Recommendation Systems

Matrix Factorization. Part I

Eugeny Malyutin / Sergey Dudorov

Problem statement

	i_1	i_2	i_3	i_4	i_5
u_1	5	?	4	1	?
u_2	?	3	?	3	?
u_3	?	2	4	4	1
u_4	4	4	5	?	?
u_5	2	4	?	5	2

- I - set of items
- U - set of users
- R - rating matrix
- A_{ij} - User-Item score

Matrix Factorization

Item

User

	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

Rating Matrix

=

A	*1.2*	*0.8*
B	*1.4*	*0.9*
C	*1.5*	*1.0*
D	*1.2*	*0.8*

User Matrix

X

W	X	Y	Z
1.5	*1.2*	*1.0*	*0.8*
1.7	*0.6*	*1.1*	*0.4*

Item Matrix

X – *Rating matrix*

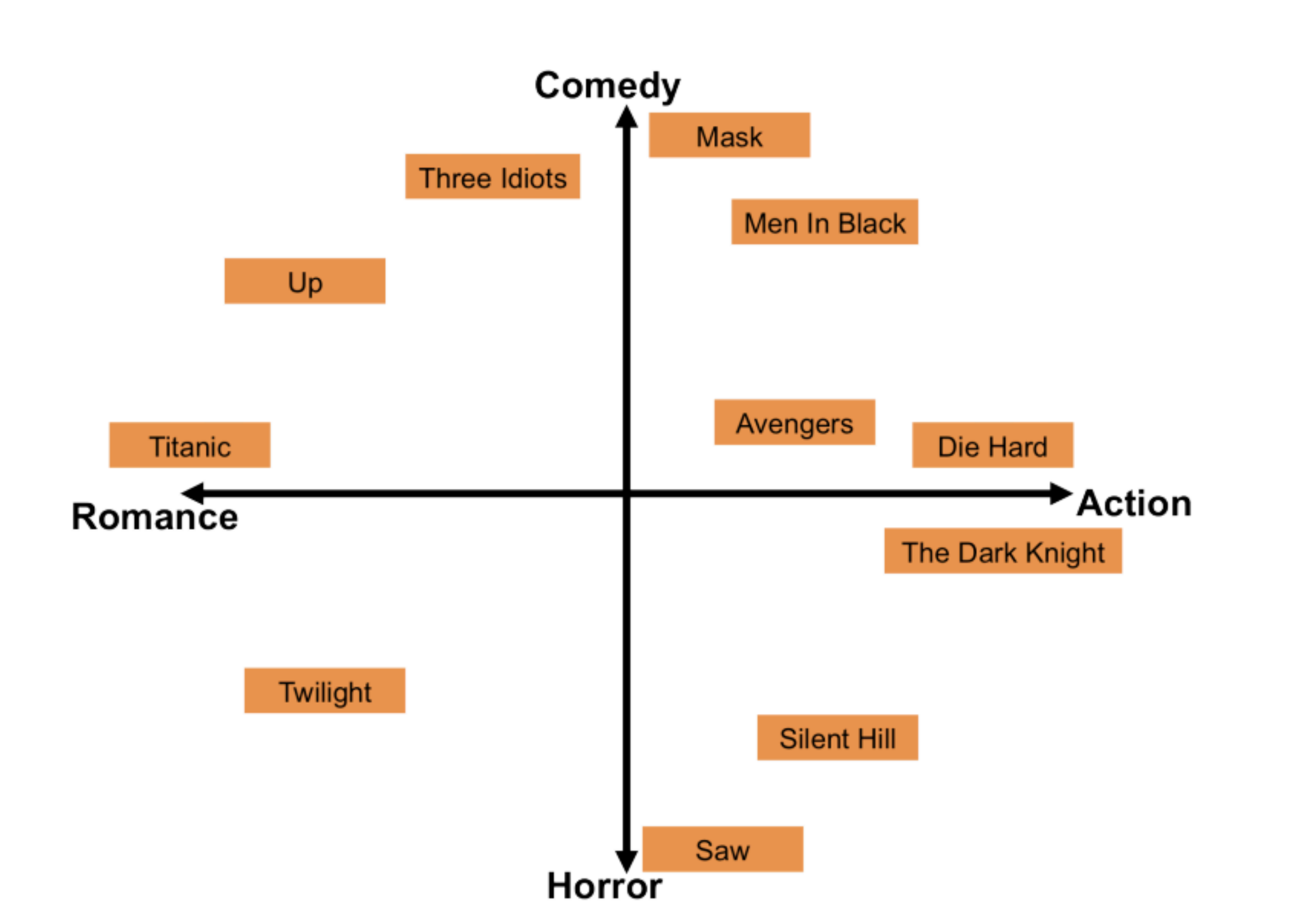
U – *User embedding matrix*

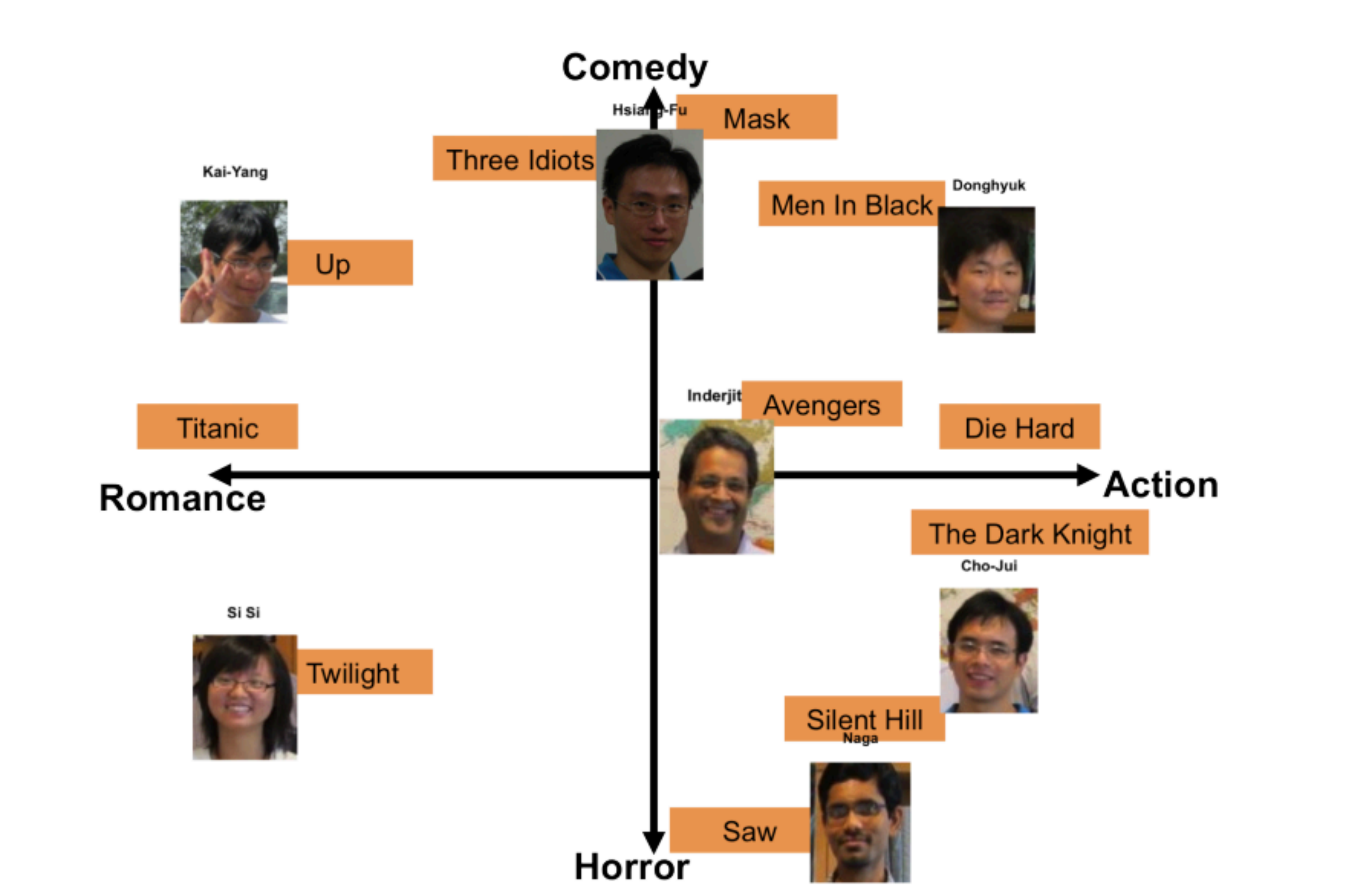
V – *Item embedding matrix*

k – *embedding size*

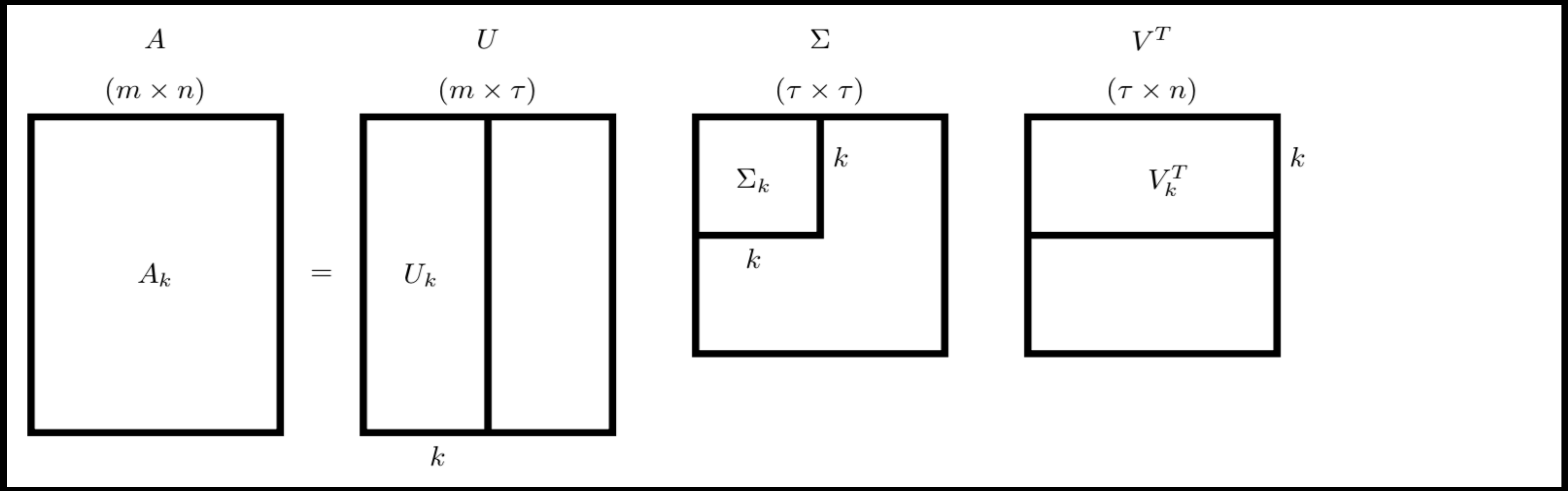
$$X_{l,n} \approx U_{l,k} \,.\, V^T_{k,n}$$

Matrix Factorization





Singular value decomposition



U, V – *Orthogonal matricies*

Σ – *Diagonal matrix*

$$A = \tilde{U}\Sigma\tilde{V}^T$$

$$U = \tilde{U}_k\Sigma_k \,,\, V = \tilde{V}_k$$

$$U = \tilde{U}_k \,,\, V = \tilde{V}_k\Sigma_k$$

$$U = \tilde{U}_k\sqrt{\Sigma_k} \,,\, V = \tilde{V}_k\sqrt{\Sigma_k}$$

Singular value decomposition

How to fill ?

	i_1	i_2	i_3	i_4	i_5
u_1	5	?	4	1	?
u_2	?	3	?	3	?
u_3	?	2	4	4	1
u_4	4	4	5	?	?
u_5	2	4	?	5	2

- Zeros
- Expectation User
- Expectation Item
- Advanced methods

SVD in Recs

Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

Rating Matrix

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

User Matrix

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

Item Matrix

$$X = UV^T$$

$$\|X - UV^T\| \to min$$

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

$$\sum_{i,j} (x_{ij} - \langle u_i, v_j \rangle)^2 \to min$$

SGD

SGD Algorithm for MF

Input: training matrix V, the number of features K, regularization parameter λ, learning rate ϵ
Output: row related model matrix W and column related model matrix H

Initialize W, H to $UniformReal(0, \frac{1}{\sqrt{K}})$
repeat
  for random $V_{ij} \in V$ **do**
    $error = W_{i*}H_{*j} - V_{ij}$
    $W_{i*} = W_{i*} - \epsilon(error \cdot H_{*j}^{\intercal} + \lambda W_{i*})$
    $H_{*j} = H_{*j} - \epsilon(error \cdot W_{i*}^{\intercal} + \lambda H_{*j})$
  end for
until convergence

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \to min$$

$$\frac{dQ}{du_i} = \sum_{i,j} \frac{d}{du_i} (\langle u_i, v_j \rangle - x_{ij})^2 = \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) \frac{d\langle u_i, v_j \rangle}{du_i} = = \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) v_j$$

$$\varepsilon_{i,j} = (\langle u_i, v_j \rangle - x_{ij})$$

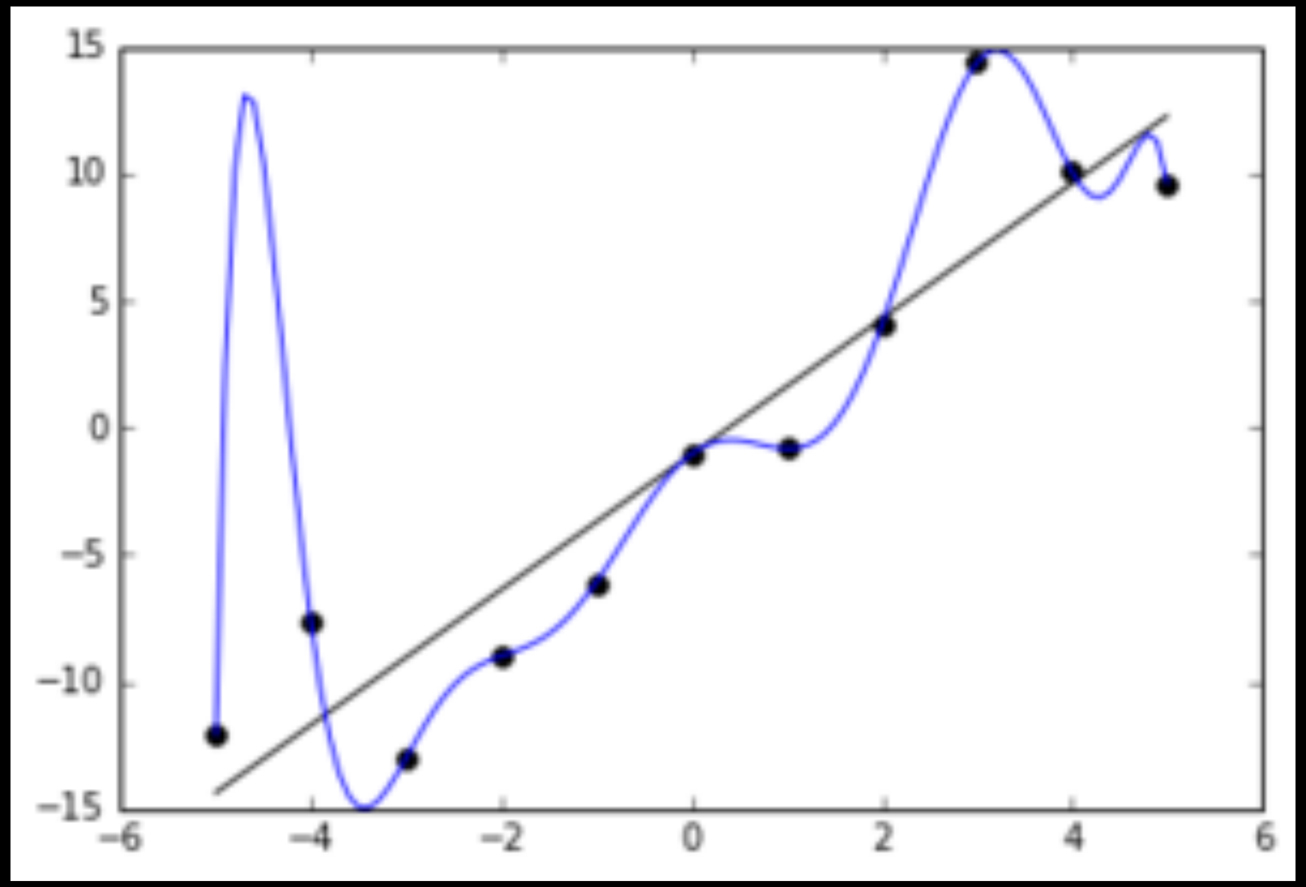
$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \sum_j \varepsilon_{i,j} v_j$$

$$v_j^{(t+1)} = v_i^{(t)} - \eta_t \sum_j \varepsilon_{i,j} u_i$$

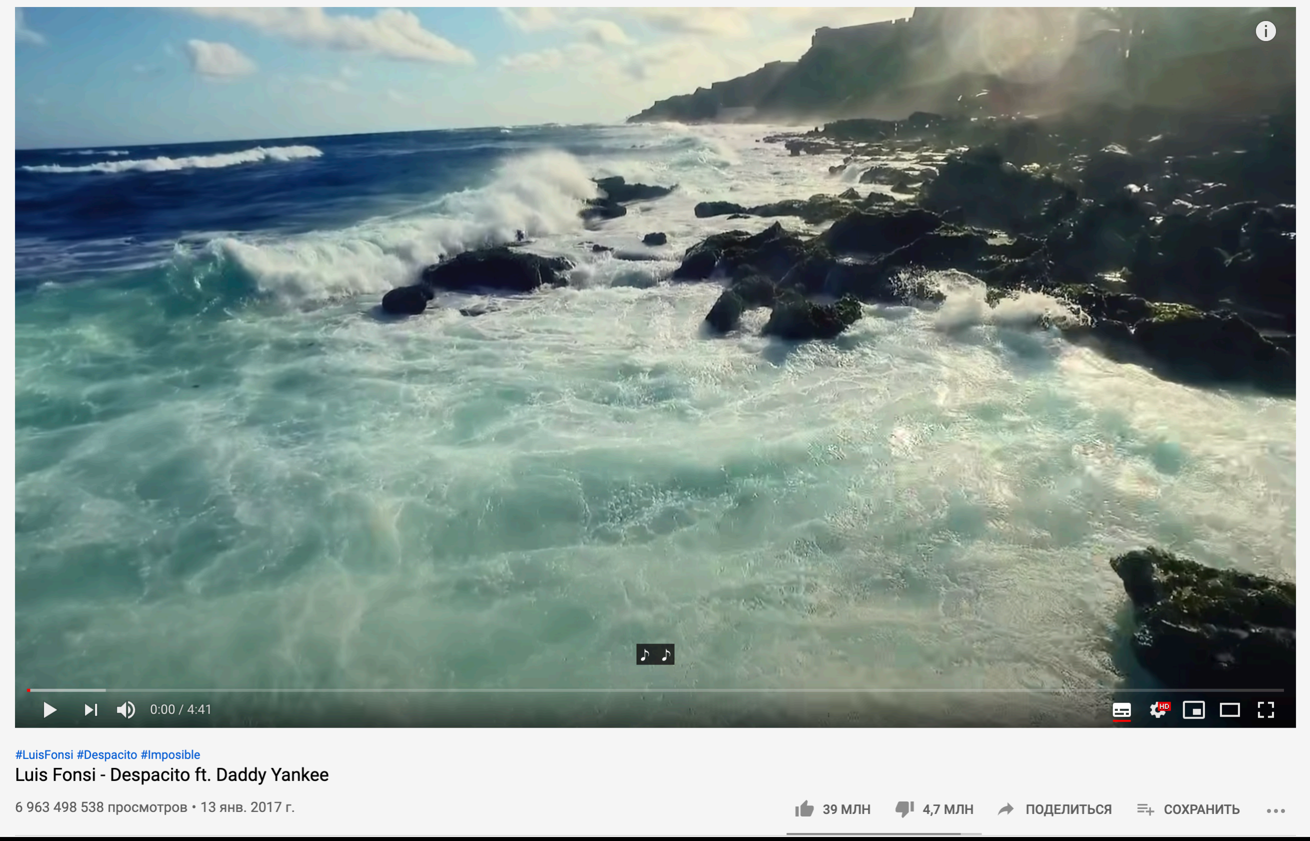
$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \varepsilon_{i,j} v_j$$

$$v_j^{(t+1)} = v_i^{(t)} - \eta_t \varepsilon_{i,j} u_i$$

Some problems?

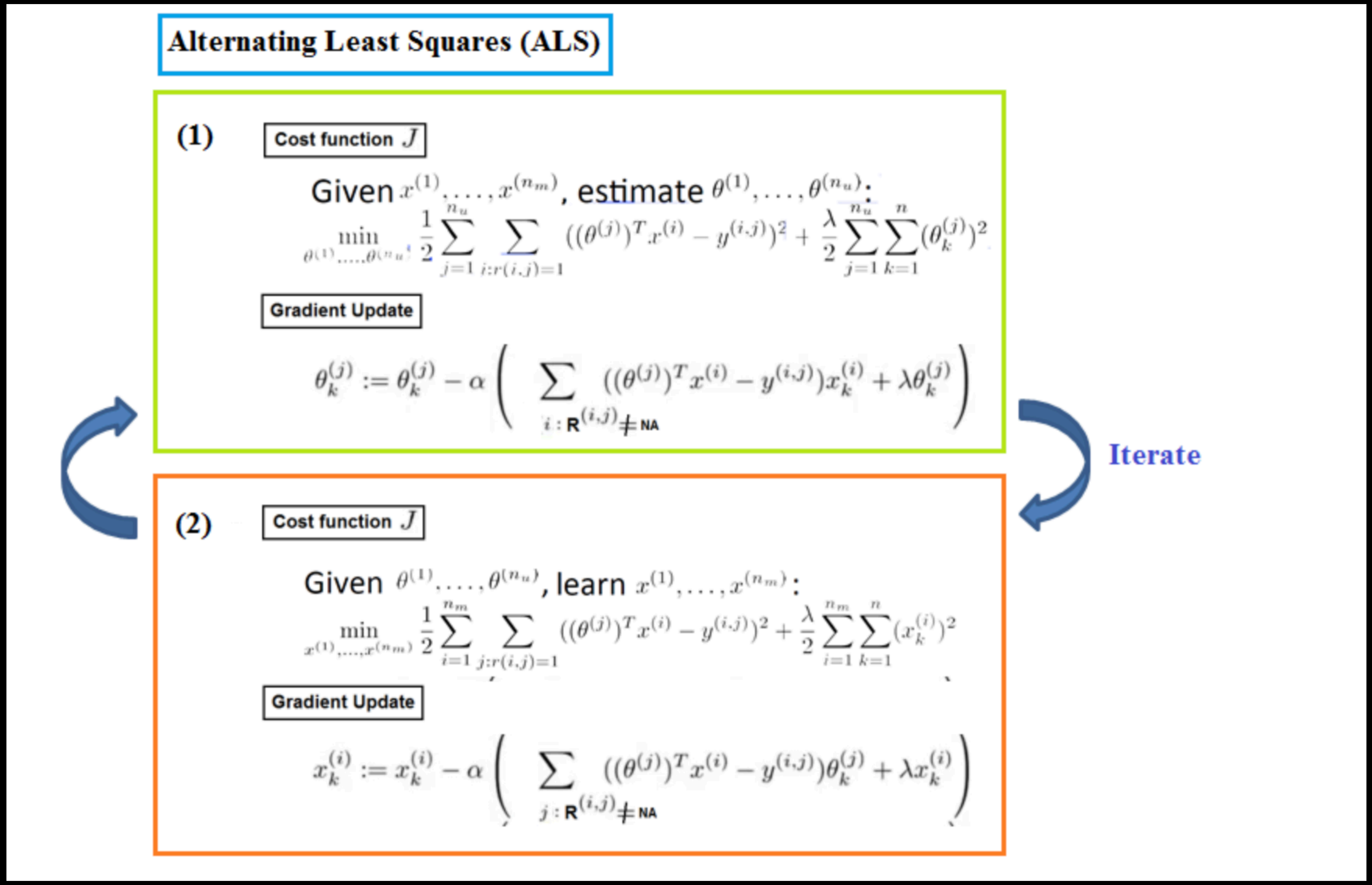


$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 \to min$$



$$Q = \sum_{i,j} (\langle u_i, v_j \rangle + b_u + b_i + \mu - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 + \gamma \|b_u\| + \theta \|b_i\| \to min$$

ALS



$$u_i = (\Upsilon^T \times \Upsilon + \lambda I)^{-}1 \times \Upsilon^T \times R_i$$

$$v_j = (U^T \times U + \lambda I)^{-1} \times U^T \times R_j$$

$$\frac{dQ}{du_i} = 0 \qquad \frac{dQ}{dv_i} = 0$$

$$\frac{dQ}{du_i} = \sum_j 2(\langle u_i, v_j\rangle - x_{i,j})v_j = 0$$

$$\sum_j v_j\langle v_j, u_i\rangle = \sum_j x_{i,j}v_j$$

$$\sum_j v_j v_j^T u_i = \sum_j x_{ij}v_j$$

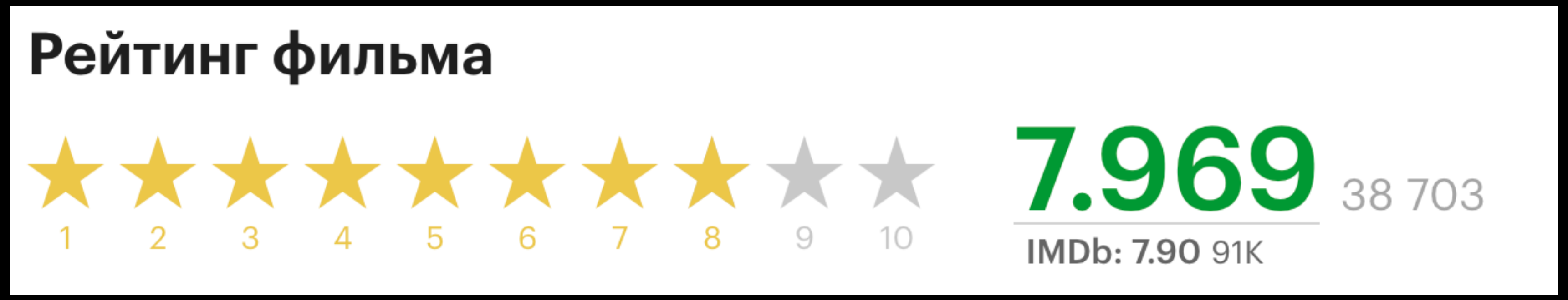
$$(\sum_j v_j v_j^T)u_i = \sum_j x_{ij}v_j$$

$$(\sum_j u_i u_i^T)v_j = \sum_i x_{ij}u_i$$

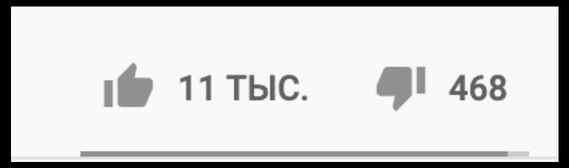
Explicit

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

Ratings



Likes/Dislikes



Timespent on item (intervals)

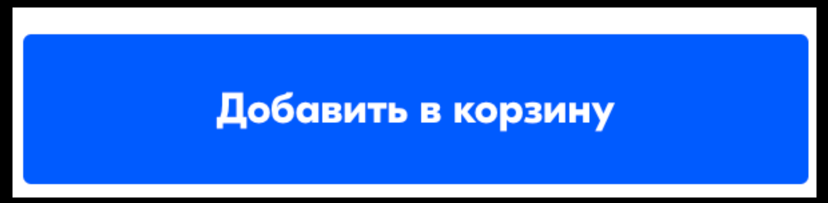
etc

Implicit

1		1			1		1
1						1	
	1		1	1			
	1					1	1
1		1	1		1		
	1			1		1	

- Likes

- Buy



- Watch full movie

- etc

How to work with implicit feedback

		1			1	
		1				1
1					1	
1						1
				1		
	1			1		
1			1		1	

$$u_i = (1,1,1,1,1...,1)$$

$$v_i = (1,1,1,1,1...,1)$$

0	0	1	0	0	1	0
0	0	1	0	0	0	1
1	0	0	0	0	1	0
1	0	0	0	0	0	1
0	0	0	0	1	0	0
0	1	0	0	1	0	0
1	0	0	1	0	1	0

$$\sum_{i,j} \omega_{ij}(\langle u_i, v_j \rangle - x_{ij})^2 \to min$$

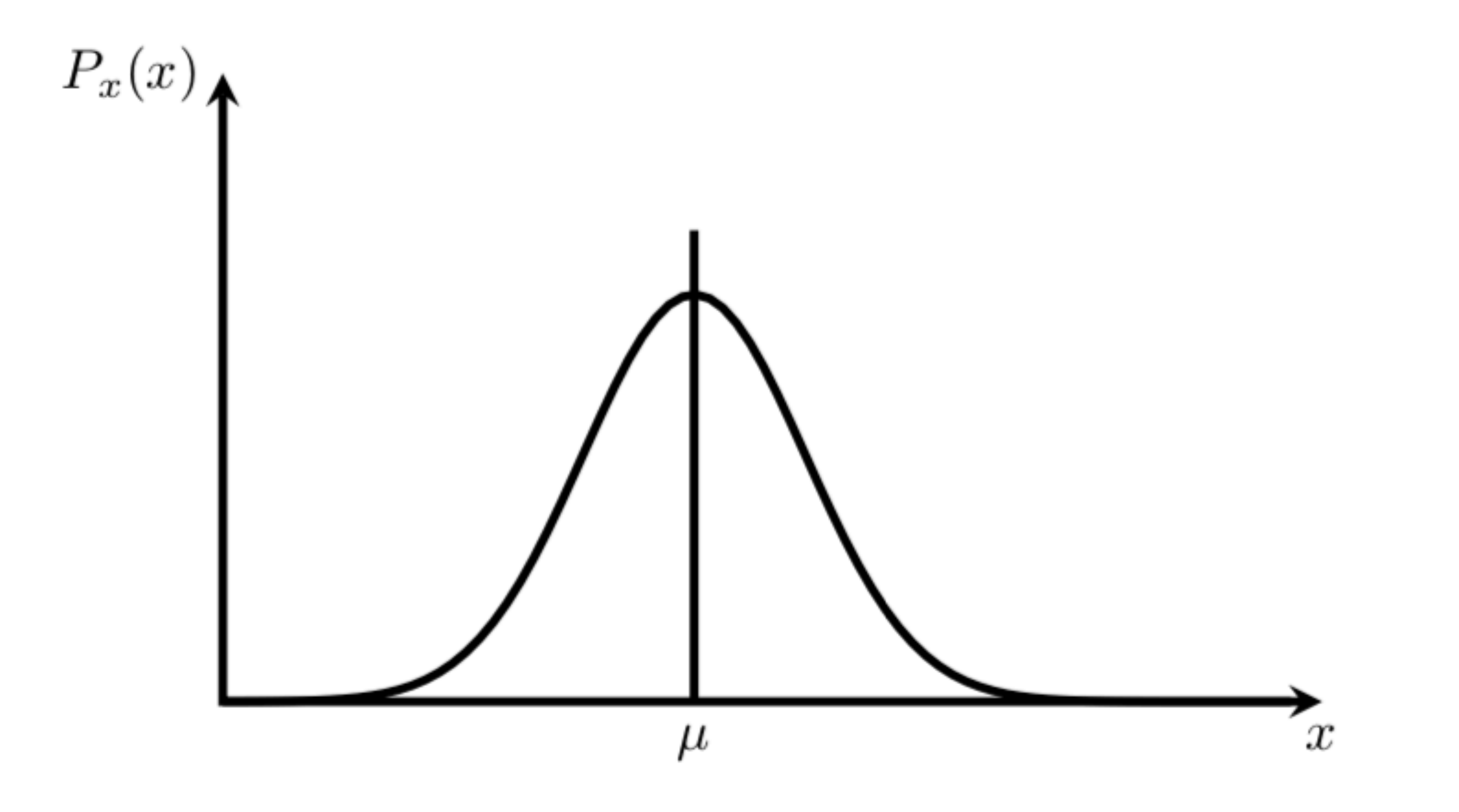
$$\omega_{i,j} = 1 + \alpha |x_{ij}|$$

$$\alpha = 10,100,1000$$

$$x_u = (Y^T Y + Y^T (C^u - I) Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$y_i = (X^T X + X^T (C^i - I) X + \lambda I)^{-1} X^T C^i p(i)$$

Distribution



$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \to min$$

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(} - \frac{(x-\mu)^2}{2\sigma^2})$$

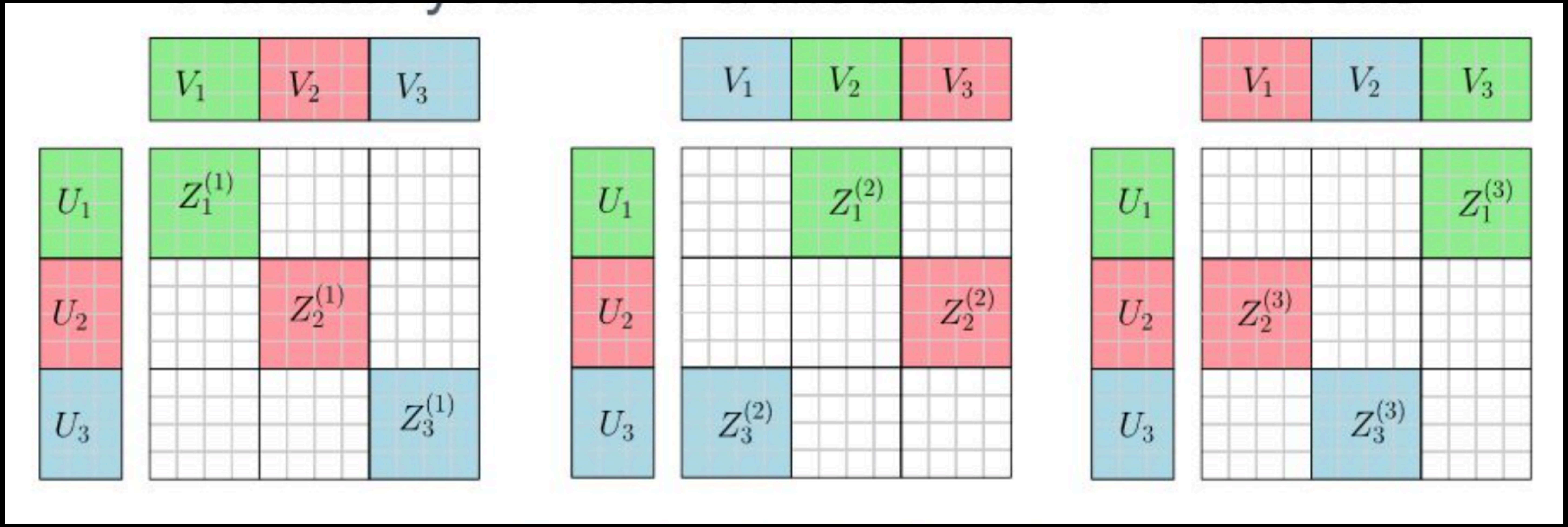
$$x_{ij} = \langle u_i, v_j \rangle + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \qquad x_{ij} \sim \mathcal{N}(\langle u_i, v_j \rangle, \sigma^2)$$

$$\prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(} \frac{(x_{i,j} - \langle u_i, v_j \rangle)^2}{2\sigma^2}) \to max$$

$$\sum_{i,j} \frac{(x_{i,j} - \langle u_i v_j \rangle)^2}{2\sigma^2} - \frac{1}{2} ln 2\pi\sigma^2 \to min$$

$$\sum_{i,j} (x_{ij} - \left\langle u_i, v_j \right\rangle)^2 \to min$$

Really BigData(DSGD)



SVD++

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + |N(u)|^{-1/2} \sum_{j \in N(u)} y_j \right).$$

$$\hat{r}_{ui} = \mu + b_u + b_i + \sum_{f=0}^{nfactors} \left(\sum_{j=0}^{nitems} r_{uj} W_{j,f} \right) W_{f,i}$$

TimeSVD++

$$\hat{r}_{ui}(t) = \mu + b_u(t) + b_i(t) + q_i^T \left(p_u(t) + |R(u)|^{-1/2} \sum_{j \in R(u)} y_j \right).$$

$$b_i(t) = b_i + b_{i,Bin(t)}$$

$$dev_u(t) = sign(t - t_u) \cdot |t - t_u|^\beta$$

$$b_i(t) = b_i + \alpha_u \cdot dev_u(t) + b_{u,t}$$

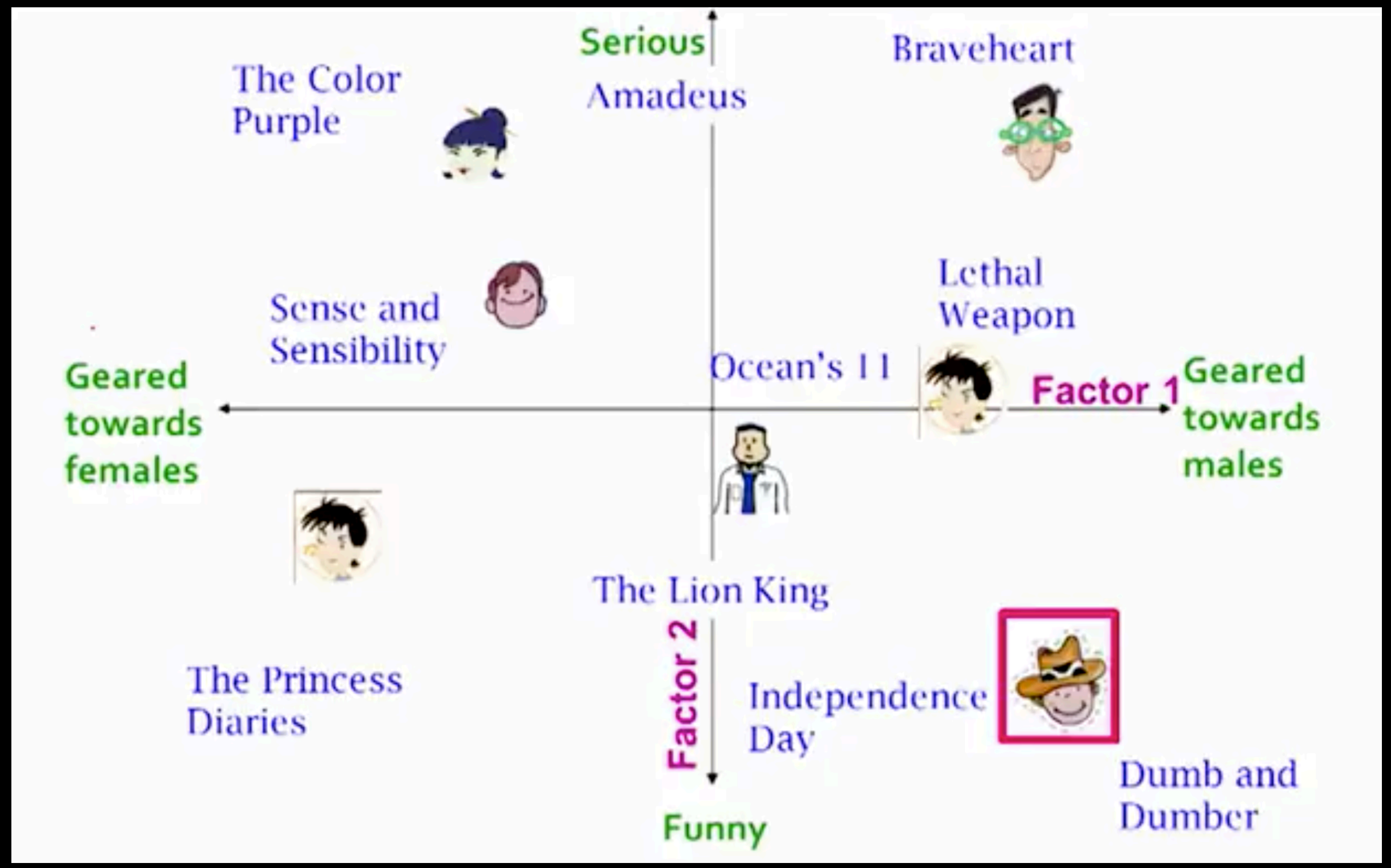
$$p_{uk}(t) = p_{uk} + \alpha_{uk} \cdot dev_u(t) + p_{uk,t}$$

Dot product to likes

- 1. We can evaluate preference on pair user-item
- 2. We can ranking all dot products to construct recommendations
- 3. We can using dot product as feature to second level models

User/Item Embeddings

- 1. We can find similar users/items
- 2. We can clustering users/items embeddings
- 3. We can find outliers



A little bit more about target

- How to boost timespent using matrix factorization?
- How to boost user's happiness using matrix factorization?
- How to boost money using matrix factorization?

Metrics

- MSE
- Precision, Recall, AUC
- NDCG
- Metrics@K

Coding

- Do it yourself
- Implicit (https://github.com/benfred/implicit)
- LightFM (https://github.com/lyst/lightfm)

- Do it yourself
- Spark MLlib (http://spark.apache.org/docs/latest/ml-guide.html)

What's next?

- 5 - 4 = 3 -2 ?
- How to compare different feedback?
- Why ALS still one of the most popular model in production?
- How improve model by the data?
- Can we add some features about user and item?

Sources

- Distributed Design and Implementation of SVD++ Algorithm for E-commerce Personalized Recommender System. Communications in Computer and Information Science. **572**. Springer Singapore. pp. 30–44
- https://github.com/benfred/implicit
- https://github.com/lyst/lightfm
- Matrix and Tensor Factorization Techniques for Recommender Systems ISBN 978-3-319-41357-0
- http://www.machinelearning.ru/wiki/images/9/95/Voron-ML-CF.pdf