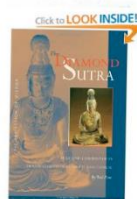


سیستم‌های توصیه‌گر

سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

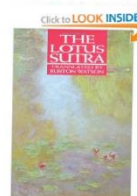
Customers Who Viewed This Item Also Bought



The Diamond Sutra
▶ Red Pine
★★★★★ (20)
Paperback
\$13.57



The Heart Sutra
▶ Red Pine
★★★★★ (21)
Paperback
\$10.17



The Lotus Sutra
▶ Burton Watson
★★★★★ (27)
Paperback
\$18.21

معرفی □

رویکردها □

پالایش گروهی □

رویکرد مبتنی بر محتوی □

رویکرد پیوندی □

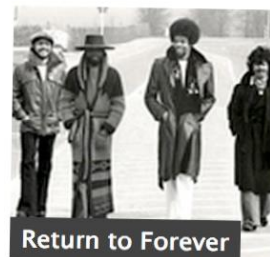
Similar Artists



Stanley Clarke & George Duke



Victor Wooten



Return to Forever



S.M.V.

پالایش گروہی

پالایش گروهی

- برجسته‌ترین رویکرد استفاده شده در سیستم‌های توصیه‌گر.
 - استفاده شده به وسیله سایت‌های تجاری بسیار بزرگ
 - شامل انواع مختلفی از الگوریتم‌ها
 - قابل استفاده در بسیاری از دامنه‌ها (کتاب، فیلم، موسیقی و ...)
- رویکرد.
 - استفاده از «**خرد جمعی**» برای توصیه کالاها
- ایده.
 - کاربران به کالاهای خریداری شده امتیاز می‌دهند. [معمولاً بین ۱ و ۵]
 - کاربرانی که در گذشته سلیقه مشابهی داشته‌اند، احتمالاً در آینده نیز دارای سلیقه‌های مشابهی خواهند بود.

پالایش گروهی

□ ورودی. یک ماتریس شامل امتیازهای داده شده به وسیله کاربران به کالاهای گوناگون.

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

□ انواع خروجی.

□ یک پیش‌بینی (عددی) در مورد میزان علاقه کاربر به یک کالای خاص

□ یک فهرست پیشنهادی از N کالای برتر

پالایش گروهی مبتنی بر کاربر

□ روش پایه.

□ با داشتن یک کاربر فعال مانند آلیس و کالای i که قبلاً به وسیله آلیس دیده نشده است:

- یک مجموعه از کاربران (نزدیک‌ترین همسایه‌ها) پیدا کن که سلیقه مشابهی با آلیس داشته‌اند و قبلاً به کالای i امتیاز داده‌اند.
- میانگین امتیاز داده شده به وسیله این کاربران به کالای i را محاسبه کن.
- از میانگین محاسبه شده به عنوان تخمینی از میزان علاقمندی آلیس به کالای i استفاده کن
- این کار را برای تمام کالاهایی که آلیس به آنها امتیاز نداده است، تکرار کن.
- کالاهای با امتیاز بیشتر را به آلیس پیشنهاد کن.

□ ایده. کاربرانی که در گذشته سلیقه مشابهی داشته‌اند، احتمالاً در آینده نیز دارای سلیقه‌های مشابهی خواهند بود.

پالایش گروهی مبتنی بر کاربر

۷

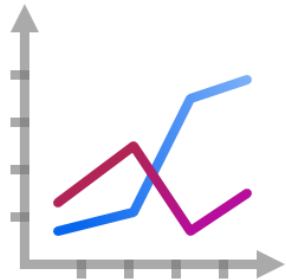
□ مثال. ورودی:

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

□ هدف. پیش‌بینی میزان علاقمندی آلیس به کالای شماره ۵

پالایش گروهی مبتنی بر کاربر

۸



□ چند پرسش ابتدایی.

□ چگونه می توان شباهت میان کاربران مختلف را محاسبه نمود؟

□ چه تعداد از همسایه ها را باید در نظر گرفت؟

□ چگونه می توان با توجه به امتیاز همسایه ها، یک پیش بینی ارائه کرد؟

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

معیارهای اندازه‌گیری شباهت کاربران

□ ضریب همبستگی پیرسون. یک معیار پر کاربرد

□ a و b : کاربرها

□ $r_{a,p}$: امتیاز داده شده به وسیله کاربر a به کالای p

□ P : یک مجموعه از کالاها که به وسیله هر دو کاربر a و b امتیازدهی شده‌اند.

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

$$\frac{cov(a, b)}{std(a) \cdot std(b)}$$

$$\frac{1}{2} + \frac{1}{2} sim(a, b) \quad \text{نرمال‌سازی}$$

معیارهای اندازه گیری شباهت کاربران

۱۰


□ ضریب همبستگی پیرسون. یک معیار پر کاربرد

□ a و b : کاربرها

□ $r_{a,p}$: امتیاز داده شده به وسیله کاربر a به کالای p

□ P : یک مجموعه از کالاها که به وسیله هر دو کاربر a و b امتیازدهی شده‌اند.

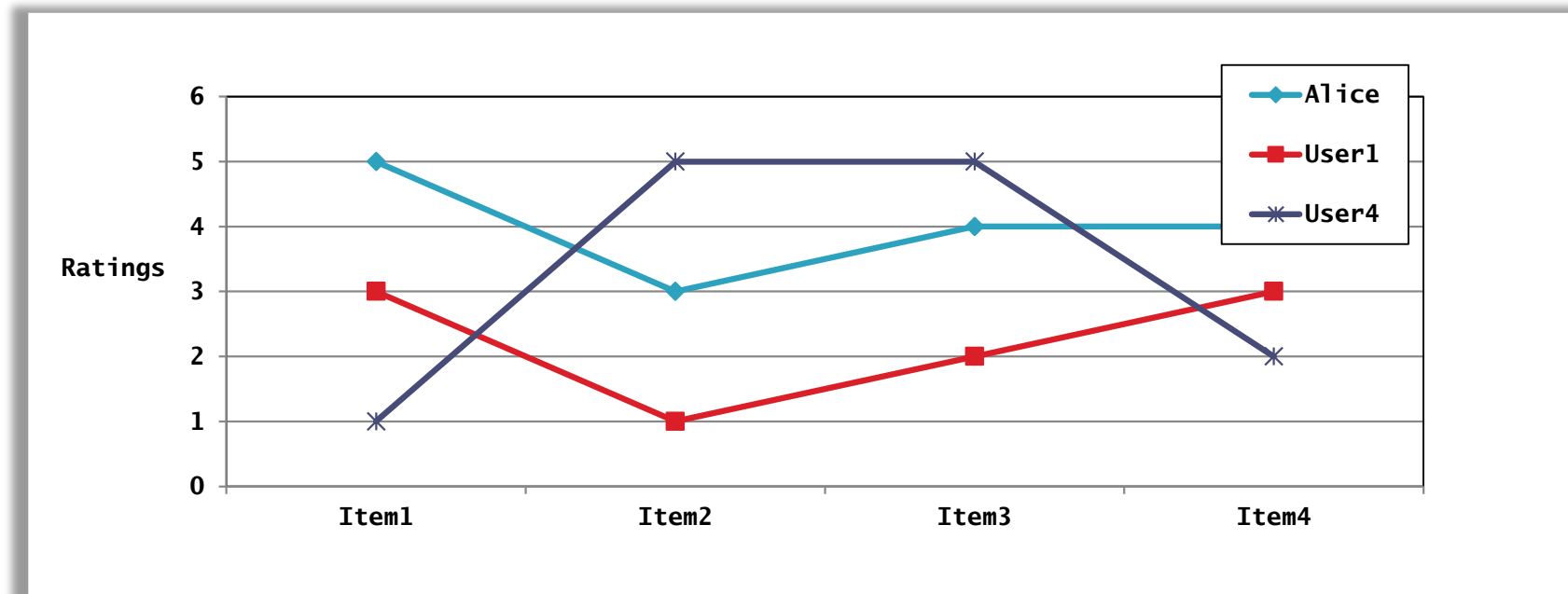
	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1



$sim = 0.85$
 $sim = 0.70$
 $sim = 0.00$
 $sim = -0.79$

ضریب همبستگی پیرسون

- مزیت. در نظر گرفتن تفاوتها در عادات امتیازدهی
- ضریب همبستگی پیرسون در بسیاری از دامنه‌ها نسبت به معیارهای دیگر عملکرد بهتری دارد.
[معیار اقلیدسی، معیار کسینوسی]



□ یک تابع متداول برای پیش‌بینی:

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

- محاسبه این که امتیاز همسایه‌ها برای کالای p کمتر یا بیشتر از میانگین آنها است.
- ترکیب اختلاف‌ها -- استفاده از معیار شباهت برای وزن‌دهی.
- اضافه کردن میانگین امتیازهای کاربر a به مقدار محاسبه شده.

رویکردهای مبتنی بر حافظه و مبتنی بر مدل

- پالایش گروهی مبتنی بر کاربر، یک روش «مبتنی بر حافظه» است.
- استفاده مستقیم از ماتریس امتیازها برای یافتن نزدیک‌ترین همسایه‌ها و پیش‌بینی
- در بسیاری از کاربردهای دنیای واقعی این رویکرد قابل استفاده نیست!
- به دلیل وجود ده‌ها میلیون کاربر و میلیون‌ها کالا

- رویکردهای مبتنی بر مدل.
- بر مبنای یادگیری مدل به صورت آفلاین (آموزش آفلاین)
- در زمان اجرا تنها از مدل یاد گرفته شده برای پیش‌بینی استفاده می‌شود.
- مدل‌ها به طور تناوبی به روز رسانی می‌شوند.
- ایجاد مدل و به روز رسانی آن می‌تواند از نظر محاسباتی بسیار پرهزینه باشد.

پالایش گروهی مبتنی بر کالا

۱۴

□ ایده.

□ استفاده از شباهت میان کالاها برای پیش‌بینی [نه شباهت میان کاربران]

□ مثال. جستجو به دنبال کالاهای مشابه با کالای ۵

□ استفاده از امتیازهای داده شده به وسیله آلیس به کالاهای مشابه برای پیش‌بینی امتیاز کالای ۵

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

معیار شباهت میان کالاها

□ معیار شباهت کسینوسی.

□ تولید نتایج بهتر در مقایسه کالا به کالا

□ امتیازهای داده شده به هر کالا به عنوان یک بردار در فضای n بعدی در نظر گرفته می شوند.

□ شباهت میان دو کالا با محاسبه کسینوس زاویه مربوط به بردار این دو کالا اندازه گیری می شود:

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

□ معیار شباهت کسینوسی تنظیم شده.

□ در نظر گرفتن میانگین امتیازهای هر کاربر

□ U : مجموعه کاربرانی که به هر دو کالای a و b امتیاز داده اند.

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

□ یک تابع متداول برای پیش‌بینی:

$$pred(u, p) = \frac{\sum_{i \in ratedItem(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItem(u)} sim(i, p)}$$

□ معمولاً اندازه همسایگی محدود است.

□ یعنی، از همه همسایه‌ها برای پیش‌بینی استفاده نمی‌شود.

□ یک قاعده تجربی: در بسیاری از کاربردهای دنیای واقعی، تعداد همسایه‌ها بین ۲۰ الی ۵۰ در نظر گرفته می‌شود. [هرلاکر، ۲۰۰۲]

مسئله خلوت بودن داده‌ها

□ مسئله شروع سرد.

- چگونه می‌توان کالاهای جدید را توصیه کرد؟
- چگونه می‌توان به کاربران جدید توصیه داد؟

□ راه‌حل‌های ساده.

- از کاربر بخواه مجموعه‌ای از کالاها را امتیازدهی کند.
- در مراحل ابتدایی از روش‌های دیگر مانند **پالایش مبتنی بر محتوی** استفاده کنید.
- مقادیر پیش‌فرض: استفاده از مقادیر پیش‌فرض برای کالاهایی که فقط یکی از دو کاربری که قرار است مقایسه شوند به آنها امتیاز داده‌اند.

انواع رویکردهای مبتنی بر مدل

- تجزیه ماتریس‌ها.
- تجزیه مقادیر منفرد، تحلیل مولفه‌های اصلی
- کاوش قواعد ارتباطی.
- مقایسه: تحلیل سبد خرید
- مدل‌های احتمالاتی.
- خوشه‌بندی، شبکه‌های بیزی و ...
- هزینه پیش‌پردازش (یادگیری مدل).
- معمولاً درباره آن صحبت نمی‌شود
- آیا به روز رسانی تدریجی ممکن است؟

تجزیه مقادیر منفرد

□ انگیزه.

- ساده‌سازی داده‌ها
- حذف نویز و افزودن
- بهبود نتایج الگوریتم

□ کاربردهای مثالی.

- جستجو و بازیابی اطلاعات [شاخص‌گذاری معنایی نهان]
- سیستم‌های توصیه‌گر

تجزیه مقادیر منفرد

□ تجزیه مقادیر منفرد.

$$Data_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

ماتریس مقادیر منفرد

□ ماتریس مقادیر منفرد.

- یک ماتریس قطری که در آن مقادیر منفرد به صورت کاهشی مرتب هستند.
- مقادیر منفرد از یک اندیس مانند r به بعد دارای مقدار صفر هستند.
- مقادیر منفرد ریشه دوم مقادیر ویژه ماتریس $Data \times Data^T$ هستند.

توصیه‌های مبتنی بر محتوی

توصیه مبتنی بر محتوی

- آموزش. برای کاربر j بردار $\theta^{(j)} \in \mathbb{R}^3$ را یاد بگیر
- پیش‌بینی. امتیاز فیلم i برای کاربر j

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

	Alice(1)	Bob(2)	Carol(3)	Dave(4)	x_1	x_2
Titanic	5	5	0	0	0.90	0.00
Sound and Music	5	?	?	0	1.00	0.01
Casablanca	?	4	0	?	0.99	0.00
Fast and Furious	0	0	5	4	0.10	1.00
Desperado	0	0	5	?	0.00	0.90

بیان رسمی مسئله

- $r(i,j) = 1$ اگر کاربر j به فیلم i امتیاز داده باشد، در غیر این صورت صفر.
- $y^{(i,j)}$ امتیاز داده شده توسط کاربر j به فیلم i
- $\theta^{(j)}$ بردار پارامترها برای کاربر j
- $x^{(i)}$ بردار ویژگی برای فیلم i
- پیش‌بینی امتیاز فیلم i برای کاربر j :

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

- $m^{(j)}$ تعداد فیلم‌های امتیازدهی شده به وسیله کاربر j

هدف بهینه‌سازی

□ یادگیری بردار $\theta^{(j)}$ -- پارامترها برای کاربر j

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

□ یادگیری بردارهای $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$

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

الگوریتم بهینه‌سازی

□ تابع هدف.

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

□ گرادیان کاهش.

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

پالایش گروہی

پالایش گروهی

	Alice(1)	Bob(2)	Carol(3)	Dave(4)	x_1	x_2
Titanic	5	5	0	0	0.90	0.00
Sound and Music	5	?	?	0	1.00	0.01
Casablanca	?	4	0	?	0.99	0.00
Fast and Furious	0	0	5	4	0.10	1.00
Desperado	0	0	5	?	0.00	0.90

$$\theta^{(1)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$$

$$\theta^{(2)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$$

$$\theta^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$$

$$\theta^{(4)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$$

$$(\theta^{(1)})^T x^{(1)} \approx 5$$

$$(\theta^{(2)})^T x^{(1)} \approx 5$$

$$(\theta^{(3)})^T x^{(1)} \approx 0$$

$$(\theta^{(4)})^T x^{(1)} \approx 0$$

هدف بهینه‌سازی

□ یادگیری $x^{(i)}$ -- با داشتن پارامترهای $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$

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

□ یادگیری $x^{(1)}, x^{(2)}, \dots, x^{(n_m)}$ -- با داشتن پارامترهای $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$

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

پالایش گروهی

۲۹

□ ایده. با داشتن ماتریس امتیازها و بردارهای $x^{(1)}, x^{(2)}, \dots, x^{(n_m)}$ می توان بردارهای $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$ را تخمین زد

□ ایده. با داشتن ماتریس امتیازها و بردارهای $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$ می توان بردارهای $x^{(1)}, x^{(2)}, \dots, x^{(n_m)}$ را تخمین زد

□ الگوریتم.

مقداردهی تصادفی

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

الگوریتم پالایش گروهی

پالایش گروهی

□ ایده. تخمین $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$ با داشتن بردارهای $x^{(1)}, x^{(2)}, \dots, x^{(n_m)}$

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

□ ایده. تخمین $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_u)}$ با داشتن بردارهای $x^{(1)}, x^{(2)}, \dots, x^{(n_m)}$

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

هدف بهینه‌سازی در پالایش گروهی

□ ایده. یادگیری همزمان بردارهای ویژگی $x^{(i)}$ و بردارهای $\theta^{(j)}$

□ تابع هدف.

$$J(x^{(1)}, \dots, x^{(n)}, \theta^{(1)}, \dots, \theta^{(n)}) =$$

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

□ هدف.

$$\min_{\substack{x^{(1)}, \dots, x^{(n_m)} \\ \theta^{(1)}, \dots, \theta^{(n_u)}}} J(x^{(1)}, \dots, x^{(n)}, \theta^{(1)}, \dots, \theta^{(n)})$$

الگوریتم پالایش گروهی

□ آموزش.

- مقداردهی اولیه به بردارهای x و θ با مقادیر تصادفی کوچک
- کمینه‌سازی تابع هزینه با استفاده از گرادیان کاهشی (یا روش‌های بهینه‌سازی پیشرفته)

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

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

□ پیش‌بینی.

- برای کاربر j با بردار پارامتر $\theta^{(j)}$ و فیلم i با بردار ویژگی $x^{(i)}$

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

نرمال سازی میانگین

	Alice(1)	Bob(2)	Carol(3)	Dave(4)	eve(5)	
Titanic	5	5	0	0	?	0
Sound and Music	5	?	?	0	?	0
Casablanca	?	4	0	?	?	0
Fast and Furious	0	0	5	4	?	0
Desperado	0	0	5	?	?	0

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

$$\theta^{(5)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow (\theta^{(5)})^T x^{(i)} = 0$$

نرمال سازی میانگین

□ نرمال سازی میانگین.

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

$$\mu = \begin{bmatrix} 2.5 \\ 2.5 \\ 2 \\ 2.25 \\ 1.25 \end{bmatrix} \rightarrow$$

$$Y_{norm} = \begin{bmatrix} 2.5 & 2.5 & -2.5 & -2.5 & ? \\ 2.5 & ? & ? & -2.5 & ? \\ ? & 2.0 & -2.0 & ? & ? \\ -2.25 & -2.25 & 2.75 & 1.75 & ? \\ -1.25 & -1.25 & 3.75 & -1.25 & ? \end{bmatrix}$$

□ پیش بینی. میزان علاقمندی کاربر j به فیلم i

$$\hat{y}(i, j) = (\theta^{(j)})^T x^{(i)} + \mu^{(i)}$$