# Lecture 25: 4/19/17

## Background: Low-Rank Factorization

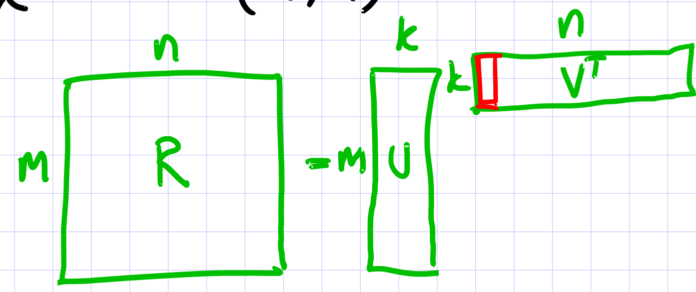
### Case #1

**Given:** $m \times n$ matrix $R$ of rank $k \ll \min(m,n)$

**Claim:** $\exists$ an $m \times k$ matrix $U$

an $n \times k$ matrix $V$

s.t. $R = UV^T$




**Note:** ① columns of $U$ are $k$ basis vectors of cols of $R$

② rows of $V^T$ are $k$ basis vectors of rows of $R$

### Case #2

**Given:** $m \times n$ matrix $R$ of rank $\ell > k$

$k$ = # cols in $U, V$

**Claim:** Can approximate $R$ with rank-$k$ matrices $U, V$

$$R \approx UV^T$$

### Approx. Error

**Def:** residual matrix $E = R - UV^T$

**MSE:** $\|E\|_2^2 = \|R - UV^T\|_2^2$

where $\|E\|_2^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} E_{ij}^2$ is the Frobenius norm of $E$.

# Unconstrained MF

## Opt. Problem #1 (fully observed R)

$$\hat{U},\hat{V} = \underset{U,V}{\operatorname{argmin}}\ J(U,V)$$

where $J(U,V) \triangleq \frac{1}{2}\|R - UV^T\|_2^2$

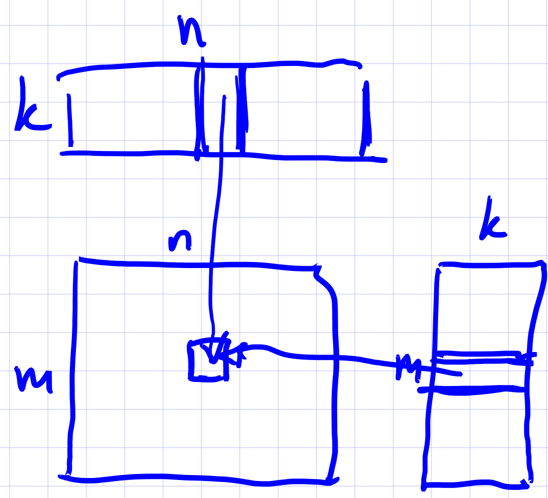
## Opt Problem #2: (partially observed R)

Let $r_{ij} \triangleq R_{ij}$ ← rating of item j by user i

$\vec{u}_i \triangleq U_i$ ← user factor

$\vec{v}_j \triangleq V_j$ ← item factor

Let $Z = \{(i,j) : r_{ij} \text{ is observed}\}$




$$J(U,V) \triangleq \frac{1}{2}\sum_{(i,j)\in Z}(r_{ij} - \vec{u}_i^{\,T}\vec{v}_j)^2$$

$$= \sum_{(i,j)\in Z} J_{ij}(U,V)$$

## Model: (Predictions)

$$\hat{r}_{ij} = \vec{u}_i^{\,T}\vec{v}_j = \sum_{l=1}^{k} u_{il}\, v_{jl}$$

## Gradient Descent:

while not converged:

$$U^{(t+1)} \leftarrow U^{(t)} - \gamma \nabla_U J(U^{(t)},V^{(t)})$$

$$V^{(t+1)} \leftarrow V^{(t)} - \gamma \nabla_V J(U^{(t)},V^{(t)})$$

in each step

## SGD

While not converged:

① Sample $(i,j)$ from $Z$ uniformly at random

② Compute $e_{ij} = r_{ij} - \vec{U}_i^T \vec{V}_j$

③ Update

$$\vec{U}_i \leftarrow \vec{U}_i - \gamma \nabla_{\vec{U}_i} J_{ij}(U,V)$$

$$\vec{V}_j \leftarrow \vec{V}_j - \gamma \nabla_{\vec{V}_j} J_{ij}(U,V)$$

W/ Regularization

$$J_{ij}(U,V) = \frac{1}{2}\left(r_{ij} - \vec{U}_i^T \vec{V}_j\right)^2 + \lambda\left(\|U_i\|_2^2 + \|V_j\|_2^2\right)$$

$$\nabla_{\vec{U}_i} J_{ij}(U,V) = -e_{ij}\vec{V}_j + \lambda \vec{U}_i$$

$$\nabla_{\vec{V}_j} J_{ij}(U,V) = -e_{ij}\vec{U}_i + \lambda \vec{V}_j$$

where $e_{ij} = r_{ij} - \vec{U}_i^T \vec{V}_j$

## Alternating Least Squares

### Block Coord. Descent:

while not converged:

① $U = \underset{U}{\operatorname{argmin}}\, J(U,V)$

② $V = \underset{V}{\operatorname{argmin}}\, J(U,V)$

convex and easy to solve in closed form

Lin. Reg.

$$J(\Theta) = \frac{1}{2}\sum_{i=1}^{N}\left(y_i - \vec{\Theta}^T \vec{x}_i\right)^2$$

$$J(U,V) = \frac{1}{2}\sum_{(i,j)\in Z}\left(r_{ij} - \vec{U}_i^T \vec{V}_j\right)^2$$

if U is fixed Least Squares in V

if V is fixed Least Squares in U

Option #1: take derivatives, set to zero and solve in closed form

☆ solving $J(U,V)$ in closed form directly isn't easy and $J(U,V)$ is nonconvex

## User/Item Bias terms

$$\hat{r}_{ij} = o_i + p_j + \vec{U}_i^T \vec{V}_j$$

matrix trick:

$$U = \begin{bmatrix} \cdots & o_1 & 1 \\ \cdots & o_2 & 1 \\ & \vdots & \vdots \\ \cdots & o_m & 1 \end{bmatrix} \qquad V = \begin{bmatrix} \cdots & 1 & p_1 \\ \cdots & 1 & p_2 \\ & \vdots & \vdots \\ \cdots & 1 & p_n \end{bmatrix}$$