Matrix factorization through successive convex approximation

Leonardo Di Nino

03 April 2024

Given the matrix $A \in \mathbb{R}^{m \times n}$, the problem of matrix factorization consists in finding $V \in \mathbb{R}^{m \times L}$ and $W \in \mathbb{R}^{L \times n}$ solving for the following:

$$\min_{(V,W)} \quad ||A - VW||_{\mathcal{F}}^2 + \lambda_V ||V||_{\mathcal{F}}^2 + \lambda_W ||W||_{\mathcal{F}}^2 \tag{1}$$

Since the objective function is block-wise convex with respect to both the matrices, the following SCA procedure naturally arises:

$$\begin{gathered} \hat{V}(V^t, W^t) = \underset{V}{\operatorname{argmin}} ||A - VW^t||_{\mathcal{F}}^2 + \lambda_V ||V||_{\mathcal{F}}^2 \\ \hat{W}(V^t, W^t) = \underset{W}{\operatorname{argmin}} ||A - V^t W||_{\mathcal{F}}^2 + \lambda_W ||W||_{\mathcal{F}}^2 \\ V^{t+1} = V^t + \eta_V [\hat{V}(V^t, W^t) - V^t] \\ W^{t+1} = W^t + \eta_W [\hat{W}(V^t, W^t) - W^t] \end{gathered} \tag{2}$$

The minimization of the surrogates can be computed in close form as follows: being $\psi(V, W)$ the loss function we have (dropping the time index to ease the notation):

$$\begin{gathered} \hat{V}(V^t, W^t) := \frac{\partial \psi(V, W^t)}{\partial V} = 0 \\ 2(VW - A)W^T + 2\lambda_V V = 0 \\ VWW^T - AW^T + \lambda_V V = 0 \\ V(WW^T + \lambda_V \mathbb{I}_L) = AW^T \\ \hat{V} = AW^T (WW^T + \lambda_V \mathbb{I}_L)^{-1} \end{gathered} \tag{3}$$

Similarly for the other block:

$$\begin{gathered} \hat{W}(V^t, W^t) := \frac{\partial \psi(V^t, W)}{\partial W} = 0 \\ 2V^T (VW - A) + 2\lambda_W W = 0 \\ V^T VW - V^T A + \lambda_W W = 0 \\ (V^T V + \lambda_W \mathbb{I}_L) W = V^T A \\ \hat{W} = (V^T V + \lambda_W \mathbb{I}_L)^{-1} V^T A \end{gathered} \tag{4}$$