Recommender system :

Given user's preferences $\theta^{(1)}, \ldots, \theta^{(nu)}$ $\forall\, \theta \in \mathbb{R}^{i \times 1}$

To learn $x^{(nm)}$ for all m movies.

$$\underbrace{\min_{x^{(1)},\ldots,x^{(nm)}} \frac{1}{2} \sum_{i=1}^{n_m} \sum_{j:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2}_{\text{Choose features } x^{(i)} \text{ so that predicted value will be similar to the real rating we observed.}} + \underbrace{\frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} \left(x_k^{(i)} \right)^2}_{\text{Regularization}}$$

Given $\theta \rightarrow$ estimate x | Guess $\theta \rightarrow x \rightarrow \theta \rightarrow x$ (. . .)

Given $x \rightarrow$ estimate θ

Calculate both at same time :

1- Initialize x and θ to small random values.

2- Minimize x and θ using gradient descent for :

$$J(x^{(1)}, \ldots, x^{(nm)}, \theta(1), \ldots, \theta^{(nu)}) =$$

$$= \frac{1}{2} \underbrace{\sum_{(i,j):r(i,j)=1}}_{\text{user-movies pairs rated by the user}} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} (x_k^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

$X \in \mathbb{R}^n, \ \theta \in \mathbb{R}^n$

3rd : For user with θ and movie with learned features x : $\theta^T x$

$$\begin{bmatrix} (x^{(1)})^T(\theta^{(1)}) & \cdots & (x^{(1)})^T(\theta^{(n_u)}) \\ \vdots & \ddots & \vdots \\ (x^{(n_m)})^T(\theta^{(1)}) & \cdots & (x^{(n_m)})^T(\theta^{(n_u)}) \end{bmatrix} = X\Theta^T \quad \text{(low rank matrix)}$$

Find K movies j related to movie i: $\min_K \|x^{(i)} - x^{(j)}\|$

<u>Mean normalization</u>: we may minimize Θ so much that then $X\Theta^T = 0$

And that is not useful at all!

Compute mean of each Y row $\Rightarrow \mu \begin{bmatrix} \text{mean}(\text{row}^{(1)}) \\ \text{mean}(\text{row}^{(2)}) \\ \vdots \\ \text{mean}(\text{row}^{(n)}) \end{bmatrix}$; and then

$y^{(i)} = y^{(i)} - \mu$ for all columns i