Predicting Movie Ratings:

May 31, 2021

Users rate 0-5 stars.

Movie	Alice	Bob	Carol	Dave
Love at Last	5	5	0	0
Romance Forever	5	?	?	0
Cute Puppies	?	4	0	?
Car Chases	0	0	5	4
Swords vs Karate	0	0	5	?

Notation:

n_u = # users

n_m = # movies

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

$y^{(i,j)}$ = rating (0-5) given by user j to movie i

↳ defined only if $r(i,j) = 1$

Goal: Predict ratings for '?' to recommend new movies.

Content-Based Recommender Systems:

$$\begin{bmatrix} 1 \\ 0.9 \\ 0 \end{bmatrix}$$

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)	x_1 (romance)	x_2 (action)
$x^{(1)}$ Love at last	5	5	0	0	0.9	0
$x^{(2)}$ Romance forever	5	?	?	0	1.0	0.01
$x^{(3)}$ Cute puppies of love	?	4	0	?	0.99	0
$x^{(4)}$ Nonstop car chases	0	0	5	4	0.1	1.0
$x^{(5)}$ Swords vs. karate	0	0	5	?	0	0.9

$n_u = 4$

$n_m = 5$

$x_0 = 1 \rightarrow$ default

$n = 2$ (# features)

$$x_1 = \begin{bmatrix} 1 \\ 0.9 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1.0 \\ 0.01 \end{bmatrix}, \text{ etc}$$

For each user j, learn a parameter $\theta^{(j)} \in \mathbb{R}^{n+1}$.
Predict user j as rating movie i w/ $(\theta^{(j)})^T x^{(i)}$ stars.

e.g.

$$x^{(3)} = \begin{bmatrix} 1 \\ 0.99 \\ 0 \end{bmatrix}, \quad \theta^{(1)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}, \quad (\theta^{(1)})^T x^{(3)} = (5)(0.99)$$
$$= 4.95$$

Problem Formulation:

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

$y^{(i,j)}$ = rating by user j on movie i if defined

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

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

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

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

To Learn $\theta^{(j)}$:

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

To Learn $\theta^{(1)}, \theta^{(2)} \ldots \theta^{(m)}$:

$J(\theta^{(1)} \ldots \theta^{(m)})$

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

Gradient Descent Update:

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

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \left(\underbrace{\sum_{i:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right) x_k^{(i)}}_{\frac{\partial}{\partial \theta_k^{(j)}} J(\theta^{(1)} \ldots \theta^{(n_u)})} + \lambda \theta_k^{(j)} \right) \quad \text{for } k \neq 0$$