Content Based Recommendation:

Sid

- n_u = # users
- n_m = # movies
- $r(i,j) = 1$ if user j has rated movie i, 0 otherwise
- $y(i,j)$ = rating given by user j to movie i

- $\theta^{(j)}$ = parameter vector for user j
- $x^{(i)}$ = feature vector for movie i

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

$m^{(j)}$ = # movies rated by user j

$$\min_{\theta^{(1)} \to \theta^{(n_u)}} J(\theta^{(1)} \to \theta^{(n_u)})$$

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

Gradient Descent

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

(for $k \neq 0$)

Collaborative Filtering : Feature Learning

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

$$J(x, \theta) =$$

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

1) Initialize $x^{(i)} \to x^{(n_m)}$, $\theta^{(1)} \to \theta^{(n_u)}$ to small random values

(breaks symmetry to have algorith learn distinct features

2) minimize $J(x, \theta)$

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

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

3) For user w/ parameters θ & features x:

prediction : $\theta^T x$

Vectorization: Low Rank Matrix Factorization

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

$$Y = \underbrace{X \Theta^T} \quad \begin{bmatrix} (\theta^{(1)})^T (x^{(1)}) & \cdots & (\theta^{(n_u)})^T (x^{(1)}) \\ \vdots & & \\ (\theta^{(1)})^T (x^{(n_m)}) & \cdots & (\theta^{(n_u)})^T (x^{(n_m)}) \end{bmatrix}$$

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

Mean Normalization:

$$\mu = [\mu_1, \cdots \mu_{n_m}] \quad , \quad \mu_i = \frac{\sum_{j:r(i,j)=1} Y_{i,j}}{\sum_j r(i,j)}$$

$$Y' = Y - \mu$$

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