Singular Value Decomposition

for recommender systems

Singular Value Decomposition

Used in LSA/LSI, PCA…

Singular Value Decomposition

Used in LSA/LSI, PCA…

$$A \quad = U \quad \Sigma \quad V^\top$$

Singular Value Decomposition

Used in LSA/LSI, PCA…

$$A \quad = \quad U \quad \Sigma \quad V^\top$$

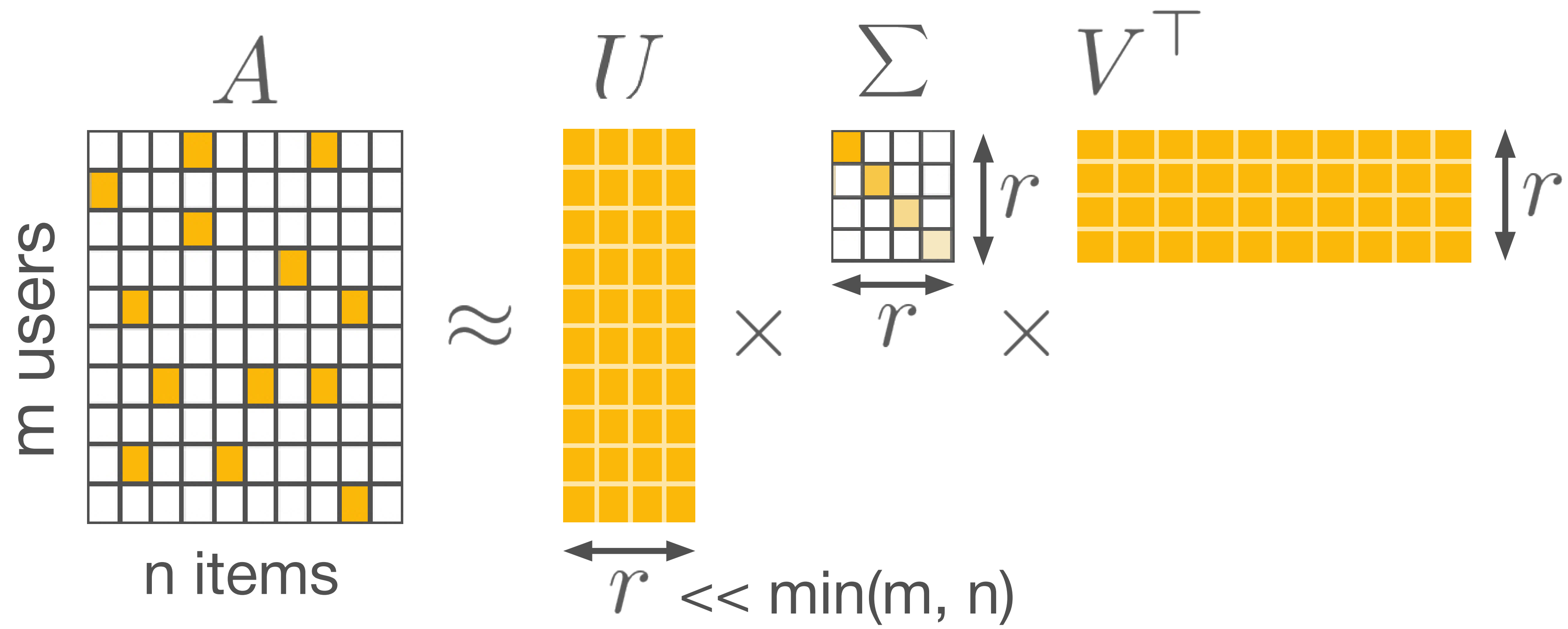
columns are orthonormal:

$$U^\top U = V^\top V = I$$

Singular Value Decomposition

Used in LSA/LSI, PCA…

Truncated SVD of rank r

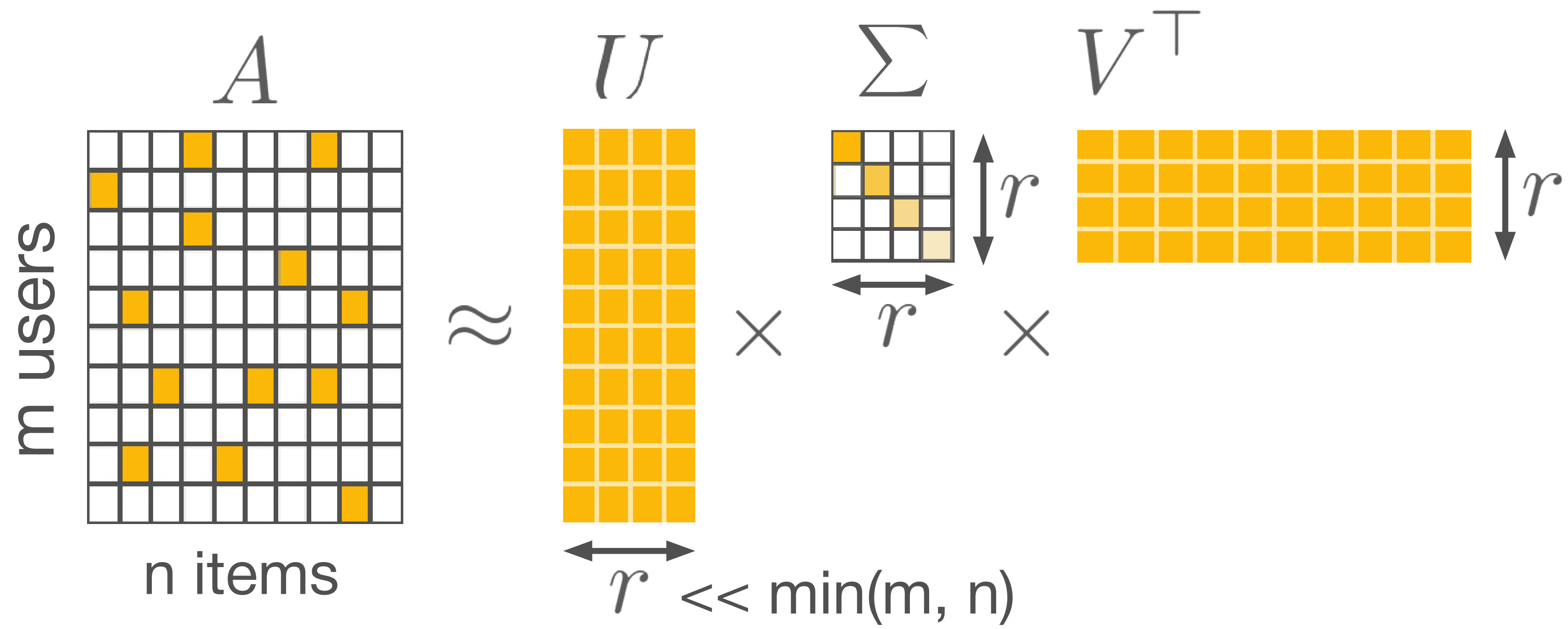


columns are orthonormal:
$$U^\top U = V^\top V = I$$

Singular Value Decomposition

Used in LSA/LSI, PCA…

Truncated SVD of rank r



columns are orthonormal:
$$U^\top U = V^\top V = I$$

$$||A - A_r||_F^2 \to \min$$

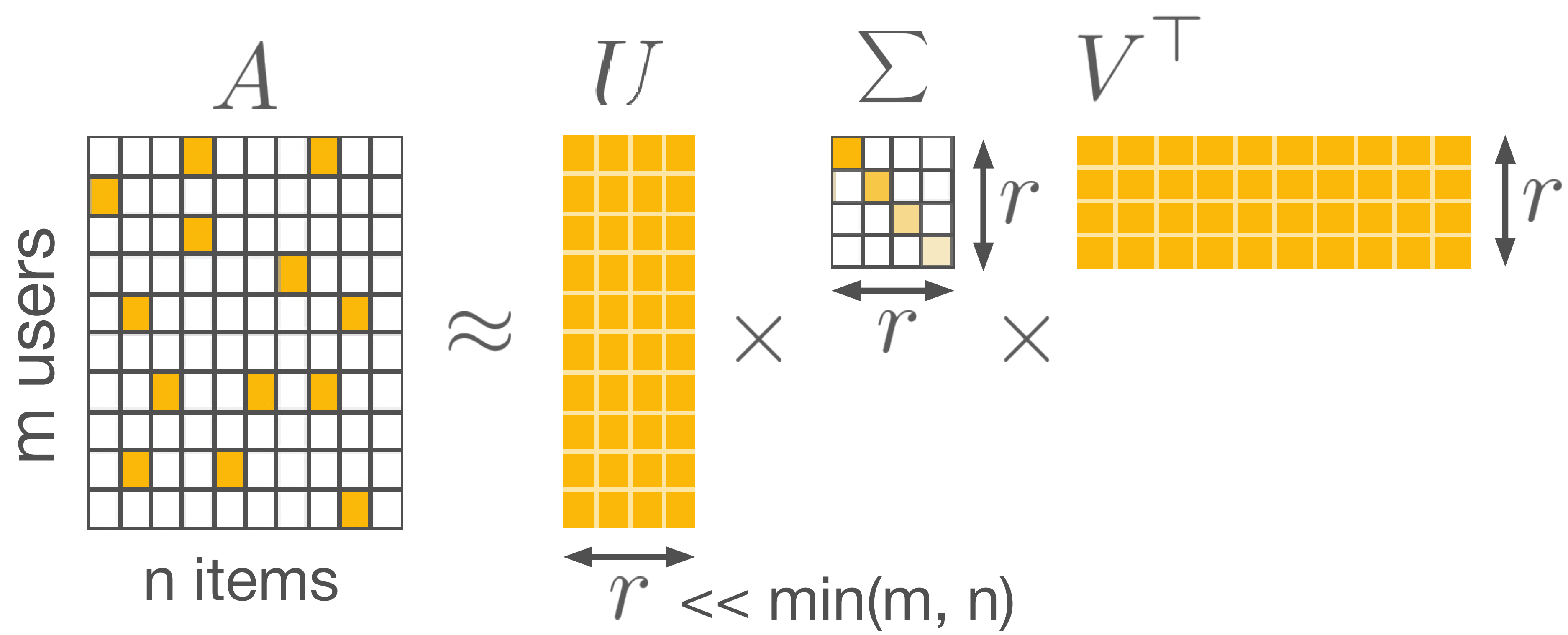
$$A_r = U \Sigma V^\top$$

$$||X||_F^2 = \sum_{ij} x_{ij}^2$$

Singular Value Decomposition

Used in LSA/LSI, PCA…

Truncated SVD of rank r



columns are orthonormal:
$U^\top U = V^\top V = I$

$$||A - A_r||_F^2 \to \min$$

$$A_r = U \Sigma V^\top$$

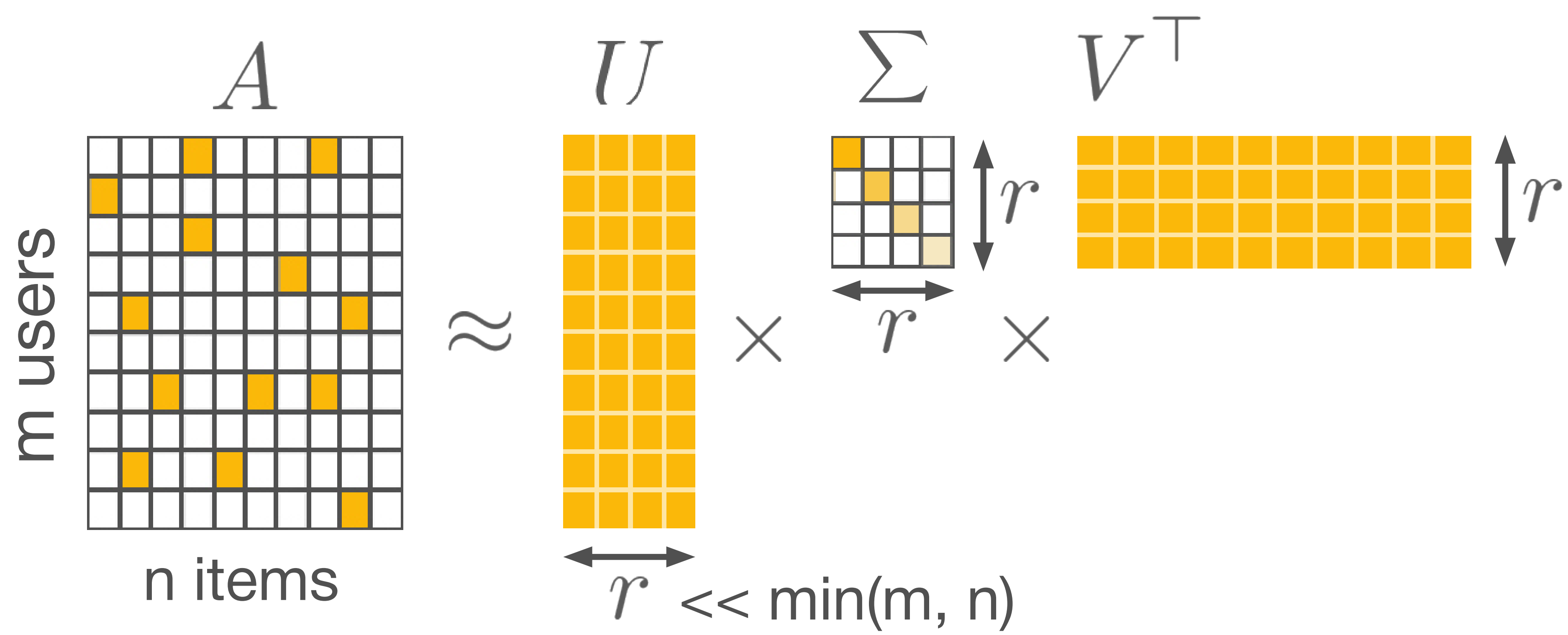
$$||X||_F^2 = \sum_{ij} x_{ij}^2$$

Undefined for incomplete matrix!

Singular Value Decomposition

Used in LSA/LSI, PCA…

Truncated SVD of rank r



columns are orthonormal:
$U^\top U = V^\top V = I$

$$||A - A_r||_F^2 \to \min$$

$$A_r = U \Sigma V^\top$$

$$||X||_F^2 = \sum_{ij} x_{ij}^2$$

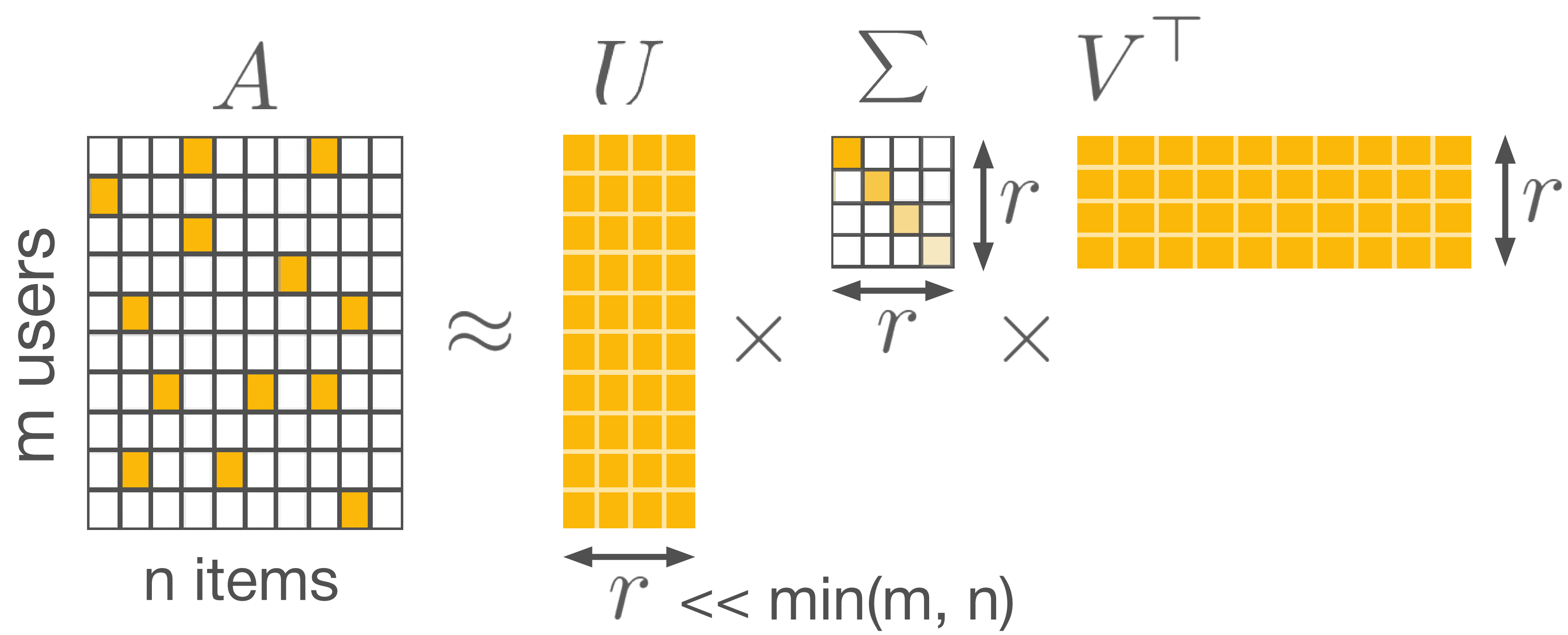
Undefined for incomplete matrix!

Let's impute zeros - PureSVD model.

Singular Value Decomposition

Used in LSA/LSI, PCA…

Truncated SVD of rank r



columns are orthonormal:
$U^\top U = V^\top V = I$

$$||A - A_r||_F^2 \to \min$$

$$A_r = U \Sigma V^\top$$

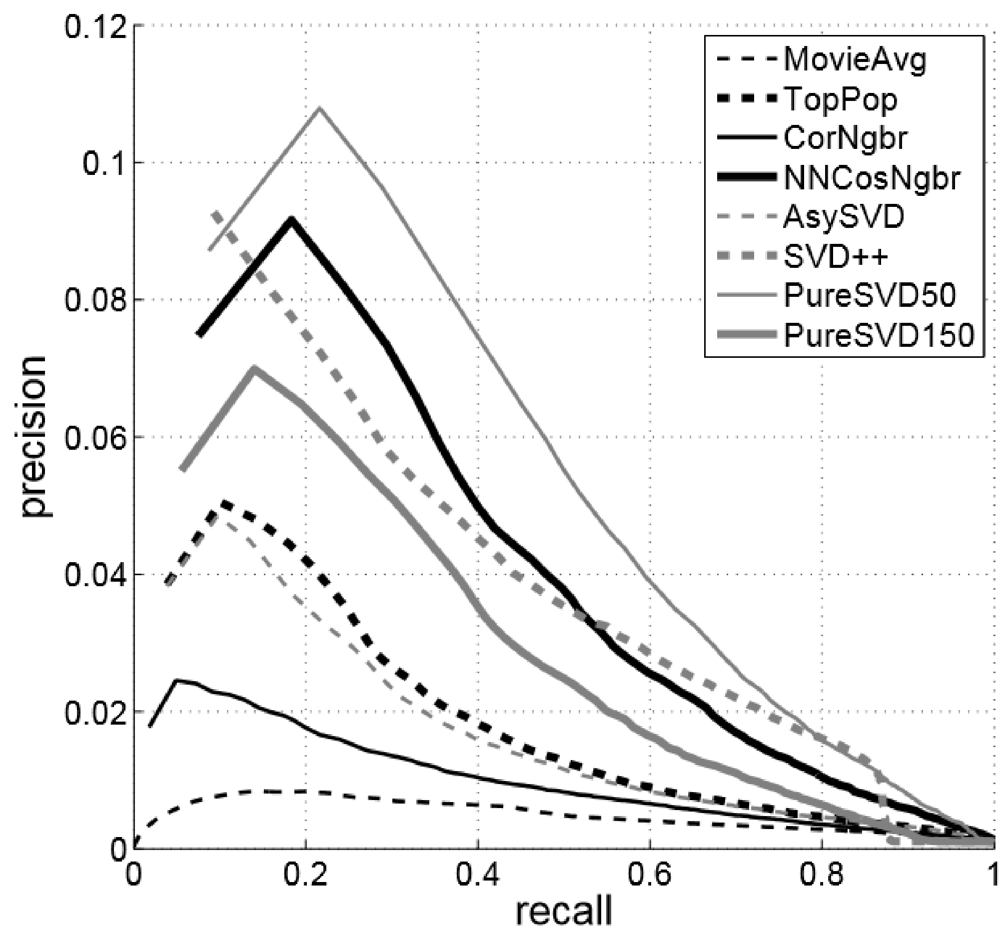
$$||X||_F^2 = \sum_{ij} x_{ij}^2$$

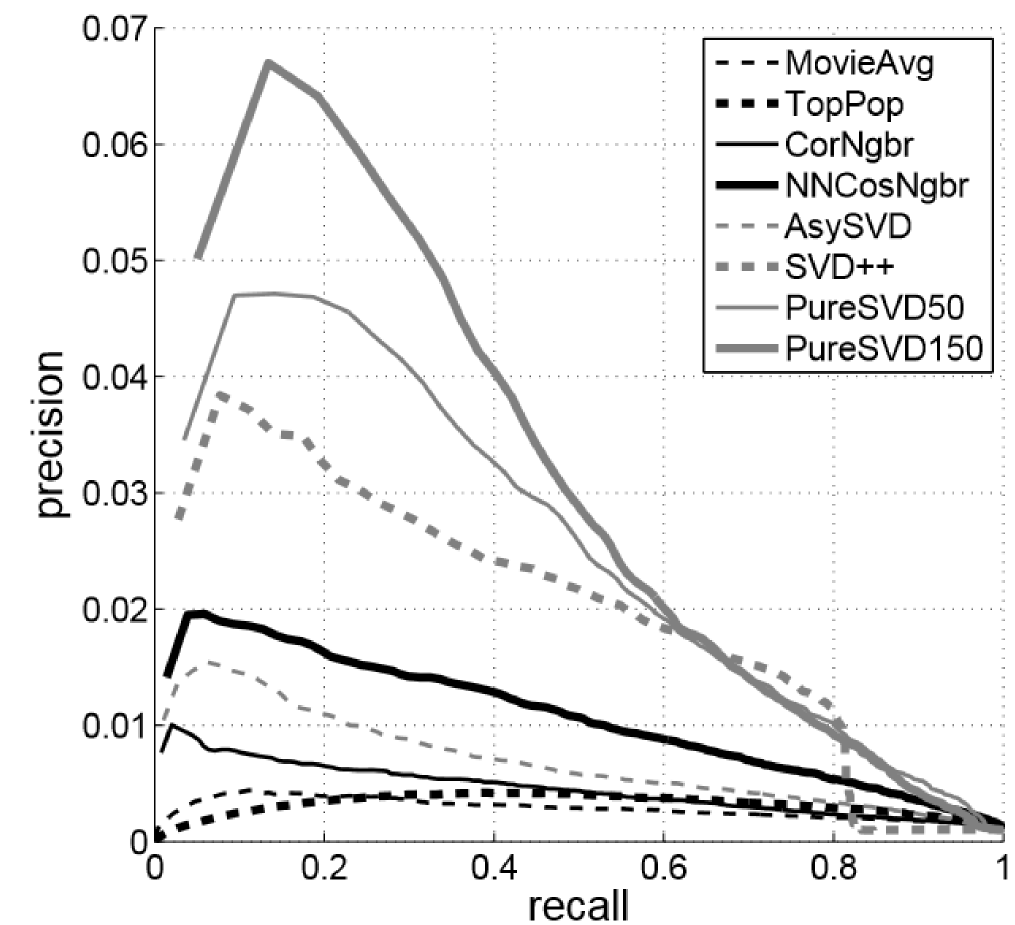
Undefined for incomplete matrix!

Let's impute zeros - PureSVD model.

- values are highly biased towards 0
- not good for rating prediction
- its not a big problem for ranking task

PureSVD – quality of recommendations



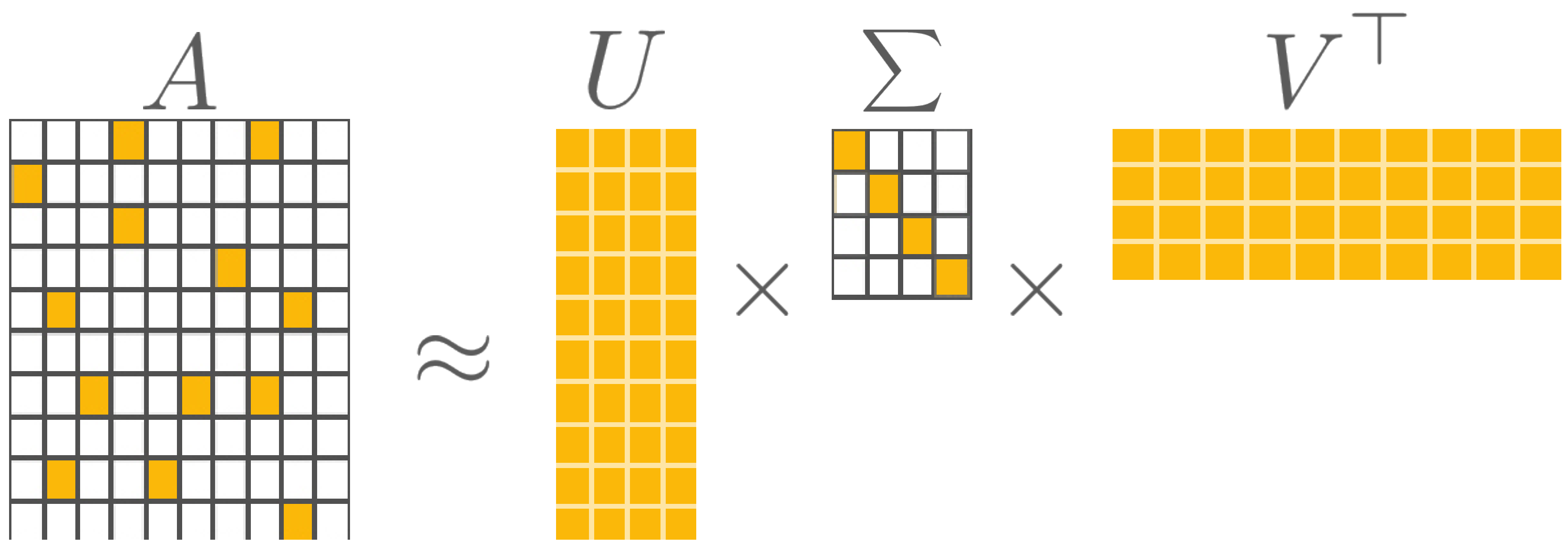


Netflix data: complete dataset (left) and “long-tail” (right).

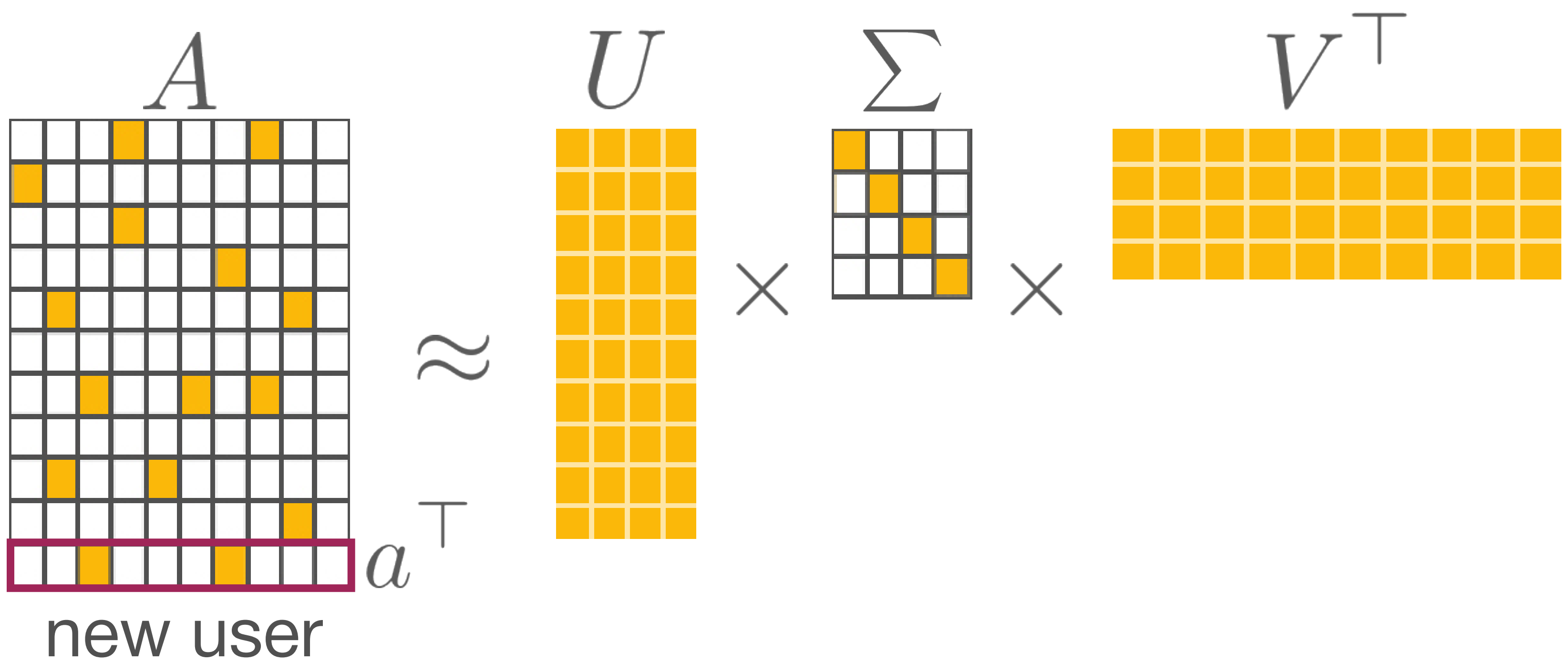
P. Cremonesi, Y.Koren, R.Turrin, “Performance of Recommender Algorithms on Top-N Recommendation Tasks“,Proceedings of the 4th ACM conference on Recommender systems, 2011.

Note: Funk SVD, SVD++, TimeSVD++, Asymmetric SVD … **are not** the SVD!

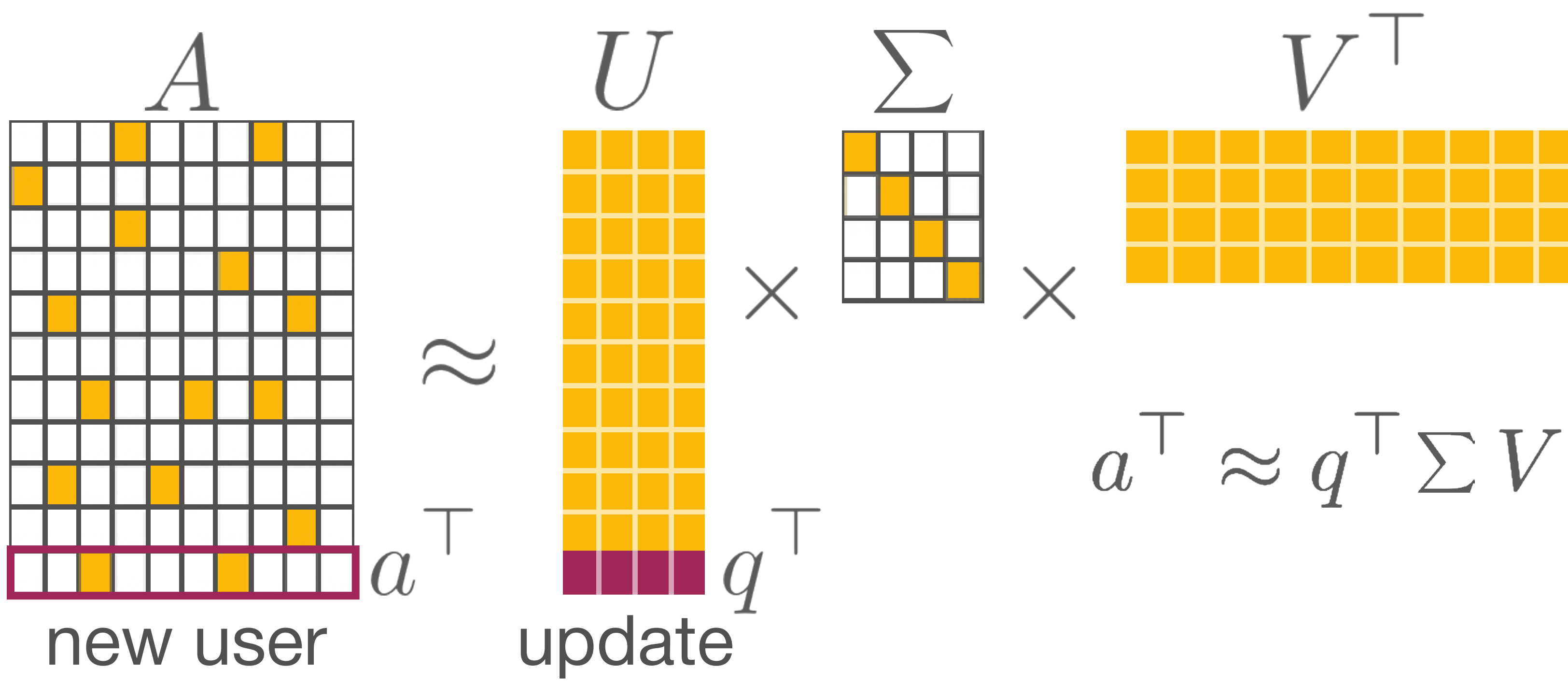
PureSVD – recommending online



PureSVD – recommending online



PureSVD – recommending online

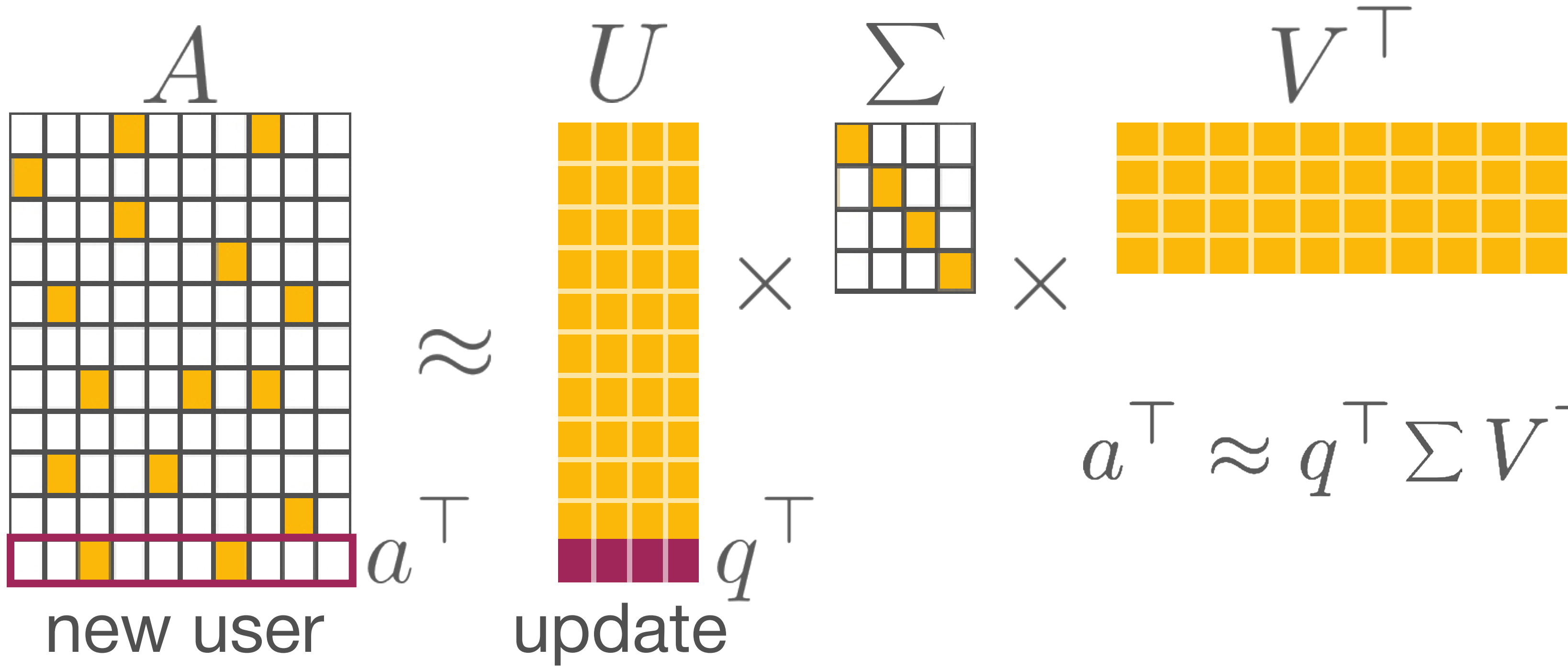


$V^\top V = I$ orthogonality of columns

$$a^\top \approx q^\top \Sigma V^\top \longrightarrow \boldsymbol{a}^\top \approx \boldsymbol{q}^\top \Sigma V^\top V = \boldsymbol{q}^\top \Sigma$$

$$\boxed{\boldsymbol{q}^\top \approx \boldsymbol{q}^\top V \Sigma^{-1}}$$ new user embedding

PureSVD – recommending online



$$V^\top V = I$$ orthogonality of columns

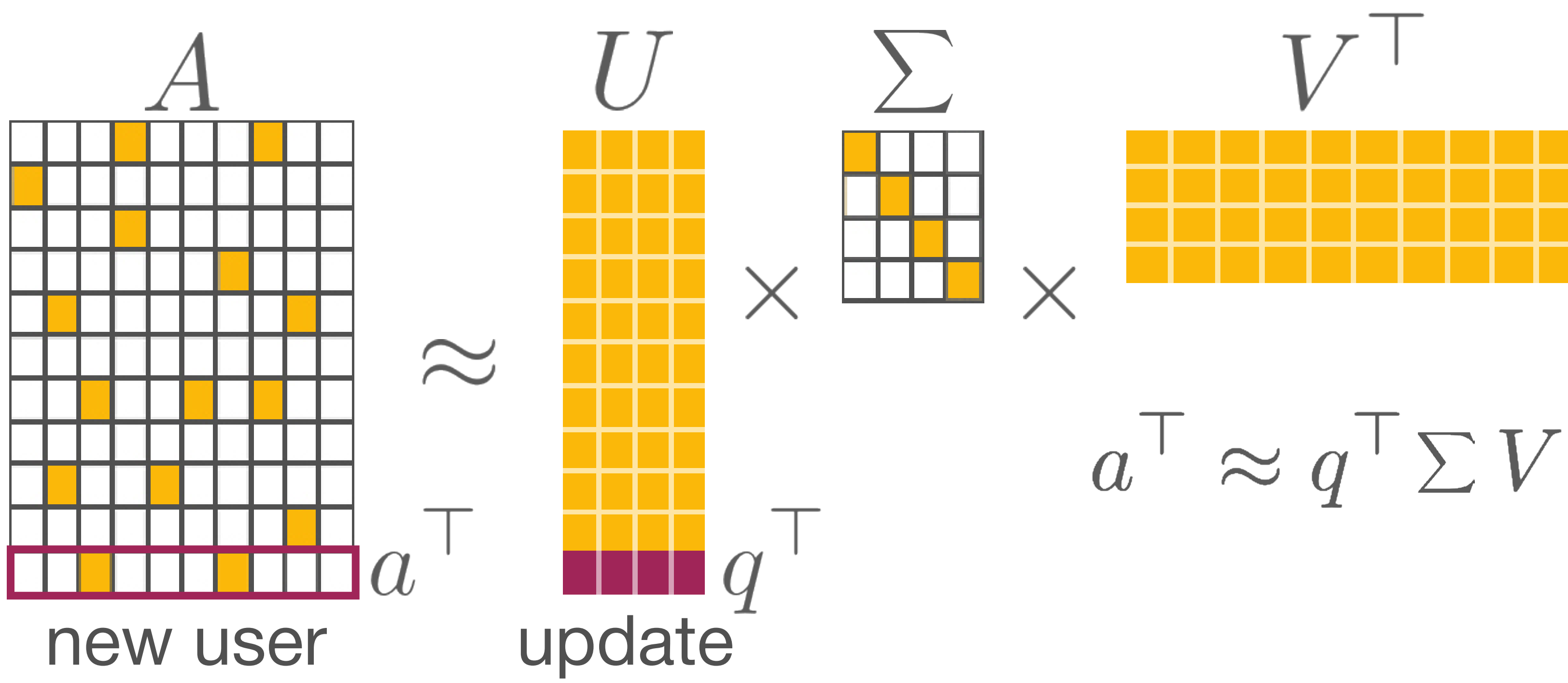
$$a^\top \approx q^\top \Sigma V^\top \rightarrow \boldsymbol{a}^\top \approx \boldsymbol{q}^\top \Sigma V^\top V = \boldsymbol{q}^\top \Sigma$$

*folding-in technique**

$$\boxed{\boldsymbol{q}^\top \approx \boldsymbol{q}^\top V \Sigma^{-1}}$$ new user embedding

*G. Furnas, S. Deerwester, and S. Dumais, "Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure," Proceedings of ACM SIGIR Conference, 1988

PureSVD – recommending online



$$V^\top V = I$$

orthogonality of columns

$$a^\top \approx q^\top \Sigma V^\top \longrightarrow \boldsymbol{a}^\top \approx \boldsymbol{q}^\top \Sigma V^\top V = \boldsymbol{q}^\top \Sigma$$

*folding-in technique**

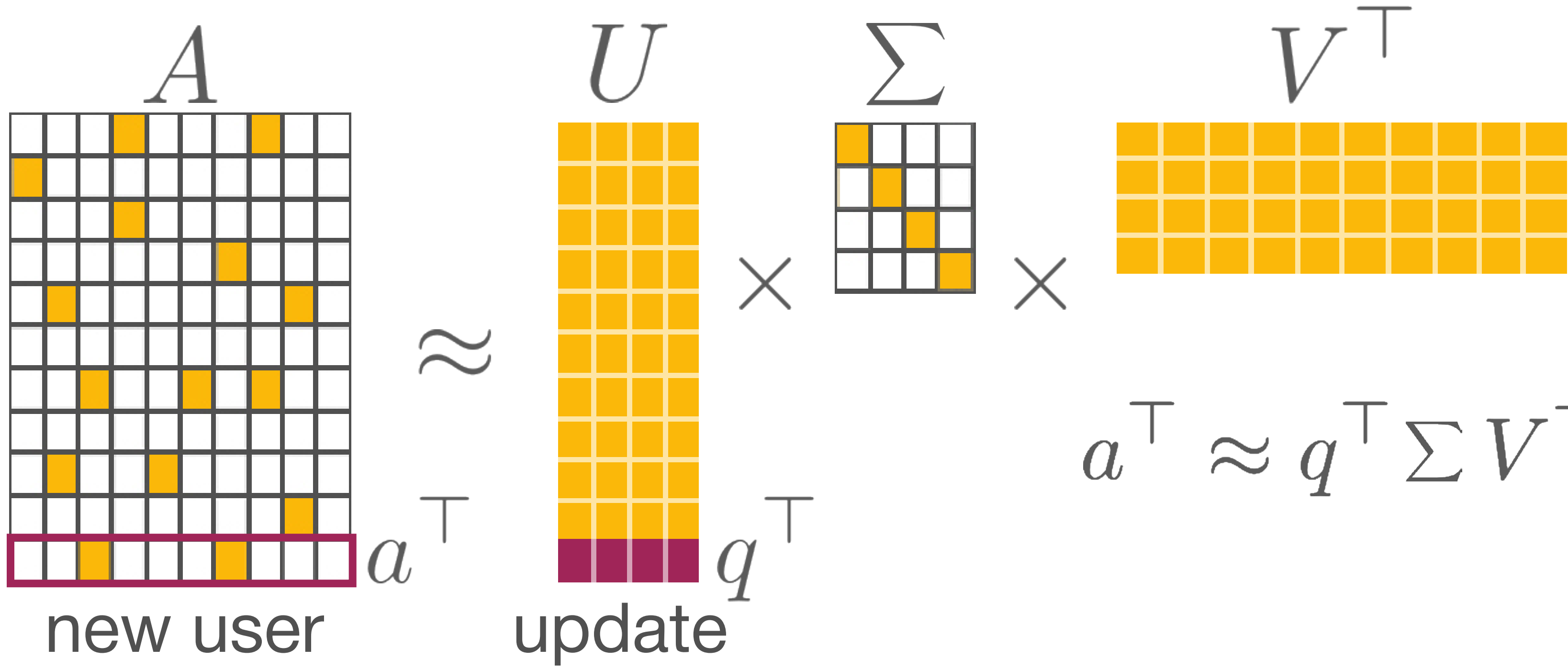
$$\boxed{\boldsymbol{q}^\top \approx \boldsymbol{q}^\top V \Sigma^{-1}}$$

new user embedding

$$r^\top = q^\top \Sigma V^\top \approx a^\top V \Sigma^{-1} V^\top = a^\top V V^\top$$

*G. Furnas, S. Deerwester, and S. Dumais, “Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure,” Proceedings of ACM SIGIR Conference, 1988

PureSVD – recommending online



$V^\top V = I$ orthogonality of columns

$$a^\top \approx q^\top \Sigma V^\top \rightarrow \boldsymbol{a}^\top \approx \boldsymbol{q}^\top \Sigma V^\top V = \boldsymbol{q}^\top \Sigma$$

*folding-in technique**

$$\boxed{\boldsymbol{q}^\top \approx \boldsymbol{q}^\top V \Sigma^{-1}}$$

new user embedding

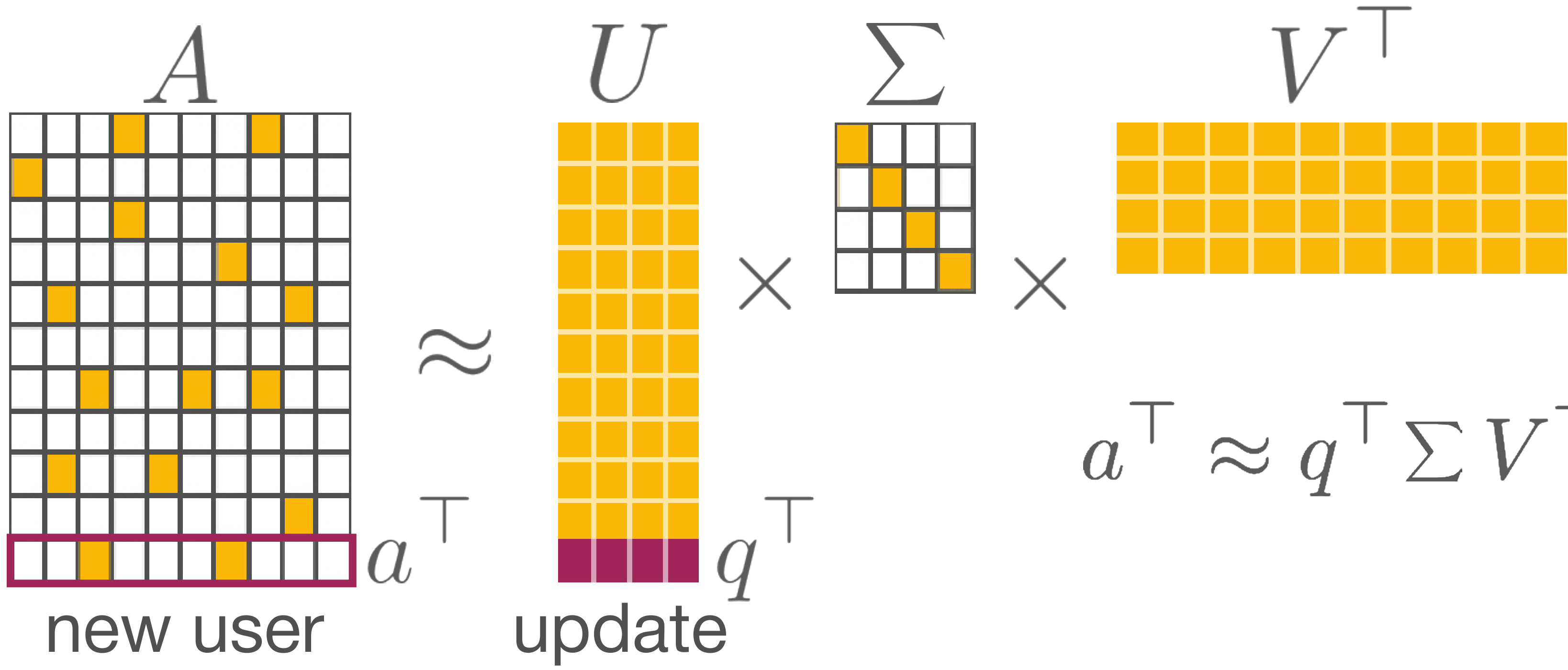
$$r^\top = q^\top \Sigma V^\top \approx a^\top V \Sigma^{-1} V^\top = a^\top V V^\top$$

vector of predicted item scores

$$\boldsymbol{r} \approx V V^\top \boldsymbol{a}$$

*G. Furnas, S. Deerwester, and S. Dumais, “Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure,” Proceedings of ACM SIGIR Conference, 1988

PureSVD – recommending online



$V^\top V = I$ orthogonality of columns

$$a^\top \approx q^\top \Sigma V^\top \rightarrow \boldsymbol{a}^\top V \approx \boldsymbol{q}^\top \Sigma V^\top V = \boldsymbol{q}^\top \Sigma$$

*folding-in technique**

$$\boxed{\boldsymbol{q}^\top \approx \boldsymbol{a}^\top V \Sigma^{-1}}$$

new user embedding

$$r^\top = q^\top \Sigma V^\top \approx a^\top V \Sigma^{-1} V^\top = a^\top V V^\top$$

SIGMA

allows for real-time recommendations
O(nr) complexity

vector of predicted item scores

$$\boldsymbol{r} \approx V V^\top \boldsymbol{a}$$

*G. Furnas, S. Deerwester, and S. Dumais, "Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure," Proceedings of ACM SIGIR Conference, 1988

Summary

Summary

- You know what SVD is and what its key features are

Summary

- You know what SVD is and what its key features are
- You understand the role it plays in the **PureSVD** model

Summary

- You know what SVD is and what its key features are
- You understand the role it plays in the **PureSVD** model
- You can explain how **folding-in** approach works and how to use it for online recommendations