Stochastic Gradient Descent, De-Mystified

Since we are implementing stochastic gradient descent, we want to compute the gradient of the following expression w.r.t $u_n, v_m \forall n \in U, m \in V$ for a given y_{ij}.

$$l = \min_{U,V,a,b} \frac{\lambda}{2N}(||U||^2_{Fro} + ||V||^2_{Fro} + ||a||^2_{Fro} + ||b||^2_{Fro}) + \frac{1}{2}(y_{ij} - \mu - (u_i^\intercal v_j + a_i + b_j))^2$$

We have the following process for stochastic gradient descent:

1. Randomly pick a y_{ij}
2. Compute the gradient of the error function $l(i, j)$ with respect to every column in U and every column in V, and every entry in a and b.
3. Subtract $\eta\nabla$ from U V, a, and b
4. Repeat steps 1-3 until $\eta\nabla < \epsilon$.

We can compute the following partial derivatives:

$$\begin{aligned}
\frac{\partial l}{\partial_{a_k}} &= \frac{\lambda}{N} a_k - \mathbb{1}^{i=k}((y_{ij} - \mu) - (u_i^\intercal v_j + a_i + b_j)) \\
\frac{\partial l}{\partial_{b_j}} &= \frac{\lambda}{N} b_j - \mathbb{1}^{j=k}((y_{ij} - \mu) - (u_i^\intercal v_j + a_i + b_j)) \\
\frac{\partial l}{\partial_{u_k}} &= \frac{\lambda}{n} u_k - \mathbb{1}^{k=i} v_j((y_{ij} - \mu) - (u_i^\intercal v_j + a_i + b_j)) \\
\frac{\partial l}{\partial_{v_k}} &= \frac{\lambda}{n} v_k + \mathbb{1}^{k=j} u_i((y_{ij} - \mu) - (u_i^\intercal v_j + a_i + b_j))
\end{aligned}$$

Choice of Parameters

As per the set, we want to over-specify the rank of U and V, so we will pick $k = 20$ given the concrete suggestion in the problem. We will run the risk of iterating too many times $\eta = 0.001$.