top u_j - \frac{\lambda}{2}\mathbf{1}_m < 0$, O will increase in (11). So the finalized update rule is:

$$v_j = \frac{\left[R_j^\top u_j - \frac{\lambda}{2}\mathbf{1}_m\right]_+}{\|u_j\|_2^2} \tag{15}$$

Optimize u_j

We can derive the update rule for u_j more simply. From (10), we take the derivative of O with respect to u_j; all regularization terms will zero out since they do not involve u_j. Hence the final update rule for u_j is:

$$u_j = \frac{[R_j v_j]_+}{\|v_j\|_2^2} \tag{16}$$