CSE 494 CSE/CBS 598 (Fall 2007): Numerical Linear Algebra for Data Exploration— Singular Value Decomposition

Instructor: Jieping Ye

1 Singular Value Decomposition

- SVD decompose the matrix A by writing it as a product of three more matrices.
- The right side of the equation yields some useful information or insight to the nature of the data matrix A.
- Useful for solving problems from a variety of applications.
- Let A be an $m \times n$ matrix, with $m \geq n$. It can be factorized as

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal

$$\Sigma = \mathrm{diag}\left(\sigma_1, \sigma_2, \cdots, \sigma_n\right), \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

  - Compute the norm of the matrix A

$$||A||_2 = \sigma_1, \quad ||A||_F = \sqrt{\sum_{i=1}^{n} \sigma_i^2}.$$

- Skinny SVD: Partition $U = (U_1, U_2)$, where $U_1 \in \mathbb{R}^{m\times n}$.

$$A = U_1 \Sigma V^T.$$

- An example:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{pmatrix} \begin{pmatrix} 9.64 & 0 \\ 0 & 5.29 \end{pmatrix} \begin{pmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{pmatrix}$$

- Solve least squares problem using SVD: $\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2$.

1.1 Singular Values and Singular Vectors

- The diagonal entries σ_i's of Σ are the **singular values** of A.
- The columns of U and V are the **left singular vectors** and **right singular vectors** respectively.

- The following can be derived from SVD:

$$\begin{aligned} Av_j &= \sigma_j U_j, \\ A^T u_j &= \sigma_j v_j. \end{aligned}$$

- Outer product form:

$$\begin{aligned} A &= U_1 \Sigma V^T = (u_1, \cdots, u_n) \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \cdots & \\ & & & \sigma_4 \end{pmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix} \\ &= (u_1, \cdots, u_n) \begin{pmatrix} \sigma_1 v_1^T \\ \sigma_2 v_2^T \\ \vdots \\ \sigma_n v_n^T \end{pmatrix} = \sum_{i=1}^{n} \sigma_i u_i v_i^T. \end{aligned}$$

  Each term $\sigma_i u_i v_i^T$ in the sum is a rank one matrix. This is thus a sum of rank one matrices.

2 Matrix Approximation

Theorem 2.1 *Let* $U_k = (u_1, \cdots, u_k)$*,* $V_k = (v_1, \cdots, v_k)$*, and* $\Sigma_k = diag(\sigma_1, \sigma_2, \cdots, \sigma_k)$*. Define*

$$A_k = U_k \Sigma_k V_k^T.$$

Then

$$\min_{B: rank(B) \leq k} ||A - B||_2 = ||A - A_k||_2 = \sigma_{k+1}.$$

- A_k is the best approximation of rank k for the matrix A.
- This low rank approximation is useful for
  - Compression
  - Noise reduction
  - finding “concepts or “topics (text mining/LSI)
  - data exploration and visualizing data
  - classification (e.g. handwritten digits)
- SVD appears under different names:
  - Principal Component Analysis (PCA)
  - Latent Semantic Indexing (LSI)/Latent Semantic Analysis (LSA)
  - Karhunen-Loeve expansion/Hotelling transform (in image processing)