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}.

$$L = \min_{U,V} \frac{\lambda}{2N}(||U||_{Fro}^2 - ||V||_{Fro}^2) + \frac{1}{2}(y_{ij} - u_i^\intercal v_j)^2$$

We have the following process for stochastic gradient descent:

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

For every column of U, i.e. $\forall u_k \in U$, we compute

$$\frac{\partial}{\partial_{u_k}}(L) = \frac{\lambda}{n}u_k + \mathbb{1}^{u_k = u_i} v_j(y_{ij} - u_i^\intercal v_j)$$

Similarly, for each volumn of V, i.e. $\forall v_k \in V$, we compute

$$\frac{\partial}{\partial_{v_k}}(L) = \frac{\lambda}{n}v_k + \mathbb{1}^{v_k = v_j} u_i(y_{ij} - u_i^\intercal v_j)$$

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 $\eta = 0.001$.