Systèmes de recommandation

Illustrations issues du cours en ligne de l’université de Stanford

Exemple Netflix Price

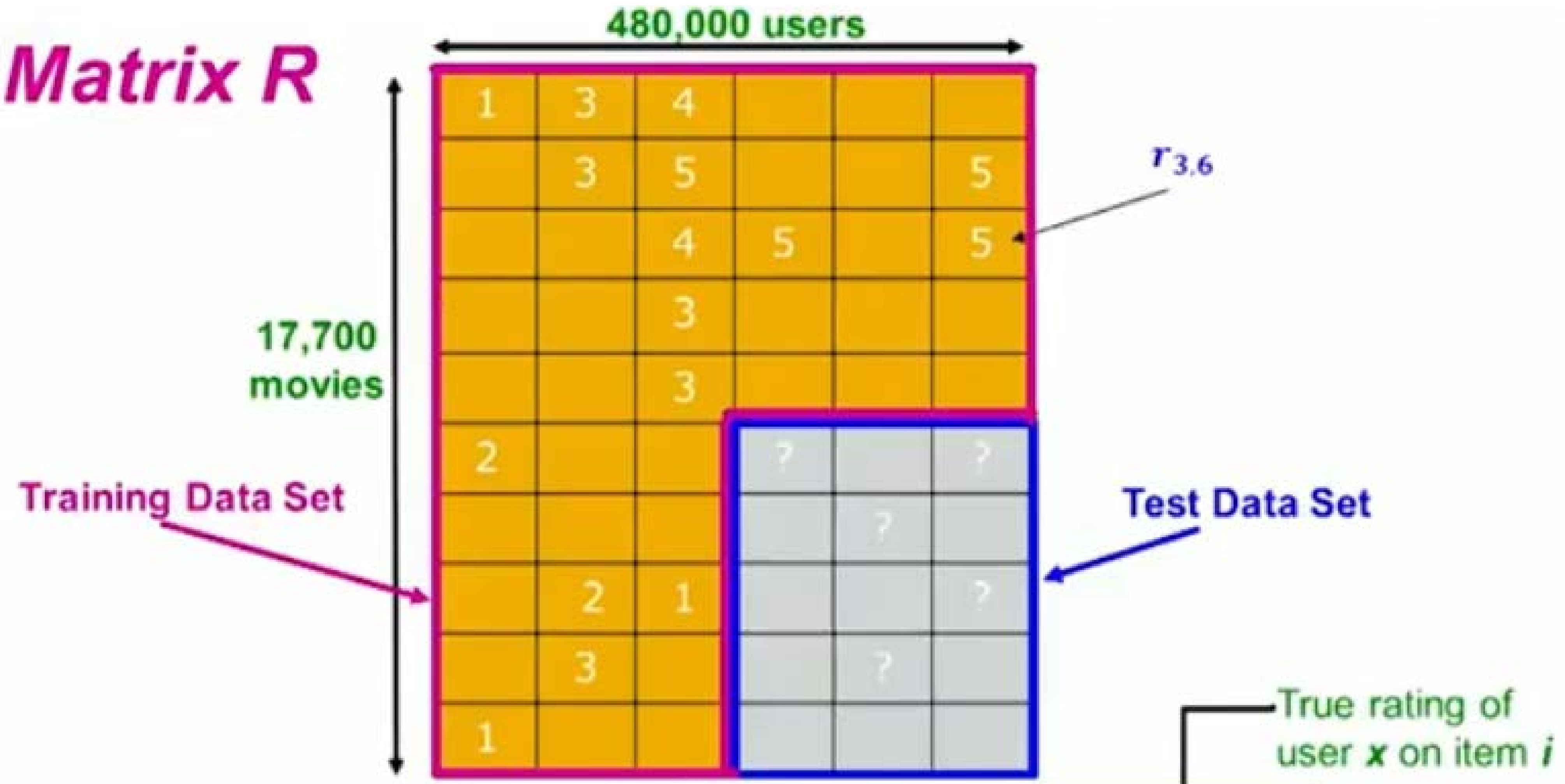
- Système de recommandation de films
- Données d’entraînement
  - 100 millions de notes
  - 480 000 utilisateurs
  - 17 770 films
  - Données de 2000-2005
- Données Test
  - Les dernières notes de chaque utilisateur (2.8 millions)

Matrix R

480,000 users

17,700 movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2			2		2
				5	
	2	1			1
	3			3	
1					



True rating of user *x* on item *i*

$$\textbf{RMSE} = \frac{1}{|R|}\sqrt{\sum_{(i,x)\in R}(\hat{r}_{xi} - r_{xi})^2}$$

Predicted rating

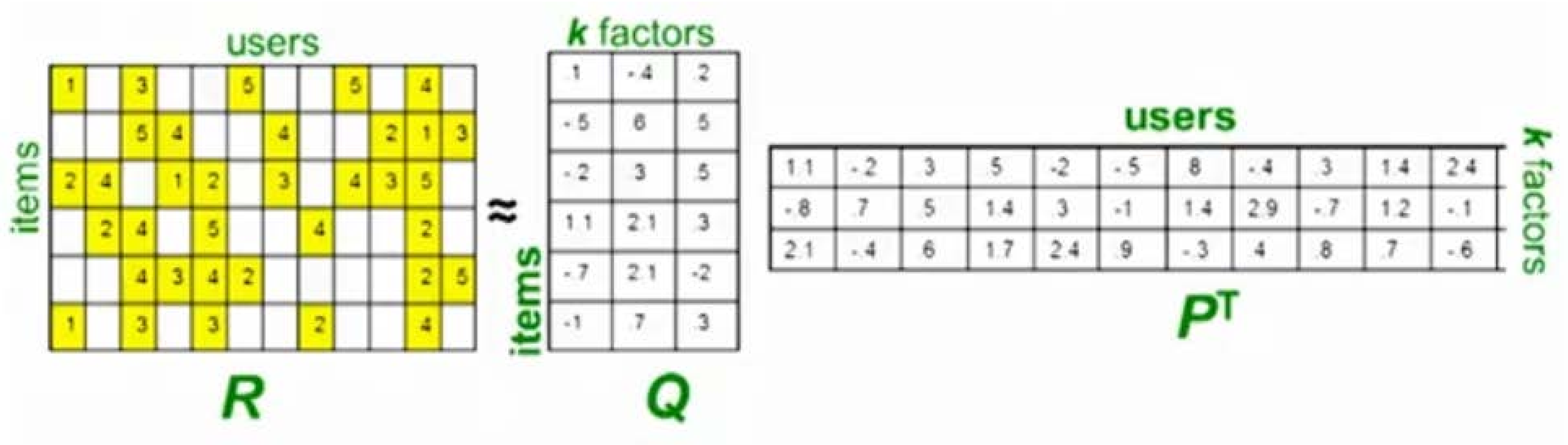
Approches

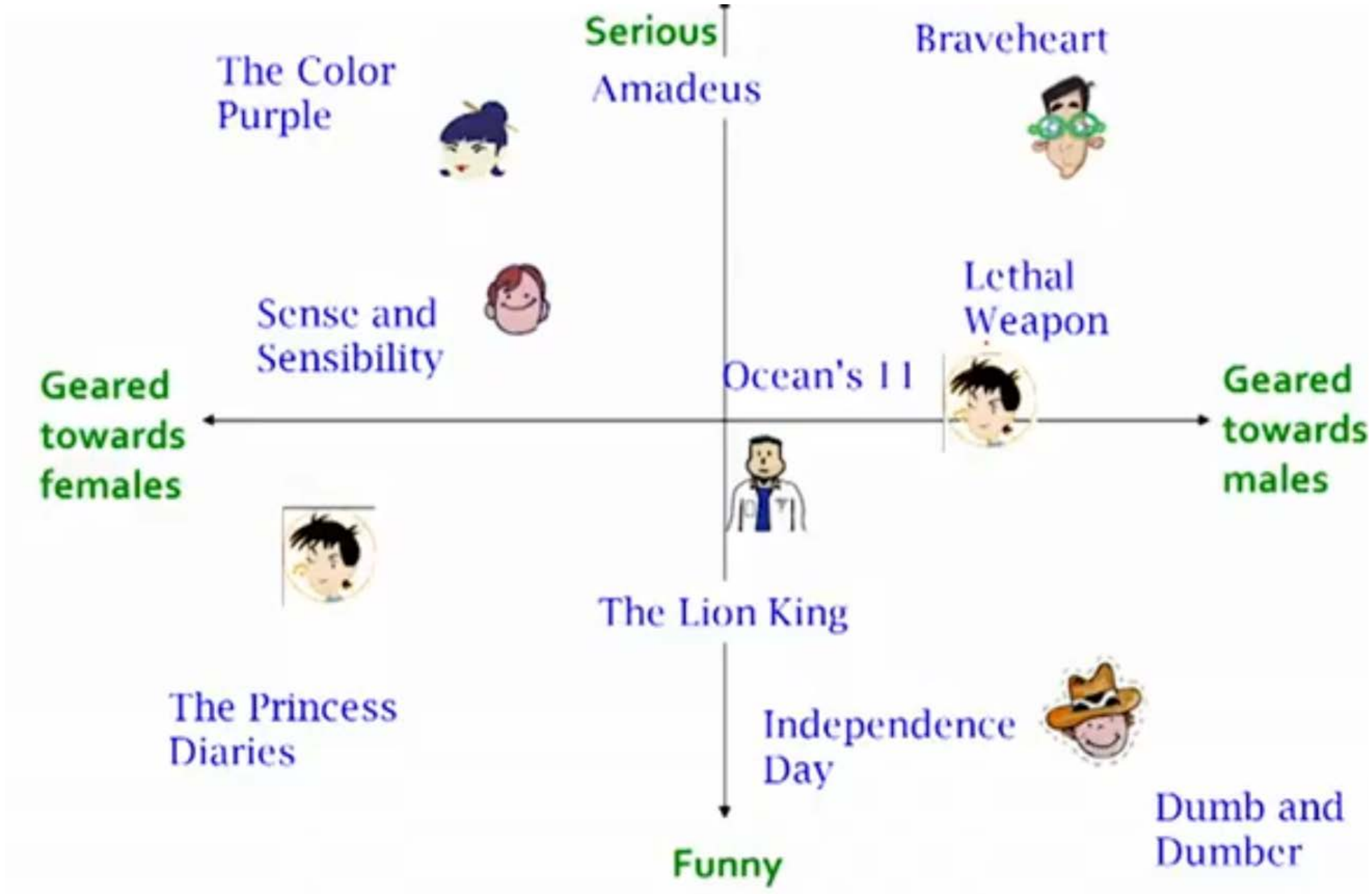
- Globale
  - Moyenne des notes des films : 3.7
  - Film i a en moyenne 0.5 points de +
  - Utilisateur x note 0.2 points en dessous de la moyenne
  - -> note de l’utilisateur x pour film i : 3.7+0.5-0.2 =4

Approche SVD

SVD: $A = U \Sigma V^T$

$$R \approx Q \cdot P^T$$





Serious
Braveheart
The Color Purple
Amadeus
Lethal Weapon
Sense and Sensibility
Ocean's 11
Geared towards females
Geared towards males
The Lion King
The Princess Diaries
Independence Day
Dumb and Dumber
Funny

users

items

1		3			5			5		4	
		5	4	?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

$\approx$

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

k factors

Q

$\bullet$

users

k factors

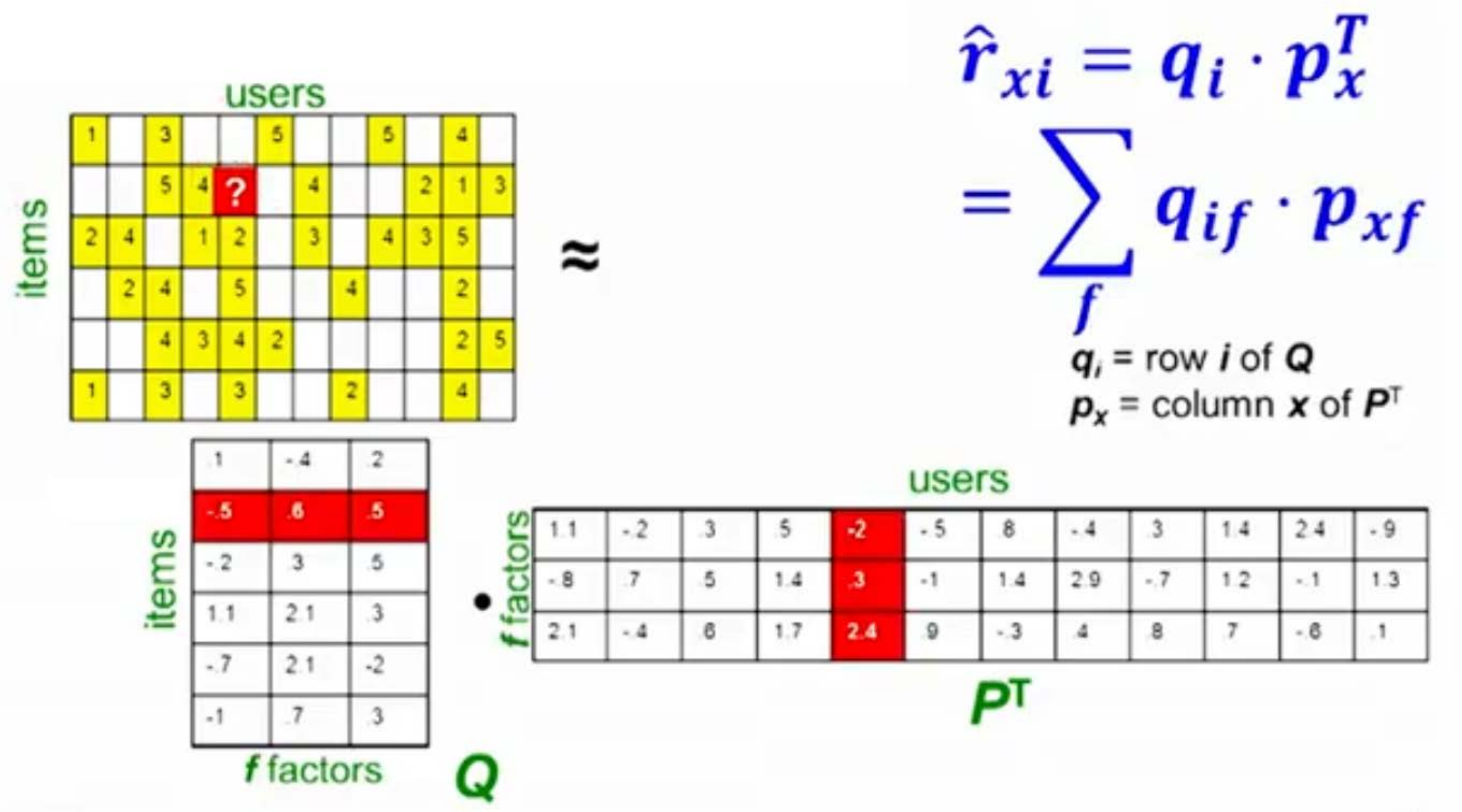
1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

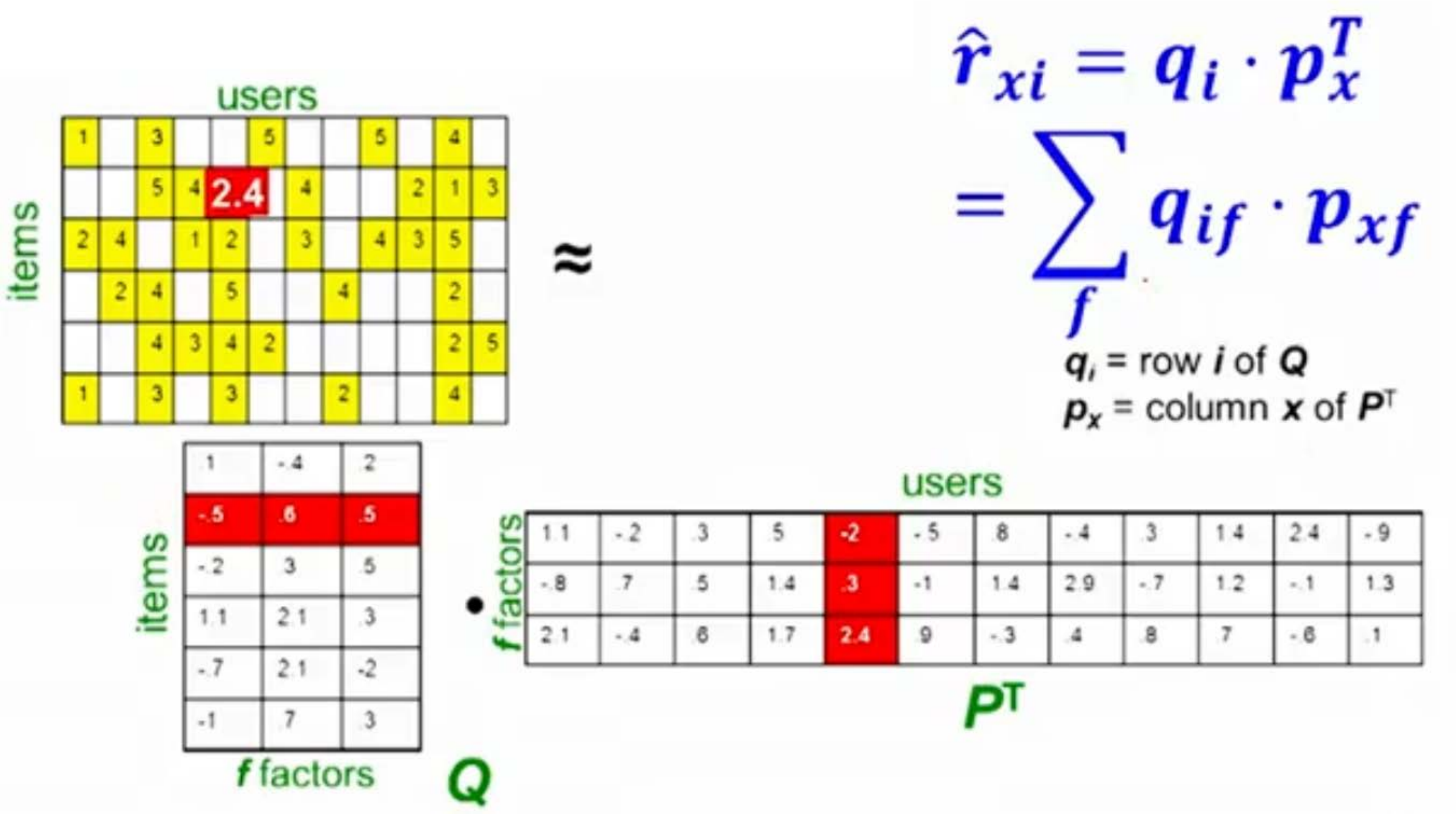
P$^{\mathrm{T}}$

$$\hat{r}_{xi} = q_i \cdot p_x^T$$
$$= \sum_k q_{ik} \cdot p_{xk}$$

q_i = row *i* of ***Q***

p_x = column *x* of P^T


users
items
≈
$\hat{r}_{xi} = q_i \cdot p_x^T$
$= \sum_f q_{if} \cdot p_{xf}$
q_i = row i of Q
p_x = column x of P^T
users
f factors
P^T
items
f factors
Q


users
items
2.4
≈
f factors
Q
users
f factors
P^T
$\hat{r}_{xi} = q_i \cdot p_x^T$
$= \sum_f q_{if} \cdot p_{xf}$
q_i = row i of Q
p_x = column x of P^T

SVD

- SVD donne l'erreur de reconstruction minimale (somme des erreurs au carré SSE)

$$\min_{U,V,\Sigma} \sum_{ij\in A} \left(A_{ij} - [U\Sigma V^{\mathrm{T}}]_{ij}\right)^2$$

- SSE et RMSE sont liés

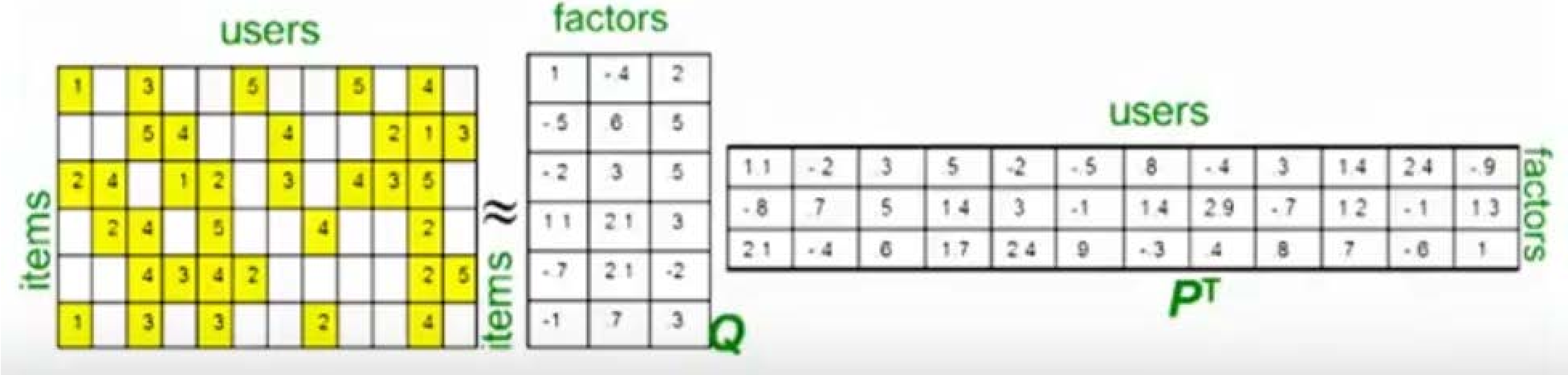
$$RMSE = \frac{1}{c}\sqrt{SSE}$$

-> SVD minimise RMSE

- Pb : somme dans le terme d'erreur : sur toutes les entrées (notes absentes interprétées comme des 0)
- Mais données manquantes

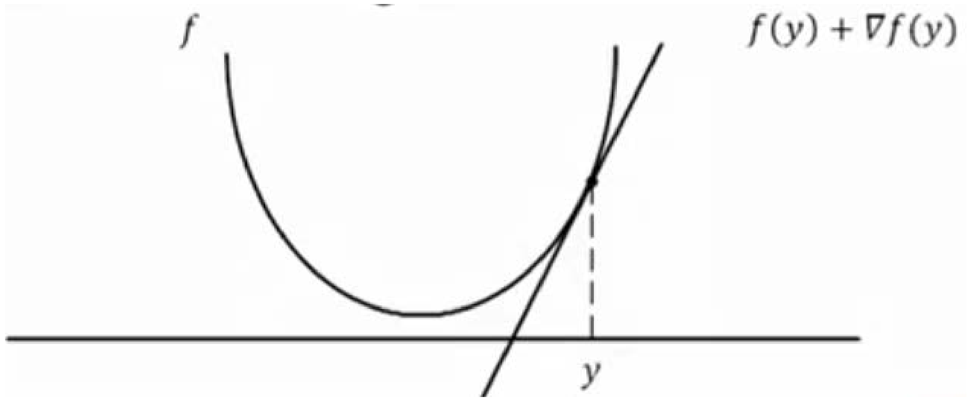
Objectif : trouver P et Q tq

$$\min_{P,Q} \sum_{(i,x)\in R} \left(r_{xi} - q_i \cdot p_x^T\right)^2$$



Rappel : Minimisation d'une fonction

- Minimisation de la fonction f(x)
  - Calcul de la dérivée ∇f
  - Partir d'un point donné et évaluer $\nabla f(y)$
  - Pas dans la direction inverse du gradient $y = y - \nabla f(y)$
  - Répéter jusqu'à convergence



- On cherche à minimiser SSE pour des données **test** non connues
- Idée : minimiser SSE sur les données d'**entraînement**
  - On cherche des valeurs de k (nb de facteurs) pour capturer tout le signal
  - Mais sur les données **test** SSE commence à augmenter pour k>2

- C'est un exemple classique d'**overfitting**
  - Avec trop de degrés de liberté (trop de paramètres libres), le modèle commence à ajuster du bruit
    - -> ajuste trop bien sur les données d'entraînement mais ne généralise pas bien aux données tests inconnues

Traitement des données manquantes

- Pour résoudre le pb d’overfitting -> on introduit une régularisation
  - Modèle riche quand les données sont suffisantes

$$\min_{P,Q} \underbrace{\sum_{training} (r_{xi} - q_i p_x^T)^2}_{\text{“error”}} + \lambda \underbrace{\left[\sum_x \|p_x\|^2 + \sum_i \|q_i\|^2 \right]}_{\text{“length”}}$$

λ… user set regularization parameter

Descente de gradient stochastique

- On cherche les matrices P et Q tq

$$\min_{P,Q} \sum_{training} (r_{xi} - q_i \, p_x^T)^2 + \lambda \left[\sum_x \|p_x\|^2 + \sum_i \|q_i\|^2 \right]$$

- Descente de gradient
  - Initialisation de P et Q (en utilisant la SVD, notes manquantes mises à 0)

- $\mathbf{P} \leftarrow \mathbf{P} - \eta \cdot \nabla \mathbf{P}$
- $\mathbf{Q} \leftarrow \mathbf{Q} - \eta \cdot \nabla \mathbf{Q}$

where $\nabla \boldsymbol{Q}$ is gradient/derivative of matrix $\boldsymbol{Q}$:

$\nabla Q = [\nabla q_{ik}]$ and $\nabla q_{ik} = \sum_{xi} -2(r_{xi} - q_i p_x^T) p_{xk} + 2\lambda q_{ik}$

- Here $\boldsymbol{q_{ik}}$ is entry $\boldsymbol{k}$ of row $\boldsymbol{q_i}$ of matrix $\boldsymbol{Q}$
- **And similarly for $\nabla \boldsymbol{P}$**