

Coursera 16-5 Vectorization Low Rank Matrix Factorization

<https://www.coursera.org/learn/machine-learning/lecture/CEXN0/vectorization-low-rank-matrix-factorization>

Collaborative filtering

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)
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	?

$n_m = 5$
 $n_u = 4$

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

$y^{(i,j)}$

-앞에서 본 예시를 행렬로 만들어준다.

Collaborative filtering

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

Predicted ratings:

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

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

$L = \begin{bmatrix} -(\theta^{(1)})^T \\ -(\theta^{(2)})^T \\ \vdots \\ -(\theta^{(n_u)})^T \end{bmatrix}$

→ Low rank matrix factorization

Andreas 1

- $X^{(1)} \dots X^{(n_m)}$ 영화의 특성들을 행으로 쌓아준다. 세타(사용자의 성향)들도 행으로 쌓아준다.
- Low rank matrix라고 부르지만 선형대수의 Low rank matrix과는 다른 개념이다.

Finding related movies

For each product i , we learn a feature vector $\underline{x^{(i)}} \in \mathbb{R}^n$.

→ $x_1 = \text{romance}$, $x_2 = \text{action}$, $x_3 = \text{comedy}$, $x_4 = \dots$

How to find movies j related to movie i ?

small $\|\underline{x^{(i)}} - \underline{x^{(j)}}\| \rightarrow$ movie j and i are "similar"

5 most similar movies to movie i :

→ Find the 5 movies j with the smallest $\|\underline{x^{(i)}} - \underline{x^{(j)}}\|$.

- 협업 필터링 알고리즘을 사용하기 위해선 영화들간의 관계를 파악해야 된다.
- 영화들간의 관계를 파악하기 위해 영화들 간의 거리를 계산한다.
- 거리가 가까울수록 연관성이 있는 관계이다.

단어공부

Property : 성질

Salient : 가장 두드러지는, 절호의