Recommendation Systems

- Maria Papadopouli, University of Crete
  https://www.csd.uoc.gr/~hy539/lectures/Recommendation_Systems.pptx
- Wikipedia
  https://en.wikipedia.org/wiki/Collaborative_filtering

Significant industrial & research interest in recommendation systems:

increases its watch times by 50% per year

35% of all sales are generated by recommendations

Ratings matrix

$r_{u,i}$

Objects (e.g., movies, products)

Users

	i_1	...	i_k	...	i_m
u_1	★★★★★		★★★★★		★☆☆☆☆
⋮					
u_j	★★★☆☆		?		★★☆☆☆
⋮					
u_n	★★☆☆☆		★★★★☆		★★☆☆☆

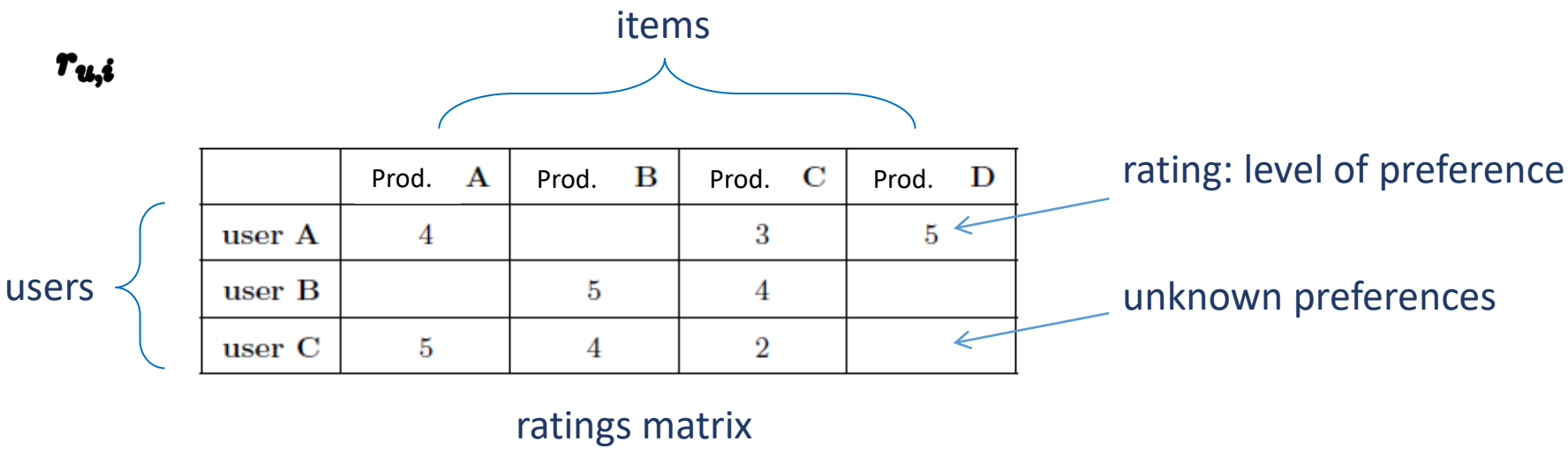
Sparse matrix: Most values are unknown

Predicting: The task of filling the unknown values

Recommendation Systems

Infer user preferences about items

Produce **highly relevant** & **personalized** lists of items



	Prod. A	Prod. B	Prod. C	Prod. D
user A	4		3	5
user B		5	4	
user C	5	4	2	

Main Approaches

Collaborative filtering (CF)

Inspect rating patters to find similar users/items

Content-based (CB)

Analyze attributes of items for building user profiles

In general, CF performs **better** than CB

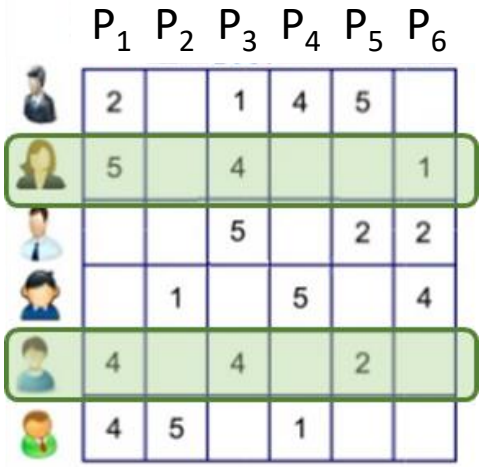
- CF fail to provide accurate predictions with insufficient ratings
- CB can alleviate the sparsity problem

Popular Recommenders: K-Nearest Neighbors Approach

User-based collaborative filtering (UBCF)

Assumption: Similar users have similar preferences

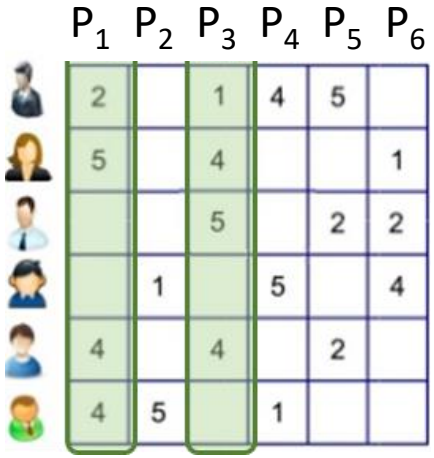
- **User similarity: agreement on co-rated items**
- Prediction: **weighted sum of similar user's ratings**



Item-based collaborative filtering (IBCF)

Assumption: Users have similar tastes for similar items

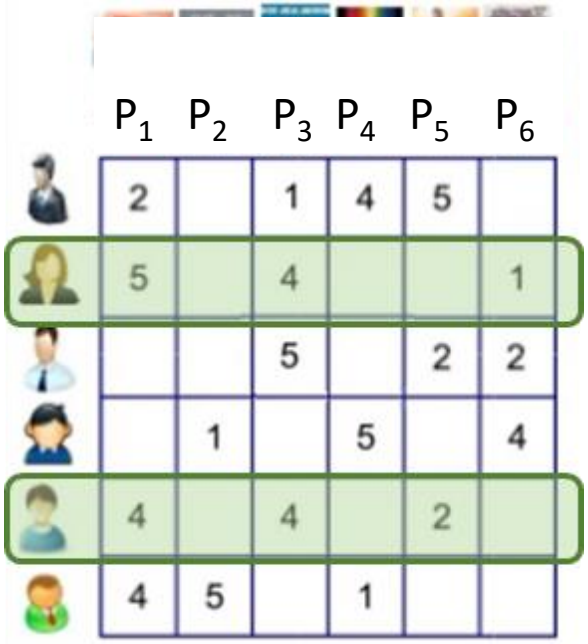
- Item similarity: agreement within users rated both items
- Prediction: weighted sum of similar items' ratings



User-based collaborative filtering (UBCF)

Assumption: Similar users have similar preferences

- **User similarity: agreement on co-rated items**
- Prediction: **weighted sum of similar user's ratings**



$r_{u,i}$

$$r_{u,i} = k \sum_{u' \in U} \text{simil}(u,u') r_{u',i} \qquad k = 1 / \sum_{u' \in U} |\text{simil}(u,u')|$$

$$\text{simil}(x,y) = \cos(\vec{x},\vec{y}) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \times ||\vec{y}||} = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}}$$

IBCF

Compute similarity between items

set of users rated both items

rating on item *i*

rating on item j

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}}$$

deviation from the average rating

Prediction of rating for item *i*

most similar items to *i*

similarity as weight

rating on similar item

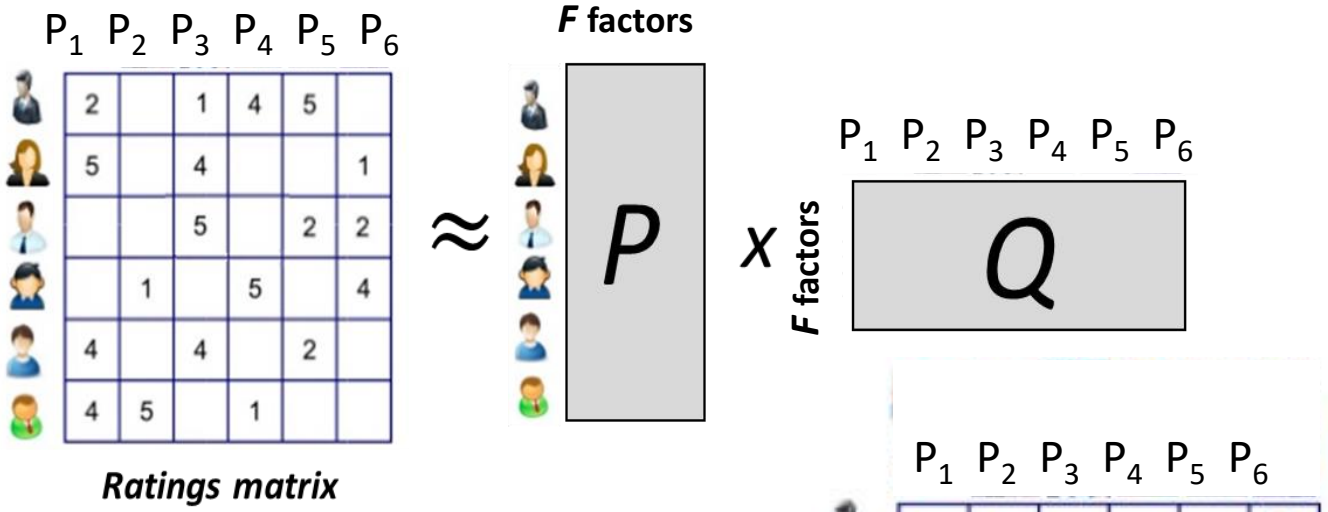
$$p_{u,i} = \frac{\sum_{j \in S} sim(i,j) R_{u,j}}{\sum_{j \in S} |sim(i,j)|}$$

SVD: Matrix Factorization Approach

Decomposition of ratings matrix ***R*** into the product of ***P*** & ***Q***

Each user & item is described with ***F*** latent features

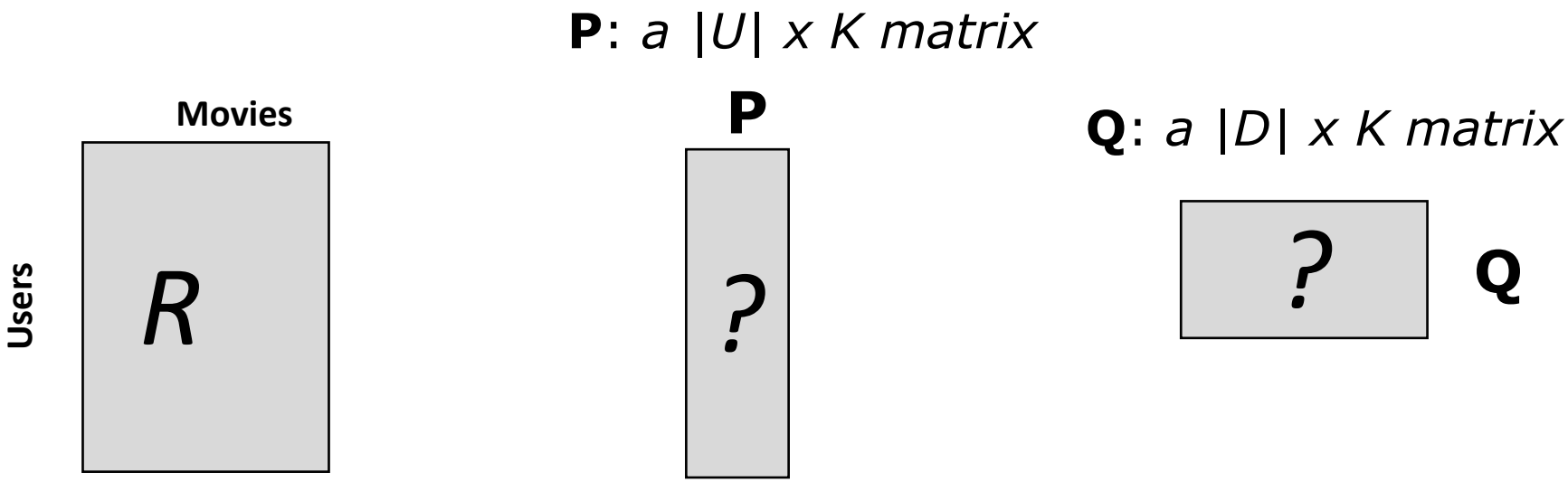
- ***P*** : user factors
- ***Q*** : item factors



Prediction for user u about item i :

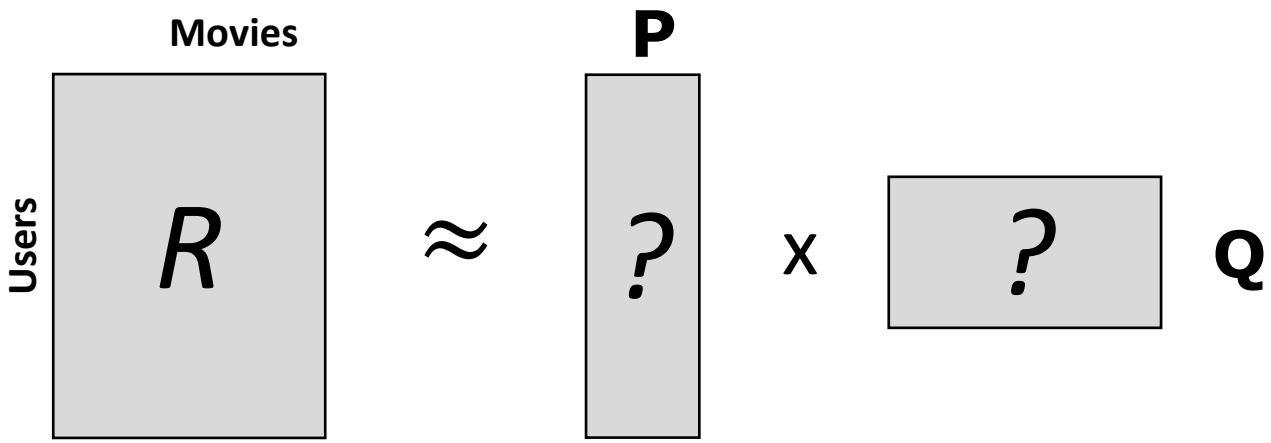
$$p_{u,i} = q_i^T p_u = \sum_{f=1}^{F} q_{i,f} \cdot p_{u,f}$$

Matrix - Factorization



R: *Matrix of size |U| x |D| that contains all the ratings that the users have assigned to the items*

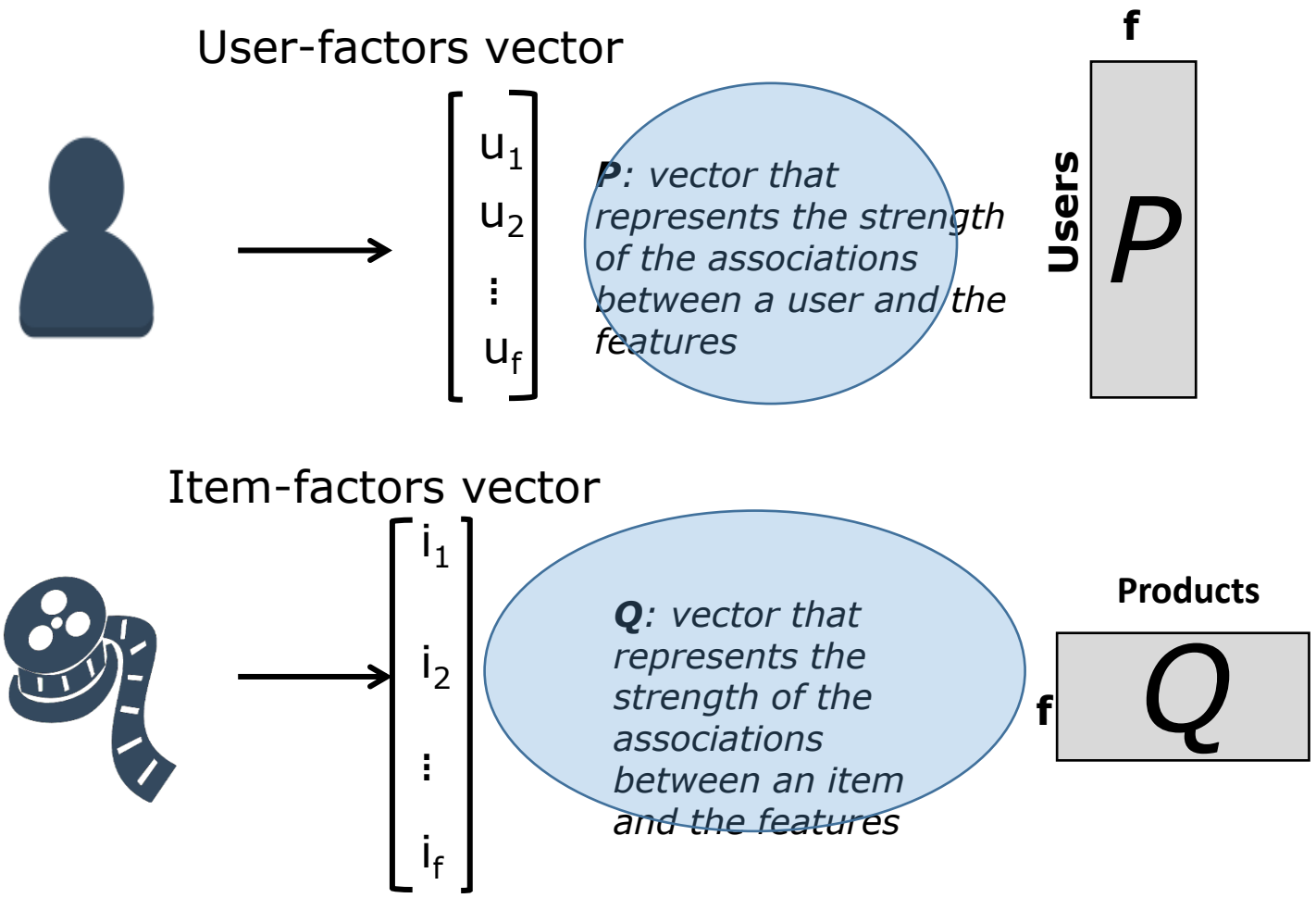
Matrix - Factorization: Ratings matrix is approximated



Find two matrices P and Q such that product approximates R

Matrix - Factorization:

Each user & item is characterized with a vector of factors



Factors - *f*: mathematic artifact for computing relationship

Matrix - Factorization:

Preference - prediction is the dot product of user & item factors

Item factors: The extent to which an item has some characteristics

User factors: Level of preference for the corresponding characteristics

Matrix - Factorization:

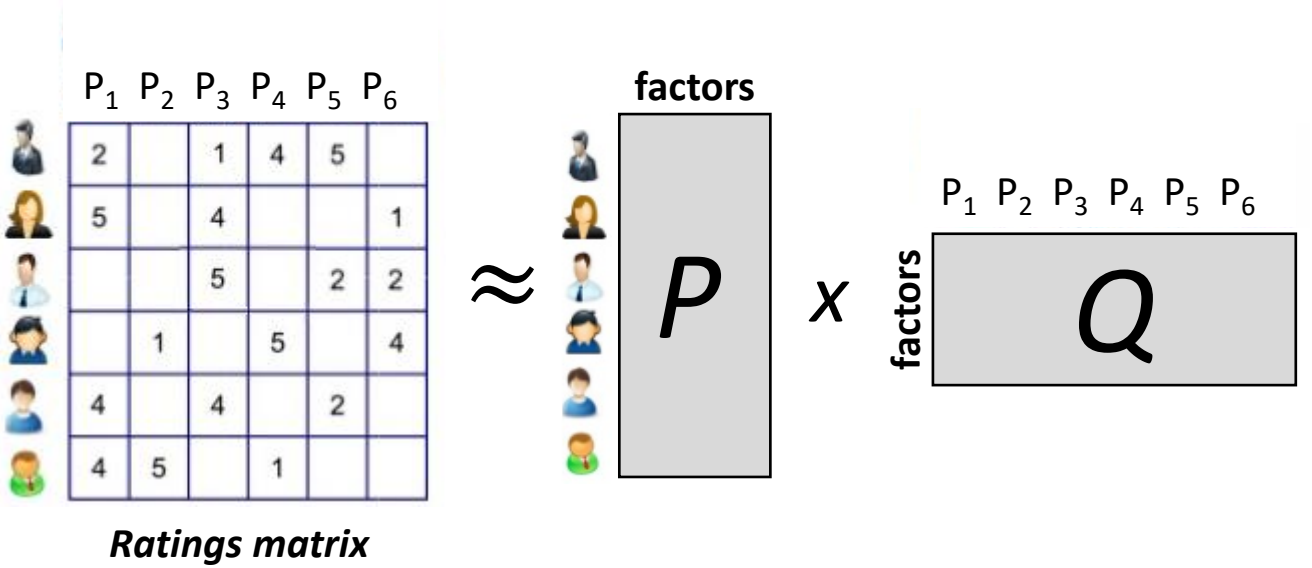
Preference - prediction is the dot product of user & item factors

To get the prediction of a rating of an **item $\mathbf{d}_j$** by a **user u_i** calculate the dot product of the two vectors corresponding to u_i and d_j

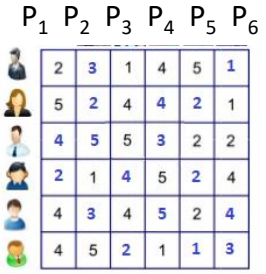
Vector corresponding to d_j

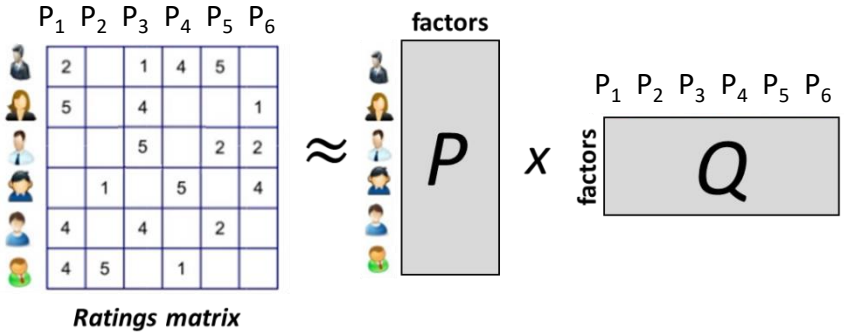
$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^{k} p_{ik} q_{kj}$$

Vector corresponding to u_i



Ratings matrix





Ratings matrix