Problem Formulation

Example. 영화 평점 예측

Movie	Alice	Bob	Carol	Dave
Love at last	5	5	0	0
Romance forever	5	?	?	0
Cute puppies of love	?	4	0	?
Nonstop car chases	0	0	5	4
Swords vs. karate	0	0	5	?

n_u = no. users $= 4$

n_m = no. movies $= 5$

$r(i,j) = 1$ if user j has rated movie i 0 or 1

$y(i,j)$ = rating given by user j to movie i 0~5

? 가 궁금하지만, 현실에서는 ? 가 더 많음

Recommender System:

automatically fills in ?

Content Based Recommendations

Movie	Alice	Bob	Carol	Dave	r x_1	a x_2
Love at last	5	5	0	0	0.9	0
Romance forever	5	?	?	0	1.0	0.01
Cute puppies of love	?	4	0	?	0.99	0
Nonstop car chases	0	0	5	4	0.1	1.0
Swords vs. karate	0	0	5	?	0	0.9

$n=2$ (x_1, x_2)

no. of features

$$x_0 = 1, \quad x^{(1)} = \begin{bmatrix} 1 \\ 0.9 \\ 0 \end{bmatrix}$$

For each user j, learn a parameter $\theta^{(j)} \in \mathbb{R}^3$.

$\theta^{(j)} \in \mathbb{R}^{n+1}$

Predict user j as rating movie i with $(\theta^{(j)})^T x^{(i)}$ stars.

ex. Alice $\theta^{(1)}$가 'Cute puppies of love'에 부여할 평점?

$$x^{(3)} = \begin{bmatrix} 1 \\ 0.99 \\ 0 \end{bmatrix} \quad \theta^{(1)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$$

$$(\theta^{(1)})^T \cdot x^{(3)} = \underbrace{\begin{bmatrix} 0 & 5 & 0 \end{bmatrix}}_{1\times 3} \underbrace{\begin{bmatrix} 1 \\ 0.99 \\ 0 \end{bmatrix}}_{3\times 1} = 4.95$$

Apply a different copy of linear regression for each user.

$r(i,j) = 1$ if user j has rated movie i 0 or 1

$y(i,j)$ = rating given by user j to movie i 0~5

$\theta^{(j)}$ = parameter vector for user j

$x^{(i)}$ = feature vector for movie i

For user j, movie i, predicted rating: $(\theta^{(j)})^T \cdot x^{(i)}$

$m^{(j)}$ = no. of movies rated by user j

To learn $\theta^{(j)}$ <parameter for user j>

$$\min_{\theta^{(j)}} \frac{1}{2\cancel{m^{(j)}}} \underbrace{\sum_{i;r(i,j)=1}}_{} \left((\theta^{(j)})^T(x^{(i)}) - y^{(i,j)}\right)^2 + \frac{\lambda}{2\cancel{m^{(j)}}} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

sum over all movies user j has rated

$\theta^{(j)} \in \mathbb{R}^{n+1}$

$1/m^{(j)}$ 는 상수니까 제거

To learn $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n_u)}$:

$$\min_{\theta^{(1)},\ldots,\theta^{(n_u)}} \frac{1}{2} \boxed{\sum_{j=1}^{n_u}} \sum_{i:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)}\right)^2 + \frac{\lambda}{2} \boxed{\sum_{j=1}^{n_u}} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

↳ Optimization objective on top $J(\theta^{(1)}, \cdots, \theta^{(n_u)})$

Gradient descent update:

for $k=0$ $\theta_k^{(j)} := \theta_k^{(j)} - \alpha \sum_{i;r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) x_k^{(i)}$

for $k \neq 0$ $\theta_k^{(j)} := \theta_k^{(j)} - \alpha \left(\sum_{i;r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) x_k^{(i)} + \lambda \theta_k^{(i)} \right)$