6.867 Machine Learning Fall 2017

Lecture 22. Non-negative Matrix Factorization

http://stellar.mit.edu/S/course/6/fa17/6.867/

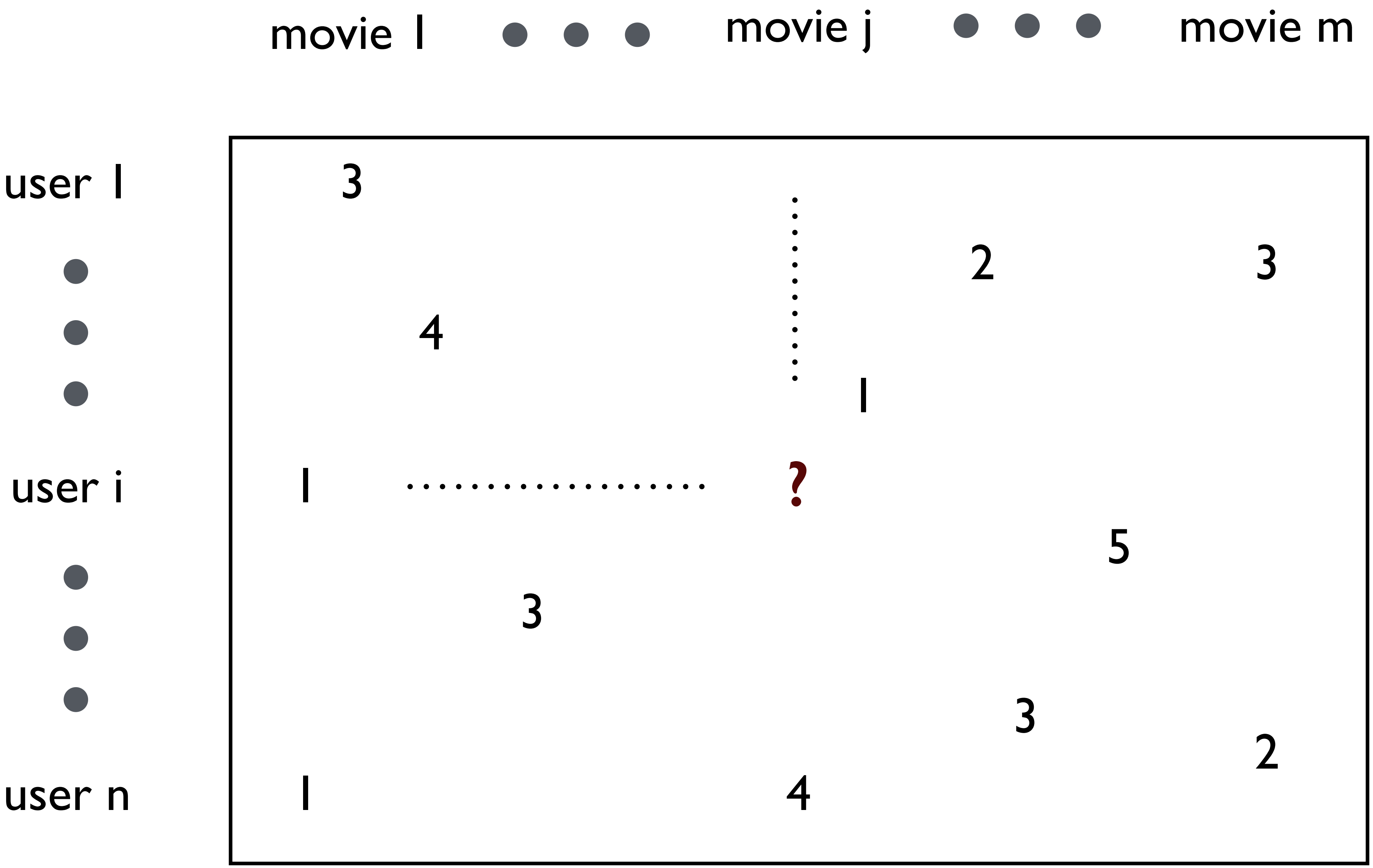
Announcements

- Quiz 2 Grades Posted
  - Re-grade request
    - Submit by Wednesday, December 6, 5pm ET
    - Without *Reason for re-grade* we will ignore it
    - Re-grade will lead to re-grade of entire question
- Projects:
  - Final project report due on December 12
  - Please reach out to TAs / Instructors if you need any help
- Exercise 12 will be posted soon: covers today's lecture/ Reinforcement learning
- Class evaluation: let us use *few* minutes *now*

Outline

- Non-negative Matrix Factorization
  - Formulation
  - Application
    - Topic Model
  - Algorithms
    - Alternative Least Squares
    - Using "Anchor Words"

Recall: Recommendation Systems



Rating Matrix A

Recall: Recommendation Systems

- Ground Truth Matrix

$$A = [A_{ij}] \in \mathbb{R}^{m \times n}$$

- Observation Matrix $Y = [Y_{ij}] \in \mathbb{R}^{m \times n}$

$$Y_{ij} = \begin{cases} \text{noisy } A_{ij} & \text{if observed} \\ \star & \text{otherwise} \end{cases}$$

- Goal: produce estimate $\hat{A} = [\hat{A}_{ij}]$ so that prediction error is small

$$\frac{1}{mn} \sum_{ij} \left(\hat{A}_{ij} - A_{ij} \right)^2$$

A Solution: Low-Rank Matrix Factorization

- Matrix Factorization: find low-rank matrices

$$U \in \mathbb{R}^{m \times r}, \ V \in \mathbb{R}^{n \times r}$$

- So that

$$Y \approx UV^T$$

- Optimization view:

$$\begin{aligned} \text{minimize} \quad & \sum_{(i,j) \in \mathcal{O}} \left(Y_{ij} - U_{i\cdot}^T V_{j\cdot} \right)^2 \\ \text{over} \quad & U \in \mathbb{R}^{m \times r} \\ & V \in \mathbb{R}^{n \times r} \end{aligned}$$

A Solution: Low-Rank Matrix Factorization

$$\underbrace{\begin{array}{c} \overset{n}{\longleftrightarrow} \\ m \updownarrow \begin{bmatrix} \quad Y \quad \end{bmatrix} \end{array}}_{} \approx \begin{array}{c} \overset{r}{\longleftrightarrow} \\ m \updownarrow \begin{bmatrix} U \end{bmatrix} \\ \geq 0 \end{array} \times \begin{array}{c} \overset{n}{\longleftrightarrow} \\ \begin{bmatrix} \quad V^T \quad \end{bmatrix} \updownarrow r \\ \geq 0 \end{array}$$

minimize $\|Y - UV^T\|_F$

over $U \in \mathbb{R}_{\geq 0}^{m \times r}$

$V \in \mathbb{R}_{\geq 0}^{n \times r}$ $\qquad \Longleftarrow \qquad \Uparrow$

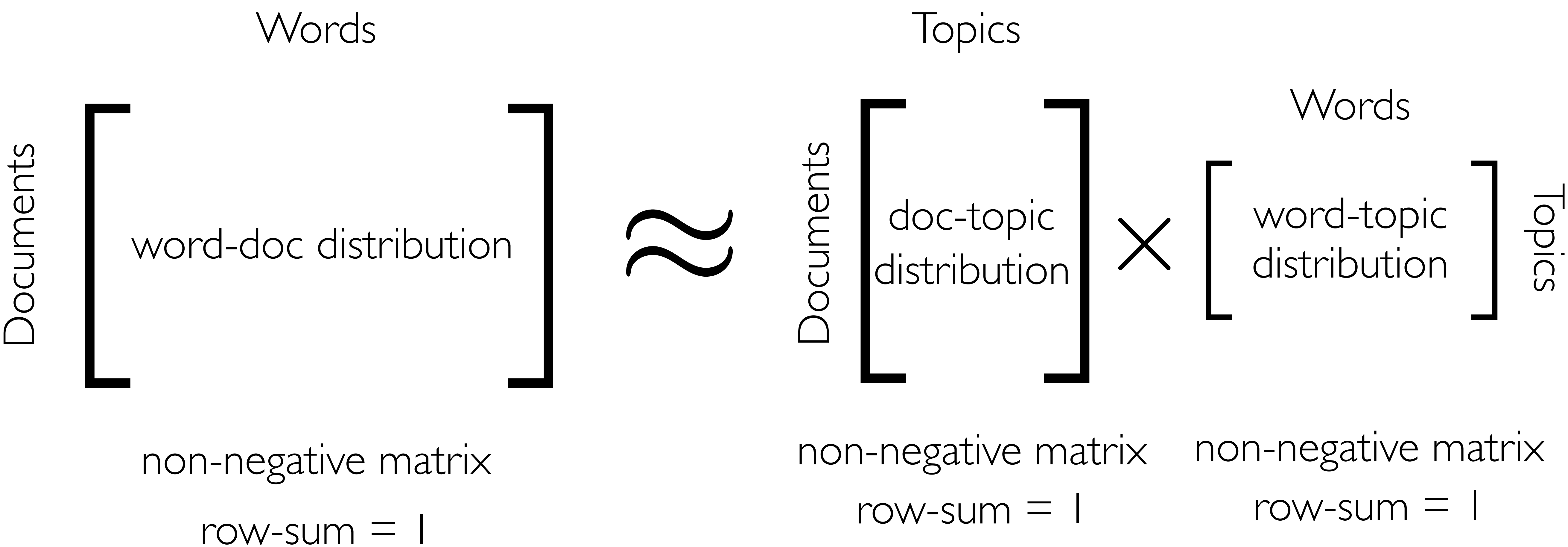
Non-negative ratings

Non-negative Matrix Factorization

- Find rank r matrix approximation of matrix ***Y***
    - with non-negative factors

$$
\begin{aligned}
\text{minimize } & \|Y - UV^T\|_F \\
\text{over } & U \in \mathbb{R}^{m \times r}_{\geq 0} \\
& V \in \mathbb{R}^{n \times r}_{\geq 0}
\end{aligned}
$$

Recall: Topic Model



Alternative Least Squares (ALS)

$$\text{minimize } \|Y - UV^T\|_F = \sum_{ij} (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } U_1, \ldots, U_m \in \mathbb{R}^r_{\geq 0}$$

$$V_1, \ldots, V_n \in \mathbb{R}^r_{\geq 0}$$

This is a HARD problem

Alternative Least Squares (ALS)

$$\text{minimize } \|Y - UV^T\|_F = \sum_{ij} (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } U_1, \ldots, U_m \in \mathbb{R}^r_{\geq 0}$$

$$V_1, \ldots, V_n \in \mathbb{R}^r_{\geq 0}$$

$$\text{minimize} \sum_j (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } U_i \in \mathbb{R}^r_{\geq 0}$$

EASY problem

$$\text{minimize} \sum_i (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } V_j \in \mathbb{R}^r_{\geq 0}$$

EASY problem

Alternative Least Squares (ALS)

$$\text{minimize } \|Y - UV^T\|_F = \sum_{ij} (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } U_1, \ldots, U_m \in \mathbb{R}^r_{\geq 0}$$

$$V_1, \ldots, V_n \in \mathbb{R}^r_{\geq 0}$$

$$\text{minimize} \sum_j (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } U_i \in \mathbb{R}^r_{\geq 0}$$

convex constraints

EASY problem

$$\text{minimize} \sum_i (Y_{ij} - U_i^T V_j)^2$$

$$\text{over } V_j \in \mathbb{R}^r_{\geq 0}$$

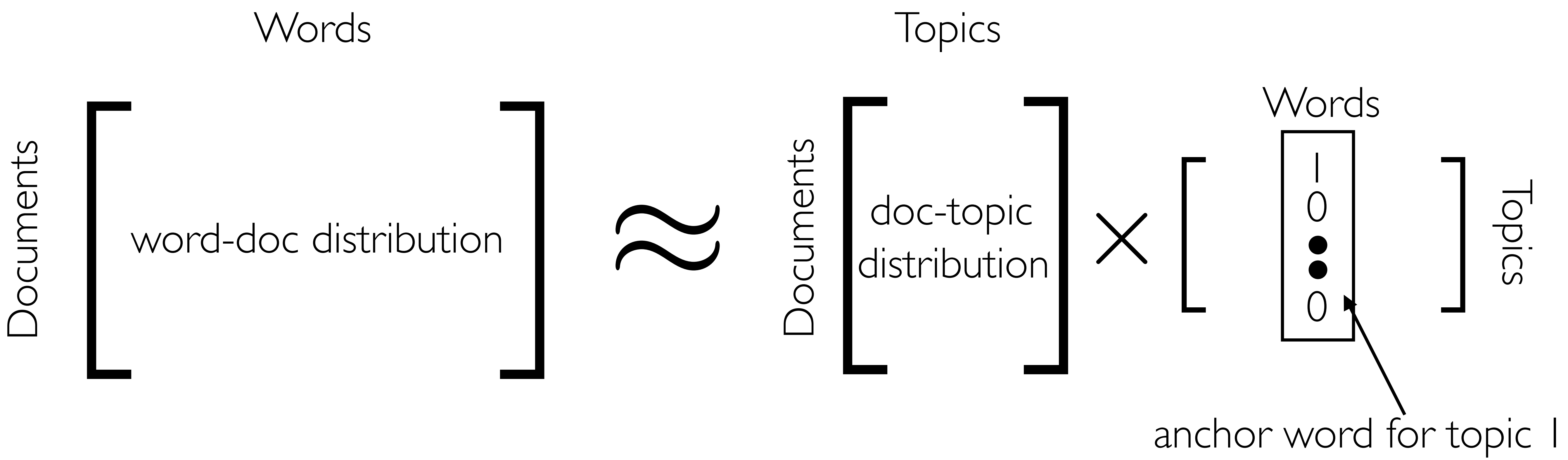
convex constraints

EASY problem

Alternative Least Squares (ALS)

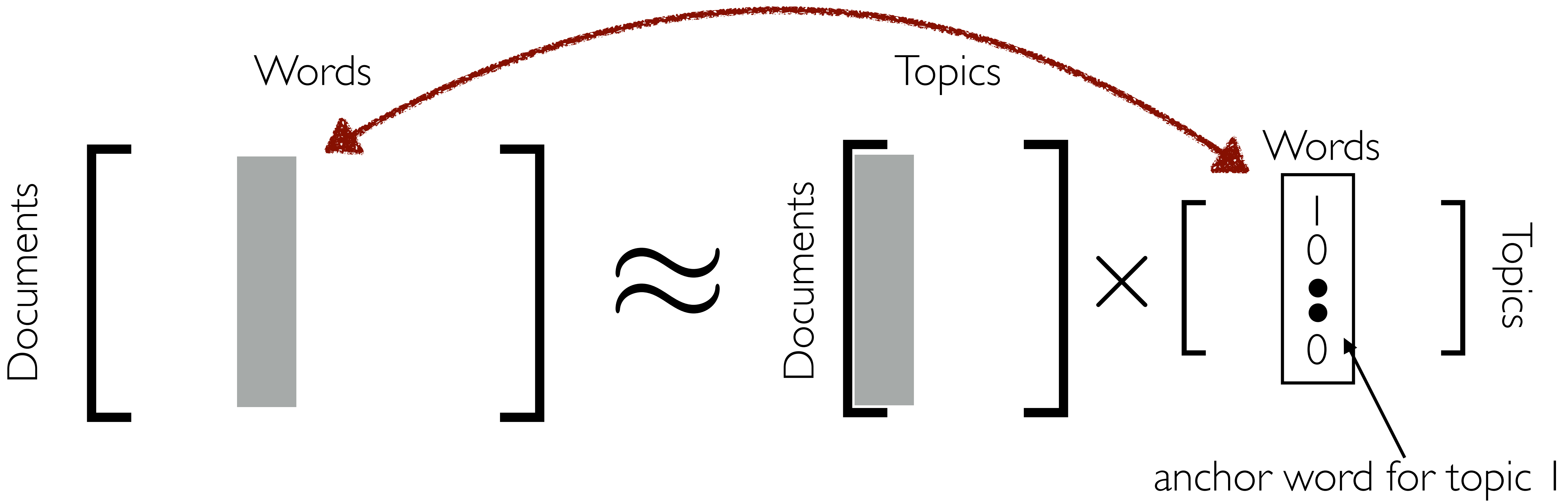
- In summary
  - Initialize $U^0 \in \mathbb{R}_{\geq 0}^{m \times r}, \; V^0 \in \mathbb{R}_{\geq 0}^{m \times r}$ appropriately
  - Iteratively:
    - set U^{t+1} assuming V^t fixed
      - Solving *m constrained, non-negative* least squares problems
    - set V^{t+1} assuming U^{t+1} fixed
      - Solving *n constrained, non-negative* least squares problems
  - Stop upon “convergence”

Algorithm using Anchor Words



- Anchor word for a topic
  - The particular word occurs *only* in that topic
  - For example: *401k* in *financial* topic, *googly* in *cricket*, etc.
  - Algebraically: column in word by topic matrix
    - has exactly on non-zero entry equal to 1
    - the row of entry 1 equals to the topic for which word is anchor

Algorithm using Anchor Words



- Let each topic have at least *one* anchor word
- Then each column of document by topic matrix (corresponding to topic)
    - is present in the document by word matrix in position
        - corresponding to the anchor word for that topic
- All other columns in document by topic matrix
    - are convex combination of the above “anchor” columns

Algorithm using Anchor Words

- Recall: we observe $Y \in \mathbb{R}_{\geq 0}^{m \times n}$ (documents by words matrix)
    - Let its columns be $Y_{\cdot 1}, \ldots, Y_{\cdot n} \in \mathbb{R}_{\geq 0}^{m}$
    - Their convex hull has $r(< n)$ extreme points
    - We need to find them
        - they will give documents by topic matrix!
        - and word by document matrix can be found using half step of ALS!

- Next, let's find the extreme points of convex hull of *n* points

Algorithm using Anchor Words

- Given n points $y_1, \ldots, y_n \in \mathbb{R}^m$
  - find *extreme* points of their convex hull

$$\mathsf{CH}(y_1, \ldots, y_n) = \Big\{ \sum_{i=1}^{n} \alpha_i y_i, \quad \alpha_i \geq 0,\ 1 \leq i \leq n, \\ \sum_i \alpha_i = 1 \Big\}$$

- Fact: a solution of linear program is achieved at extreme point
- Algorithm: do the following *many* times and collect solution set
  - sample $w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  - find solution of $\quad$ **minimize** $w^T y$ **over** $y \in \mathsf{CH}(y_1, \ldots, y_n)$