Homework 2:

Collaborative Filtering:

In this question, we will use the alternating projections algorithm for low-rank matrix factorization, which aims to minimize:

$$J(u,v) = \underbrace{\frac{1}{2}\sum_{(a,i)\in D}\left(Y_{ai} - [UV^T]_{ai}\right)^2}_{\text{Squared Error}} + \underbrace{\frac{\lambda}{2}\sum_{a=1}^{n}\sum_{j=1}^{k} U_{aj}^2 + \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{j=1}^{k} V_{ij}^2}_{\text{Regularization}}$$

$$Y = \begin{bmatrix} 5 & ? & 7 \\ ? & 2 & ? \\ 4 & ? & ? \\ ? & 3 & 6 \end{bmatrix}$$

D is defined as the set of indices (a,i), where Y_{ai} is not missing.

U and V are initialized as:

$U^{(0)} = [6, 0, 3, 6]^T \quad V^{(0)} = [4, 2, 1]^T$

$$X^{(0)} = UV^T = \begin{bmatrix} 6 \\ 0 \\ 3 \\ 6 \end{bmatrix}\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 3 \\ 6 \end{bmatrix}\begin{bmatrix} 4 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 24 & 12 & 6 \\ 0 & 0 & 0 \\ 12 & 6 & 3 \\ 24 & 12 & 6 \end{bmatrix}$$

Squared error term?

$$J_{square} = \sum_{i,j\in D} (Y_{ij} - X_{ij})^2/2$$

$$= \frac{1}{2}\left((5-24)^2 + (7-6)^2 + (2-0)^2 + (4-12)^2 + (3-12)^2 + (6-6)^2\right) = 255{,}5$$

Regularization term?

$$J_{reg} = \frac{\lambda}{2}\|U\|_F^2 + \frac{\lambda}{2}\|V\|_F^2$$

$$= \frac{\lambda}{2}\sum_{a=1}^{n}(U_a)^2 + \frac{\lambda}{2}\sum_{i=1}^{m}(V_i)^2$$

$$= \frac{\lambda}{2}\left((6^2) + (3)^2 + (6)^2\right) + \frac{\lambda}{2}\left((4)^2 + (2)^2 + (1)^2\right) = 51$$

Suppose V is kept fixed. Run one step algorithm to find the new estimate $U^{(1)}$:

$$U^{(1)} = [U_1^{(1)} \; U_2^{(1)} \; U_3^{(1)} \; U_4^{(1)}]$$

V is Kept Fixed:

$V = [4 \; 2 \; 1]^T$

$$X = UV^T = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} [4 \; 2 \; 1] = \begin{bmatrix} 4u_1 & 2u_1 & u_1 \\ 4u_2 & 2u_2 & u_2 \\ 4u_3 & 2u_3 & u_3 \\ 4u_4 & 2u_4 & u_4 \end{bmatrix}$$

$$U^{(1)} = \arg\min_U J(U)$$

$$= \arg\min_U \sum_{(a,i)\in D} (Y_{ai} - (UV^T)_{ai})^2/2 + \sum_{a=1}^{4} \frac{1}{2}\|U_a\|^2$$

$$= \arg\min_U [(5-4u_1)^2 + (7-u_1)^2 + (2-2u_2)^2 + (4-4u_3)^2 + (3-2u_4)^2 + (6-u_4)^2]/2 + \sum_1^4 u_a^2/2$$

To minimize the loss, we take the gradient with respect to U and equate it to zero:

$$0 = \nabla J(U) = \begin{pmatrix} -4(5-4u_1) - (7-u_1) + u_1 \\ -2(2-2u_2) + u_2 \\ -4(4-4u_3) + u_3 \\ -2(3-2u_4) - (6-u_4) + u_4 \end{pmatrix} = \begin{pmatrix} -27 + 18u_1 \\ -4 + 5u_2 \\ -16 + 17u_3 \\ -12 + 6u_4 \end{pmatrix}$$

$$\Rightarrow U = \begin{pmatrix} \frac{3}{2} \\ \frac{4}{5} \\ \frac{16}{17} \\ 2 \end{pmatrix}$$

Linear Regression: Ridge Regression

$$L(\theta, \theta_0) = \sum_{t=1}^{n} (y^{(t)} - \theta x^{(t)} - \theta_0)^2 + \lambda\theta^2$$

$\frac{\partial L}{\partial \theta}$?

$$\frac{\partial L}{\partial \theta} = \sum_t (-2x^t(y^t - \theta x^t - \theta_0)) + 2\lambda\theta$$

$$\frac{\partial L}{\partial \theta_0} = \sum_t (-2(y^t - \theta x^t - \theta_0))$$

Ridge Regression:

$$L(\theta,\theta_0)=\sum_{t=1}^{m}(y^t-\theta.x^t-\theta_0)+\lambda\theta^2$$

$$\frac{\partial L}{\partial\theta_0}=-2\sum_{t=1}^{m}(y^t-\theta.x^t-\theta_0)=-2\sum_{t=1}^{m}(y^t-\theta x^t)+2\sum_{t=1}^{m}\theta_0=0$$

$$\Longrightarrow -2m\theta_0=-2\sum_{t=1}^{m}(y^t-\theta x^t)$$

$$\theta_0=\frac{1}{m}\sum_{t=1}^{m}(y^t-\theta x^t)$$

$$\frac{\partial L}{\partial\theta}=2\lambda\theta-2\sum_{t=1}^{m}(y^t-\theta.x^t-\theta_0)x^t$$

$$=2\lambda\theta-2\sum_{t=1}^{m}\left(y^t-\theta x^t-\left[\frac{1}{m}\sum_{s=1}^{m}(y^s-\theta x^s)\right]x^t=0\right.$$

$$\Longrightarrow \lambda\theta-\sum_{t=1}^{m}x^ty^t+\theta\sum_{t=1}^{m}x^{(t)2}+\frac{1}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}(y^s-\theta x^s)x^t=0$$

$$\lambda\theta-\sum_{t=1}^{m}x^ty^t+\theta\sum_{t=1}^{m}x^{(t)2}+\frac{1}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}y^sx^t-\frac{1}{m}\theta\sum_{t=1}^{m}\sum_{s=1}^{m}x^sx^t=0$$

$$\hat{\theta}=\frac{\sum_{t=1}^{m}x^ty^t-\frac{1}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}y^sx^t}{\lambda+\sum_{t=1}^{m}x^{(t)2}-\frac{1}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}x^sx^t}$$

$$\bar{x}=\frac{1}{m}\sum_{t=1}^{m}x^{(t)}$$

$$\Longrightarrow \hat{\theta}=\frac{\sum_{t=1}^{m}(x^t-\bar{x})y^t}{\lambda+\sum_{t=1}^{m}x^t(x^t-\bar{x})}$$