Gradient Descent for Matrix Factorization

Funk SVD

$$\hat{Y} = U \cdot V^T$$

U is a matrix of user embeddings

$$U = \begin{bmatrix} - & u_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_i & - \\ & \vdots & \\ - & u_{n_u} & - \end{bmatrix} = \begin{array}{c} \\ i \end{array} \begin{array}{c} \quad k \\ \begin{bmatrix} u_{ik} & \\ & \\ & \end{bmatrix} \end{array} K$$

$$V = \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ & \vdots & \\ - & v_j & - \\ & \vdots & \\ - & v_{n_m} & - \end{bmatrix} = \begin{array}{c} \\ j \end{array} \begin{array}{c} k^{th} \\ \begin{bmatrix} & | & \\ & v_{jk} & \\ & & \end{bmatrix} \end{array}$$

$$\begin{bmatrix} | \\ v_j \\ | \end{bmatrix}$$

Mask Matrix

$R = \{r_{ij}\}_{ij}$

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is not empty} \\ 0 & \text{else} \end{cases}$$

$$E(U,V) = \frac{1}{N} \sum_{(i,j): r_{ij}=1} (y_{ij} - u_i \cdot v_j)^2$$

$N = \sum_{ij} r_{ij}$

Gradient Descent

param

$w \leftarrow w - \eta \, \partial E / \partial w$

Goal:

updating eqs for arbitrary elements of U & V:

u_{lk} $\quad$ v_{jk}

$$E(U,V) = \frac{1}{N} \sum_{(i,j):r_{ij}=1} (y_{ij} - \hat{y}_{ij})^2$$

$$\frac{\partial E}{\partial u_{lk}} = \frac{\partial}{\partial u_{lk}} \frac{1}{N} \sum_{(i,j):r_{ij}=1} (y_{ij} - \hat{y}_{ij})^2$$

$$= \frac{-2}{N} \sum_{(i,j):r_{ij}=1} (y_{ij} - \hat{y}_{ij}) \cdot \left(\frac{\partial \hat{y}_{ij}}{\partial u_{lk}} \right)$$

$$\frac{\partial \hat{y}_{ij}}{\partial u_{lk}} = \begin{cases} 0 & \text{if } l \neq i \\ v_{jk} & \text{if } l = i \end{cases}$$

$$\hat{y}_{ij} = u_i \cdot v_j = \sum_s u_{is} v_{js}$$

$$\frac{\partial \hat{y}_{ij}}{\partial u_{lk}} = \sum_s \frac{\partial}{\partial u_{lk}} u_{is} v_{js}$$

$$= \frac{\partial}{\partial u_{lk}} \left(u_{i1} v_{j1} + u_{i2} v_{j2} + \dots + u_{ik} v_{jk} + \dots + u_{ik} v_{jk} \right)$$

$$= \frac{\partial}{\partial u_{lk}} u_{ik} v_{jk}$$

$$= \begin{cases} v_{jk} & \text{if } l = i \\ 0 & \text{if } l \neq i \end{cases}$$

$$\frac{\partial E}{\partial u_{\ell k}} = \frac{\partial}{\partial u_{\ell k}} \frac{1}{N} \sum_{(i,j):r_{ij}=1} \left(y_{ij} - \hat{y}_{ij}\right)^2$$

$$= -\frac{2}{N} \sum_{(i,j):r_{ij}=1} \left(y_{ij} - \hat{y}_{ij}\right) \cdot \left(\frac{\partial \hat{y}_{ij}}{\partial u_{\ell k}}\right)$$

$$= -\frac{2}{N} \sum_{(j:r_{\ell j}=1)} \left(y_{ij} - \hat{y}_{ij}\right) v_{jk} = \frac{\partial E}{\partial u_{\ell k}}$$

$$u_{\ell k} \longleftarrow u_{\ell k} + \eta \frac{2}{N} \sum_{(j:r_{\ell j}=1)} \left(y_{ij} - \hat{y}_{ij}\right) v_{jk}$$

A similar (& symmetric) argument for item embeddings gives:

$$v_{jk} \longleftarrow v_{jk} + \frac{2\eta}{N} \sum_{(i:r_{ij}=1)} \left(y_{ij} - \hat{y}_{ij}\right) \cdot u_{ik}$$

Vectorized Grad Descent for MF

Define

elementwise multiplication

$$\Delta = (Y - U \cdot V^T) \otimes R$$

What are the elements of Δ?

$$\delta_{ij} = (y_{ij} - u_i \cdot v_j) \cdot r_{ij}$$

Define the matrices

$\frac{\partial E}{\partial U}$ = The matrix w/ elements $\frac{\partial E}{\partial u_{ik}}$

$\frac{\partial E}{\partial V}$ = " " " $\frac{\partial E}{\partial v_{jk}}$

Claim:

$$\frac{\partial E}{\partial U} = \frac{-2}{N} \Delta V$$

$$\frac{\partial E}{\partial V} = \frac{-2}{N} \Delta^T U.$$

What you want to code for your GD

Let's go element-by-element:

$$\left(\frac{\partial E}{\partial U}\right)_{lk} = \frac{\partial E}{\partial u_{lk}} \stackrel{?}{=} \frac{-2}{N} \sum_j \Delta_{ij} V_{jk}$$

1 or 0!

$$= \frac{-2}{N} \sum_j (y_{ij} - u_i v_j) r_{ij} v_{jk}$$

$$= \frac{-2}{N} \sum_{(j : r_{ij} = 1)} (y_{ij} - u_i v_j) v_{jk}$$