CSE291B 2/10

Matrix factorization

How to represent / approximate a large matrix M as a product of smaller or simpler matrices?

Today: singular value decomposition,
low-rank approximations (again),
metric multidimensional scaling

Office hours: 3-5:30pm today.
Come with HW questions.

Singular value decomposition

Thm: suppose M is a real m×n matrix
Then there exists a factorization

$$\underset{(m\times n)}{M} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T}$$

Here: U is an m×m orthogonal matrix
V " " n×n " "

$UU^T = U^TU = I_m$
$VV^T = V^TV = I_n$

Σ is m×n non-negative "diagonal" matrix

if m<n: $$\Sigma = \left(\begin{array}{cccc|c} \sigma_1 & \sigma_2 & & \emptyset & \\ & \emptyset & \ddots & & \emptyset \\ & & & \sigma_m & \\ \multicolumn{4}{c}{\longleftarrow m \longrightarrow} & \end{array}\right) \begin{array}{c}\uparrow \\ m \\ \downarrow\end{array}$$

if n<m: $$\Sigma = \begin{array}{c} \longleftarrow n \longrightarrow \\ \left(\begin{array}{cccc} \sigma_1 & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & & \\ & \emptyset & & \\ & & & \end{array}\right) \end{array} \begin{array}{c}\uparrow \\ n \\ \downarrow \\ \\ \\ \\ \end{array}$$

Intuition:

V contains orthonormal "input" analysis vectors
U contains orthonormal "output" vectors.
σ_i singular values are "gain controls".

Hence: matrix M maps $\vec{x} = \sum_{i=1}^{n} \alpha_i \vec{V_i}$ ← colums of V

to $M\vec{x} = \sum_{j=1}^{m} \sigma_j \alpha_j \vec{u_j}$ ← colums of U

Convention:
sort singular values from largest to smallest:

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots\cdots \geq 0$$

Matrix approximation

- Problem: given matrix M, compute $\tilde{M}$ with $\mathrm{rank}(\tilde{M}) = r < \mathrm{rank}(M)$ that is closest to M in minimizing

$$\min_{\tilde{M}} \|M - \tilde{M}\|_F^2 = \sum_{ij} (M_{ij} - \tilde{M}_{ij})^2 = tr\left[(M-\tilde{M})(M-\tilde{M})^T\right]$$

Rank constraint: $\tilde{M}$ has r linearly independent rows or columns.

- Solution: take SVD of $M = U \Sigma V^T$

$$\|M - \tilde{M}\|_F^2 = tr\left[(M-\tilde{M})(M-\tilde{M})^T\right]$$

$$= tr\left[(U\Sigma V^T - \tilde{M})(U\Sigma V^T - \tilde{M})^T\right] = tr\left[(U\Sigma V^T - \tilde{M})(V\Sigma^T U^T - \tilde{M}^T)\right]$$

$$= tr\left[(UU^T)(U\Sigma V^T - \tilde{M})(VV^T)(VV^T)(V\Sigma^T U^T - \tilde{M}^T)(UU^T)\right]$$

$$= tr\left[U(\Sigma - U^T\tilde{M}V)\underbrace{V^T V}_{I}(\Sigma^T - V^T\tilde{M}^T U)U^T\right]$$

Can cycle matrices inside traces
$tr(ABC) = tr(BCA)$

$$= tr\left[(\Sigma - U^T\tilde{M}V)(\Sigma^T - V^T\tilde{M}^T U)(\underbrace{U^T U}_{I})\right]$$

$$= \|\Sigma - U^T \tilde{M} V\|_F^2$$

- Change of variables:

  let $S = U^T \tilde{M} V$

  Note: $\text{rank}(S) = \text{rank}(\tilde{M})$.

  Also: $\tilde{M} = USV^T$

- Optimize over S subject to $\text{rank}(S) = r$:

$$\|M-\tilde{M}\|_F^2 = \|\Sigma - U^T(USV^T)V\|_F^2$$

$$= \|\Sigma - S\|_F^2$$

Assume S is diagonal since Σ is diagonal (See handout for justification.)

$$= \sum_{i=1}^{n} (\sigma_i - s_i)^2$$

where σ_i, s_i are diagonal elements.

From rank constraint: S has (at most) r non-zero diagonal elements.

$$\min_{\{s_i\}} \left[\sum_{i=1}^{n} (\sigma_i - s_i)^2 \right] = \min_{\{s_i\}} \left[\sum_{i=1}^{r} (\sigma_i - s_i)^2 + \underbrace{\sum_{i=r+1}^{n} (\sigma_i - s_i)^2}_{\text{remaining error}} \right]$$

Choose $s_i = \begin{cases} \sigma_i & \text{if } i \le r \\ 0 & \text{otherwise} \end{cases}$

Thus best approximating $\tilde{M}$

$$M = U\Sigma V^T \qquad \Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$$

$$\tilde{M} = USV^T$$

$$S = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \emptyset \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & & \emptyset & & \emptyset \end{array}\right)$$

Metric multidimensional scaling

- Problem: given an $N \times N$ symmetric matrix D_{ij}, how to compute vectors $\vec{x}_i \in \mathbb{R}^N$ such that $\|\vec{x}_i - \vec{x}_j\| = D_{ij}$?

Simplifying assumptions:

(i) $D_{ii} = 0$ for all i

(ii) $D_{ij} \geq 0$ for all i, j

(iii) $D_{ij} + D_{jk} \geq D_{ik}$ for all i, j, k (triangle inequality)

otherwise, no solution exists.

Solution:

Step 1. Derive inner product ("Gram") matrix from distance matrix.

Let $S_{ij} = D_{ij}^2 = \|\vec{x}_i - \vec{x}_j\|^2$ store square distances.

Let $G_{ij} = \vec{x}_i \cdot \vec{x}_j$ store inner products.

How to recover G from S?

other direction (from G to S) is trivial:

$$S_{ij} = \|\vec{x}_i - \vec{x}_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2\vec{x}_i\vec{x}_j$$

$$\boxed{S_{ij} = G_{ii} + G_{jj} - 2G_{ij}} \qquad (*)$$

Without loss of generality, assume vectors are centered at origin:

$$\sum_{i=1}^{N} \vec{x}_i = \vec{0} \qquad \left(\begin{array}{l}\text{distances only determine vectors} \\ \text{up to translation and rotation}\end{array}\right)$$

From centering assumption:

$$\sum_i G_{ij} = \left(\cancelto{0}{\sum_i \vec{x}_i}\right) \cdot \vec{x}_j = 0$$

$$\sum_j G_{ij} = \vec{x}_i \cdot \left(\cancelto{0}{\sum_j \vec{x}_j}\right) = 0$$

$$\sum_{ij} G_{ij} = \left(\sum_i \vec{x}_i\right)\left(\sum_j \vec{x}_j\right) = 0$$

- Diagonal elements of Gram matrix

(i) sum (*) over i: $\sum_i S_{ij} = tr(G) + N G_{jj} - \cancelto{0}{2\sum_i G_{ij}}$

(ii) Sum (*) over i,j: $\sum_{i,j} S_{ij} = 2N\, tr(G)$

From (i): $G_{jj} = \frac{1}{N}\left[\sum_i S_{ij} - tr(G)\right]$

Apply (ii): $\boxed{G_{jj} = \frac{1}{N}\left[\sum_i S_{ij} - \frac{1}{2N}\sum_{kl} S_{kl}\right]}$

Useful shorthand:

$S = \frac{1}{N^2}\sum_{ij} S_{ij}$ row & column average

$S_i = \frac{1}{N}\sum_j S_{ij}$ row average

From above: $\boxed{G_{jj} = S_j - \frac{1}{2} S}$

For non-diagonal demants:

(*) $S_{ij} = G_{ii} + G_{jj} - 2G_{ij}$

$= (S_i - \frac{1}{2}S) + (S_j - \frac{1}{2}S) - 2G_{ij}$

$\Longrightarrow$ $\boxed{G_{ij} = \frac{1}{2}\left[S_i + S_j - S - S_{ij}\right]}$

- Compact notation

Let $I = \begin{pmatrix} 1 & & \emptyset \\ & \ddots & \\ \emptyset & & 1 \end{pmatrix}$ identity matrix

$\vec{u} = \frac{1}{\sqrt{N}}(1, 1, \dots, 1)$ N-dim vector of all "ones"

matrix S

$$\boxed{G = -\frac{1}{2}(I - uu^T)S(I - uu^T)}$$

Now you have a similarity matrix from a distance matrix.

Step2. Perform SVD on Gram matrix

$$G = V\Sigma V^T \quad \text{(here } U=V \text{ because } G \text{ is symmetric)}$$

Let $X = \Sigma^{\frac{1}{2}} V^T$ (since Σ is diagonal, $\Sigma^{\frac{1}{2}}$ is easy to compute)

Then: $G = X^T X$

$$G_{ij} = \sum_k X_{ki} X_{kj} = \vec{x}_i \cdot \vec{x}_j$$ where $\vec{x}_i \in \mathbb{R}^N$ is i^{th} column of $X = \Sigma^{\frac{1}{2}} V^T$

Hence: $\|\vec{x}_i - \vec{x}_j\|^2 = S_{ij}$

- Approximately distance-preserving embedding

Problem: given distances $D_{ij} \in \mathbb{R}^{N\times N}$, how to compute vectors $\vec{x}_i \in \mathbb{R}^d$ (with $d<N$) such that $\|\vec{x}_i - \vec{x}_j\| \approx D_{ij}$?

"Obvious" cost function:

$$\text{error}(X) = \sum_{ij} (D_{ij} - \|\vec{x}_i - \vec{x}_j\|)^2$$ This is difficult to optimize, with many local minima.

Instead minimize

$$\text{error}(X) = \sum_{ij} (G_{ij} - \vec{x}_i \vec{x}_j)^2 \quad \text{where} \begin{cases} S_{ij} = D_{ij}^2 \\ G = -\frac{1}{2}(I - uu^T) S (I - uu^T) \end{cases}$$

This asks for best rank d approximation of G.

Top d components from SVD of G give global minimum:

$$G = V\Sigma V^T \text{ with } \Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_N \end{pmatrix}$$

$\vec{x}_i \in \mathbb{R}^d$ with $\underset{\substack{i=1..N \\ \alpha=1..d}}{X_{i\alpha}} = \sqrt{\sigma_\alpha}\, V_{i\alpha}$