Collaborative filtering.

Movie	Alice (θ^{1})	Bob (θ^{2})	Carol θ^{3}	Dave θ^{4}	x_1 (romance)	x_2 (action)
$x^{(1)}$	~~5~~	~~5~~	~~0~~	0	?	?
$x^{(2)}$	5	?	?	0	?	?
⋮	?	4	0	?	?	?
⋮	0	0	5	4	⋮	⋮
	0	0	5	?		⋮

$$\theta^{(1)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} \quad \theta^{(2)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} \quad \theta^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} \quad \theta^{(4)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$$

$$\therefore \; X^{(1)} = \begin{bmatrix} 0 \\ 1.0 \\ 0 \end{bmatrix}.$$

Given $\theta^{(1)} \ldots \theta^{(n_u)}$, to learn $x^{(i)}$:

$$\min_{x^{(i)}} \frac{1}{2} \underbrace{\sum_{j:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2}_{\text{squared error}} + \frac{\lambda}{2} \sum_{k=1}^{n} \left(x_k^{(i)} \right)^2$$

Given $\theta^{(1)}, \ldots \theta^{(n_u)}$, to learn $x^{(1)}, x^{(2)} \ldots x^{(n_m)}$

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

for movie i

sum over all users who have watched that movie.

In the first algo,
if you have movie ratings, & $x^{(1)}, \dots x^{(n_m)}$
then predict $\theta^{(1)}, \dots \theta^{(n_u)}$

In 2nd,

if your users have $\theta^{(1)}, \dots \theta^{(n_u)}$
can estimate $x^{(1)}, \dots x^{(n_m)}$
← features for different movie.

Guess $\theta \rightarrow x \rightarrow \theta \rightarrow x \dots$

So, we are putting both of them together as -

$$J(x^{(1)}, \dots x^{(n_m)}, \theta^{(1)}, \dots \theta^{(n_u)}) = \frac{1}{2} \sum_{(i,j): r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 + \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} (x_k^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

$\theta \in R^n$
$X \in R^n$ now R^{n+1} ... $\{ \cancel{\theta_0 = 1} \; \cancel{x_0 = 1} \}$

Algo

1) Initialize $x^{(1)}, \dots x^{(n_m)}, \theta^{(1)}, \dots \theta^{(n_u)}$ to small random values.
2) Minimize $J(x^{(1)}, \dots x^{(n_m)}, \theta^{(1)}, \dots \theta^{(n_u)})$ using gradient descent (or advanced opt. algorithm) Eg. for every $j = 1 \dots n_u$ $i = 1, \dots n_m$:

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

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

3) For a user with parameters θ & a movie with learned features x, predict a star rating. $\theta^T x$

$(\theta^{(j)})^T (x^{(i)})$

<u>Vectorization Example</u> (Low rank matrix factorization)

Movie	Alice	Bob	Carol	Dave
Love at last	5	5	0	0
Roman...	5	?	?	0
	?	4	0	?
	0	0	5	4
	0	0	5	?

$n_m = 5$
$n_u = 4$

$$Y = \begin{bmatrix} 5 & 5 & 0 & 0 \\ 5 & ? & ? & 0 \\ ? & 4 & 0 & ? \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 5 & ? \end{bmatrix}$$

$y^{(i,j)}$ user movie

$$Y = \begin{bmatrix} 5 & 5 & 0 & 0 \\ 5 & ? & ? & 0 \\ ? & 4 & 0 & ? \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 5 & 0 \end{bmatrix}$$

Predicted ratings

$$\begin{bmatrix} (\theta^{(1)})^T(x^{(1)}) & (\theta^{(2)})^T(x^{(1)}) & \cdots & (\theta^{(n_u)})^T x^{(1)} \\ \vdots & & & \\ (\theta^{(1)})^T(x^{(n_m)}) & (\theta^{(2)})^T(x^{(n_m)}) & \cdots & ((\theta^{(n_u)})^T) x^{(n_m)} \end{bmatrix}$$

$$\underset{\text{user } j}{(\theta^{(j)})^T} \underset{\text{movie } i}{(x^{(i)})}$$

$$X = \begin{bmatrix} - x^{(1)T} - \\ - (x^{(2)})^T - \\ \vdots \\ -(x^{(n_m)})^T - \end{bmatrix} \qquad \theta = \begin{bmatrix} -(\theta^{(1)})^T - \\ - (\theta^{(2)})^T - \\ \vdots \\ -(\theta^{(n_u)})^T - \end{bmatrix}$$

$\therefore X\theta^T$

<u>Low rank Matrix vectorization</u>

How to find movie j is related to movie i

$\| x^{(i)} - x^{(j)} \| \rightarrow$ is small

$\therefore$ movie j is similar to i.

Mean Normalization

New user

$$Y = \begin{bmatrix} 5 & 5 & 0 & 0 & ? \\ 5 & ? & ? & 0 & ? \\ ? & 4 & 0 & ? & ? \\ 0 & 0 & 5 & 4 & ? \\ 0 & 0 & 5 & 0 & ? \end{bmatrix} \qquad \mu = \begin{bmatrix} 2.5 \\ 2.5 \\ 2 \\ 2.25 \\ 1.25 \end{bmatrix}$$

New user (Eve)

$Y = Y - \mu$

$$\therefore Y = \begin{bmatrix} 2.5 & 2.5 & -2.5 & -2.5 & ? \\ 2.5 & ? & ? & -2.5 & ? \\ ? & 2 & -2 & ? & ? \\ -2.25 & -2.25 & 2.75 & 1.75 & ? \\ -1.25 & -1.25 & 3.75 & -1.25 & ? \end{bmatrix}$$

learn $\theta^{(j)}$ & $x^{(i)}$

$\therefore$ For user j, on movie i predict:

$$\rightarrow (\theta^{(j)})^T (x^{(i)}) + \mu_i$$

Eve -

$$\theta^{(5)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \therefore (\theta^{(5)})^T (x^{(i)}) + \mu_i$$