# Matrix Completion

# Objectives

- define the matrix completion problem
- approach missing data using low-rank models
- introduce iterative singular value thresholding

# Use "patterns" to fill in missing entries 

Ratings matrix

$\underline{X} \in \mathbb{R}^{N \times M}$

$$\begin{bmatrix} 5 & 4 & 9 & 1 & \times \\ 9 & 6 & \times & \times & 7 \\ \times & 10 & \times & 2 & 4 \\ 3 & 7 & 3 & \times & \times \\ 8 & \times & \times & 6 & 2 \end{bmatrix}$$

N movies ↓

M users →

Can we predict the missing entries?

Model: assume $\underline{X}$ is well approximated with a small number of patterns

$$\underline{X} \approx \sum_{i=1}^{r} \underline{t}_i \underline{s}_i^T = \underline{T}\,\underline{S}$$

genres, actors, director...

hobbies, age, address...

Matrix completion: use known data to find patterns and predict missing entries

$\Omega = \{ i,j : \underline{X}_{ij} \text{ given} \}$ indices of known values

1) Rank minimization $\underline{X} = \underset{\underline{M}}{\operatorname{argmin}} \operatorname{rank}(\underline{M})$ s.t. $\underline{M}_{ij} = \underline{X}_{ij}$ $\forall i,j \in \Omega$

minimum number of patterns matching given values

Intractable !

$\operatorname{rank} \underline{M} = \#\{ \ell : \sigma_\ell > 0 \}$

2) Nuclear norm minimization

$\underline{X} = \underset{\underline{M}}{\operatorname{argmin}} \|\underline{M}\|_* \text{ s.t. } \underline{M}_{ij} = \underline{X}_{ij} \ \forall i,j \in \Omega$

Nuclear/trace norm

$\|M\|_* = \sum_\ell \sigma_\ell$

Computationally tractable

# Iterative Singular Value Thresholding is one possible algorithm



## Initialize

$\underline{M}^{(0)} = \underline{0}$

Set threshold or $r$

## Iterate

for $k = 1, 2, 3, \ldots$

$\underline{M}^{(k)} = \underline{M}^{(k-1)}$

$\underline{M}^{(k)}_{\Omega} = \underline{X}_{\Omega}$ (fill in known values)

$[\underline{U}, \underline{\Sigma}, \underline{V}] = \text{svd}(\underline{M}^{(k)})$

$$\hat{\Sigma}_{ii} = \Sigma_{ii} \cdot \begin{cases} 1 & \Sigma_{ii} > \text{threshold} \\ 0 & \Sigma_{ii} \le \text{threshold} \end{cases}$$

— or —

$$\hat{\Sigma}_{ii} = \begin{cases} \Sigma_{ii}, & i \le r \\ 0, & i \ge r+1 \end{cases}$$

$\underline{M}^{(k)} = \underline{U}\,\hat{\underline{\Sigma}}\,\underline{V}^T$

if $\|\underline{M}^{(k)} - \underline{M}^{(k-1)}\|_F < \varepsilon$

stop

else

next $k$

# Matrix completion is an open problem



- choosing $r$ or threshold in ISVT
- multiple algorithms:
  - convergence
  - complexity
  - noise
- results depend on distribution of missing entries
- applications include missing pixels in images, position from partial distance info, ...