03-02 Systèmes de recommandation II

NOUS ÉCLAIRONS.
VOUS BRILLEZ.

FORMATION CONTINUE
ET SERVICES AUX ENTREPRISES

Cégep de Sainte-Foy

Sommaire

Sommaire

Recommandation basée sur le contenu

- On suppose que chaque série est représentée par un **vecteur "profile"** x

	Alice (1)	Bob (2)	Mike (3)	Alex (4)	x_1 (crime)	x_2 (horreur)
Money Heist $x^{(1)}$	5	5	1	?		
Mindhunter $x^{(2)}$	4	5	?	0		
The Walking Dead $x^{(3)}$	1	?	5	4		
The Haunting of Hill House $x^{(4)}$	?	2	4	5		
Ash vs Evil Dead $x^{(5)}$	?	0	5	?		

Recommandation basée sur le contenu

- On suppose que chaque série est représentée par un **vecteur "profile"** x

	Alice (1)	Bob (2)	Mike (3)	Alex (4)	x_1 (crime)	x_2 (horreur)
Money Heist $x^{(1)}$	5	5	1	?	0.90	0
Mindhunter $x^{(2)}$	4	5	?	0	0.95	0.2
The Walking Dead $x^{(3)}$	1	?	5	4	0	0.80
The Haunting of Hill House $x^{(4)}$	?	2	4	5	0.05	0.99
Ash vs Evil Dead $x^{(5)}$	?	0	5	?	0	0.85

$$x^{(3)} = \begin{bmatrix} 1 \\ 0 \\ 0.80 \end{bmatrix} \quad \rightarrow x_0 = 1$$

$n = 2$

Recommandation basée sur le contenu

- Prédiction de la note d'une série j par l'utilisateur i

	Bob (2)	x_1 (crime)	x_2 (horreur)
Money Heist $\boldsymbol{x}^{(1)}$	5	0.90	0
Mindhunter $\boldsymbol{x}^{(2)}$	5	0.95	0.2
The Walking Dead $\boldsymbol{x}^{(3)}$	?	0	0.80
The Haunting of Hill House $\boldsymbol{x}^{(4)}$	2	0.05	0.99
Ash vs Evil Dead $\boldsymbol{x}^{(5)}$	0	0	0.85

Pour prédire la note que donnerait Bob à "The Walking Dead", **il faut apprendre un vecteur "profile utilisateur"**

Recommandation basée sur le contenu

- Prédiction de la note d'une série j par l'utilisateur i

	Bob (2) $\theta^{(2)}$	x_1 (crime)	x_2 (horreur)
Money Heist $\boldsymbol{x}^{(1)}$	5	0.90	0
Mindhunter $\boldsymbol{x}^{(2)}$	5	0.95	0.2
The Walking Dead $\boldsymbol{x}^{(3)}$	?	0	0.80
The Haunting of Hill House $\boldsymbol{x}^{(4)}$	2	0.05	0.99
Ash vs Evil Dead $\boldsymbol{x}^{(5)}$	0	0	0.85

$$\theta^{(2)} = \begin{bmatrix} 0 \\ 5 \\ 0.1 \end{bmatrix} \qquad \boldsymbol{x}^{(3)} = \begin{bmatrix} 1 \\ 0 \\ 0.80 \end{bmatrix}$$

$$(\theta^{(2)})^T \boldsymbol{x}^{(3)} = 0 \times 1 + 5 \times 0 + 0.1 \times 0.8$$
$$= 0.08$$

- Pour chaque utilisateur j, on apprend un vecteur de paramètres $\theta^{(j)} \in \mathbb{R}^{n+1}$

Recommandation basée sur le contenu

- Prédiction de la note d'une série j par l'utilisateur i

	Bob (2) $\theta^{(2)}$	x_1 (crime)	x_2 (horreur)
Money Heist $x^{(1)}$	5	[illegible]	[illegible]
Mindhunter $x^{(2)}$	5	0.95	[illegible]
The Walking Dead $x^{(3)}$	?	0	0.80
The Haunting of Hill House $x^{(4)}$	2	0.05	0.99
Ash vs Evil Dead $x^{(5)}$	0	0	0.85

$$\theta^{(2)} = \begin{bmatrix} 0 \\ 5 \\ [illegible] \end{bmatrix} \qquad \boldsymbol{x}^{(3)} = \begin{bmatrix} 1 \\ 0 \\ 0.80 \end{bmatrix}$$

$$(\theta^{(2)})^T \boldsymbol{x}^{(3)} = 0 \times 1 + 5 \times 0 + 0.1 \times 0.8$$
$$= 0.08$$

On prédit que l'utilisateur j donnerait $(\theta^{(j)})^{\mathsf{T}} x^{(i)}$ étoiles à la série i

- Pour chaque utilisateur j, on apprend un vecteur de paramètres $\theta^{(j)} \in \mathbb{R}^{n+1}$

Exercice

- Considérons les notes suivantes. Quelle pourrait être une valeur de $\theta^{(3)}$?

	Alice (1)	Bob (2)	Mike (3)	Alex (4)	x_1 (crime)	x_2 (horreur)
Money Heist $x^{(1)}$	5	5	1	?	1	0
Mindhunter $x^{(2)}$	4	5	?	0	1	0
The Walking Dead $x^{(3)}$	1	?	5	4	0	1
The Haunting of Hill House $x^{(4)}$	?	2	5	5	0	1
Ash vs Evil Dead $x^{(5)}$	?	0	5	?	0	1

- Réponse: par exemple $\theta^{(3)} = \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}$

Recommandation basée sur le contenu

- $r^{(i,j)}$ = 1 si l'utilisateur j a noté la série i, $r^{(i,j)}$ = 0 sinon
- $y^{(i,j)}$ = note donnée par l'utilisateur j sur la série i (si et seulement si $r^{(i,j)}$ = 1)
- $\boldsymbol{\theta}^{(j)}$ = vecteur de paramètres / profile de l'utilisateur j ($\boldsymbol{\theta}^{(j)} \in \mathbb{R}^{n+1}$)
- $\boldsymbol{x}^{(j)}$ = vecteur "profile" de la série i
- Prédiction $(\boldsymbol{\theta}^{(j)})^{\mathsf{T}}\boldsymbol{x}^{(i)}$ pour la série i de l'utilisateur j
- $s^{(j)}$ représente le nombre de séries notées par l'utilisateur j
- $\rightarrow$ **Comment apprendre $\boldsymbol{\theta}^{(j)}$?**

Recommandation basée sur le contenu

- $r^{(i,j)}$ = 1 si l'utilisateur j a noté la série i, $r^{(i,j)}$ = 0 sinon
- $y^{(i,j)}$ = note donnée par l'utilisateur j sur la série i (si et seulement si $r^{(i,j)}$ = 1)
- $\theta^{(j)}$ = vecteur de paramètres de l'utilisateur j ($\theta^{(j)} \in \mathbb{R}^{n+1}$)
- $x^{(j)}$ = vecteur de fe
- Prédiction $(\theta^{(j)})^{\mathsf{T}} x^{(i)}$ pour la série i de l'utilisateur j
- $s^{(j)}$ représente le nombre de séries notées par l'utilisateur j
- **→ Comment apprendre $\theta^{(j)}$?**

Par régression linéaire simple !

Recommandation basée sur le contenu

- Pour apprendre $\theta^{(j)}$:

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

$$\min_{\theta^{(j)}} J(\theta^{(j)})$$

Recommandation basée sur le contenu

- Pour apprendre $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(nu)}$:

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

$$\min_{\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n_u)}} J(\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n_u)})$$

Recommandation basée sur le contenu

- Comme en régression linéaire simple, nous utilisons la **descente de gradient**

$$\text{Répéter jusqu'à convergence}\{$$

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \sum_{i:r^{(i,j)}=1}^{s^{(j)}} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \times x_k^{(i)} \qquad (k = 0)$$

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \left[\sum_{i:r^{(i,j)}=1}^{s^{(j)}} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \times x_k^{(i)} + \lambda \theta_k^{(j)} \right] \qquad (k \neq 0)$$

$$\}$$

Recommandation basée sur le contenu

- **Avantages**
  - Pas besoin des données des autres utilisateurs
  - Capable de recommander les utilisateurs aux goût uniques
  - Capable de recommander des nouveaux produits ou des produits impopulaires
  - **Modèle interprétable**, car nous avons les vecteurs “profile”

- **Inconvénients**
  - Nécessite le vecteur “profile” qui peut être très difficile d’obtenir (ex. Musique, films, …)
  - N’exploite pas les notation des autres utilisateurs
  - Problème du démarrage à froid (sera adressé par la normalisation par la moyenne)
  - Ne fonctionne que pour un seul type de produit

Sommaire

Filtrage collaboratif

- La recommandation basée sur le contenu suppose un **vecteur “profile”** x

	Alice (1)	Bob (2)	Mike (3)	Alex (4)	x_1 (crime)	x_2 (horreur)
Money Heist	5	5	1	?	0.90	0
Mindhunter	4	5	?	0	0.95	0.2
The Walking Dead	1	?	5	4	0	0.80
The Haunting of Hill House	?	2	4	5	0.05	0.99
Ash vs Evil Dead	?	0	5	?	0	0.85

Filtrage collaboratif

?

- **La réalité est toute autre !**

	Alice (1)	Bob (2)	Mike (3)	Alex (4)	x_1 (crime)	x_2 (horreur)
Money Heist	5	5	1	?	?	?
Mindhunter	4	5	?	0	?	?
The Walking Dead	1	?	5	4	?	?
The Haunting of Hill House	?	2	4	5	?	?
Ash vs Evil Dead	?	0	5	?	?	?

Filtrage collaboratif

- **La réalité est toute autre !**

?

	Alice $\theta^{(1)}$	Bob $\theta^{(2)}$	Mike $\theta^{(3)}$	Alex $\theta^{(4)}$	x_1	x_2
$x^{(1)}$	5	5	1	?	?	?
$x^{(2)}$	4	5	?	0	?	?
$x^{(3)}$	1	?	5	4	?	?
$x^{(4)}$	?	2	4	5	?	?
$x^{(5)}$	?	0	5	?	?	?

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

Filtrage collaboratif

$\boldsymbol{x}^{(3)}$ tel que $(\theta^{(1)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 1$, $(\theta^{(3)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 5$ et $(\theta^{(4)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 4$

- **La réalité est toute autre !**

	Alice $\theta^{(1)}$	Bob $\theta^{(2)}$	Mike $\theta^{(3)}$	Alex $\theta^{(4)}$	x_1	x_2
$x^{(1)}$	5	5	1	?	?	?
$x^{(2)}$	4	5	?	0	?	?
$x^{(3)}$	1	?	5	4	?	?
$x^{(4)}$	?	2	4	5	?	?
$x^{(5)}$	?	0	5	?	?	?

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

Filtrage collaboratif

$\boldsymbol{x}^{(3)}$ tel que $(\theta^{(1)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 1$, $(\theta^{(3)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 5$ et $(\theta^{(4)})^\mathsf{T}\boldsymbol{x}^{(3)} \approx 4$

- **La réalité est toute autre !**

	Alice $\theta^{(1)}$	Bob $\theta^{(2)}$	Mike $\theta^{(3)}$	Alex $\theta^{(4)}$	x_1	x_2
$x^{(1)}$	5	5	1	?	?	?
$x^{(2)}$	4	5	?	0	?	?
$x^{(3)}$	1	?	5	4	**0**	**0.80**
$x^{(4)}$	?	2	4	5	?	?
$x^{(5)}$	?	0	5	?	?	?

- Supposons: $\theta^{(1)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$ $\theta^{(2)} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$ $\theta^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$ $\theta^{(4)} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$ $x^{(3)} = \begin{bmatrix} 1 \\ 0 \\ 0.80 \end{bmatrix}$

Exercice

- Considérons les notes suivantes et une feature unique x_1.

	Utilisateur 1	Utilisateur 2	Utilisateur 3	x_1
Série 1	0	1.5	2.5	?

- On suppose $\theta^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\theta^{(2)} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ $\theta^{(3)} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$

- Quelle est une valeur possible de $x_1^{(1)}$?

- Réponse: 0.5

Exercice

- Considérons les notes suivantes et une feature unique x_1.

	Utilisateur 1	Utilisateur 2	Utilisateur 3	x_1
Série 1	0	1.5	2.5	?

Comment apprendre le vecteur x_1 ?

- On suppose $\theta^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\theta^{(2)} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ $\theta^{(3)} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$

- Quelle est une valeur possible de ${x_1}^{(1)}$?

- Réponse: 0.5

Exercice

- Considérons les notes suivantes et une feature unique x_1.

	Utilisateur 1	Utilisateur 2	Utilisateur 3	x_1
Série 1	0	1.5	2.5	?

Par régression linéaire simple !

- On suppose $\theta^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\theta^{(2)} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ $\theta^{(3)} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$

- Quelle est une valeur possible de ${x_1}^{(1)}$?

- Réponse: 0.5

Filtrage collaboratif

- Pour apprendre $x^{(i)}$, étant donné $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(nu)}$:

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

$$\min_{x^{(i)}} J(x^{(i)})$$

Filtrage collaboratif

- Pour apprendre $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$, ..., $\boldsymbol{x}^{(nm)}$, étant donné $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, ..., $\boldsymbol{\theta}^{(nu)}$:

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

$$\min_{x^{(1)}, x^{(2)}, \cdots, x^{(n_s)}} J(x^{(1)}, x^{(2)}, \cdots, x^{(n_s)})$$

Exercice

- Supposons que la descente de gradient soit utilisée pour minimiser:

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

- Pour $k \neq 0$, quelle opération de mise à jour est correcte ?

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

B. $x_k^{(i)} := x_k^{(i)} - \alpha \sum_{j:r^{(i,j)}=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \times \theta_k^{(j)}$

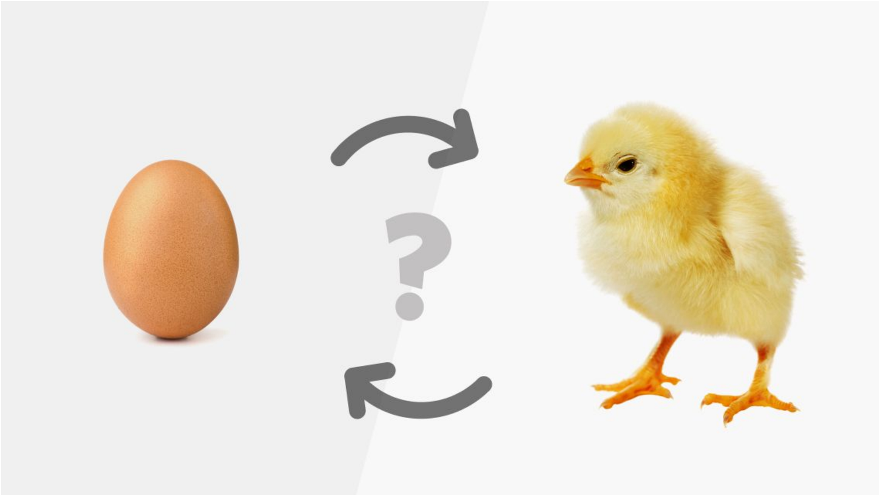
C. $x_k^{(i)} := x_k^{(i)} + \alpha \left[\sum_{j:r^{(i,j)}=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)}) \times \theta_k^{(j)} + \lambda x_k^{(i)} \right]$

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

- Réponse: D

Pour résumer ...

- Estimer $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(nm)}$, étant donné $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(nu)}$ et les notes
- Estimer $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(nu)}$, étant donné $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(nm)}$ et les notes



Filtrage collaboratif

- Estimer $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ..., \boldsymbol{x}^{(nm)}$, étant donné $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(nu)}$ et les notes
- Estimer $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(nu)}$, étant donné $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ..., \boldsymbol{x}^{(nm)}$ et les notes
- Une possibilité est de partir de **suppositions sur $\boldsymbol{\theta}$** et d'estimer $\boldsymbol{x}$, puis d'estimer θ, etc...

Supposition $\boldsymbol{\theta} \rightarrow \boldsymbol{x} \rightarrow \boldsymbol{\theta} \rightarrow \boldsymbol{x} \rightarrow \boldsymbol{\theta} \rightarrow \boldsymbol{x} \rightarrow$ ...

- L'algorithme converge magiquement ...
- … mais il y a mieux !

Filtrage collaboratif

- Minimiser selon $x^{(1)}, x^{(2)}, ..., x^{(nm)}$ et $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(nu)}$ **simultanément** !

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

- Plus besoin de $x_0 \rightarrow \theta$ et $x \in \mathbb{R}^n$

Algorithme de filtrage collaboratif

1. Initialiser $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(nm)}$ et $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(nu)}$ avec des **faibles valeurs aléatoires**
2. Minimiser J($\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(nm)}, \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(nu)}$) par l'algorithme de **descente du gradient**
3. Réaliser la **prédiction** $\theta^{\mathsf{T}}\boldsymbol{x}$ de la note pour un utilisateur de paramètres $\boldsymbol{\theta}$ et une série de profile $\boldsymbol{x}$

Descente de gradient: dérivées partielles

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

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

Exercice

- Dans l'algorithme présenté précédemment, pourquoi est-il nécessaire d'initialiser $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$, ..., $\boldsymbol{x}^{(nm)}$ et $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, ..., $\boldsymbol{\theta}^{(nu)}$ avec des **faibles valeurs aléatoires ?**

- Réponse: il est important de briser la symétrie afin que l'algorithme puisse apprendre des paramètres différents les uns des autres (similaire à l'initialisation des réseaux de neurones)

Exercice

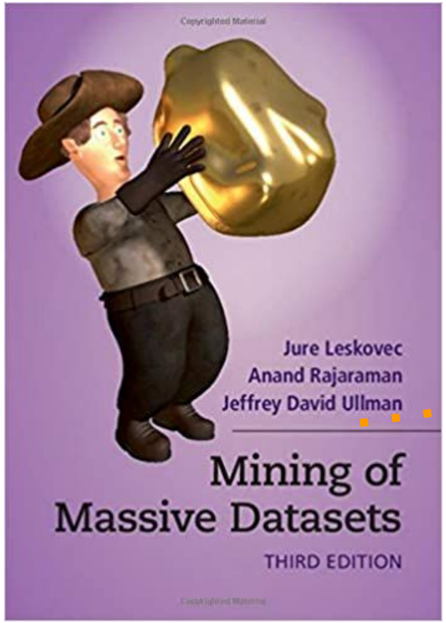
- Pouvez vous nommer quelques **hyperparamètres** propres à l’algorithme de filtrage collaboratif ?

- Réponse:
  - Learning rate α
  - Nombre d’itérations de la descente de gradient
  - Paramètres d’initialisation
  - Nombre de features n

Filtrage collaboratif

- **Avantages**
  - Fonctionne pour n'importe quel type de produits
- **Inconvénients**
  - Parcimonie (matrice d'utilité creuse)
  - Problème de la première note pour un nouveau produit ou sans notations
  - Biais de popularité
  - Problème du démarrage à froid (il faut un certain nombre d'utilisateurs)

Sommaire



9 Recommendation Systems
p. 319-353

Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman**, Mining of Massive Datasets, 3rd edition**

Références

[1] CS229: Machine Learning - Stanford University

[2] Mining of Massive Datasets, 3rd edition