SVD

SINGULAR VALUE DECOMPOSITION

Review …

- Last time
  - Recommendation Systems
  - Decompositions of the Utility Matrix
  - Gradient Descent
- Today
  - Singular Value Decomposition

Dimensionality reduction

- Utility Matrix, M, is low rank → Singular Value Decomposition
- $M \rightarrow n \times m$
- $M = UV$
- $U \rightarrow n \times d, \quad V \rightarrow d \times m$

Singular Value Decomposition

SWISS ARMY KNIFE OF LINEAR ALGEBRA

Goal: Given a $m \times n$ matrix A,

$$A = U \Sigma V^* = \sum_{j=1}^{n} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*$$

$m \times n$ $m \times n$ $n \times n$ $n \times n$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the singular values of A

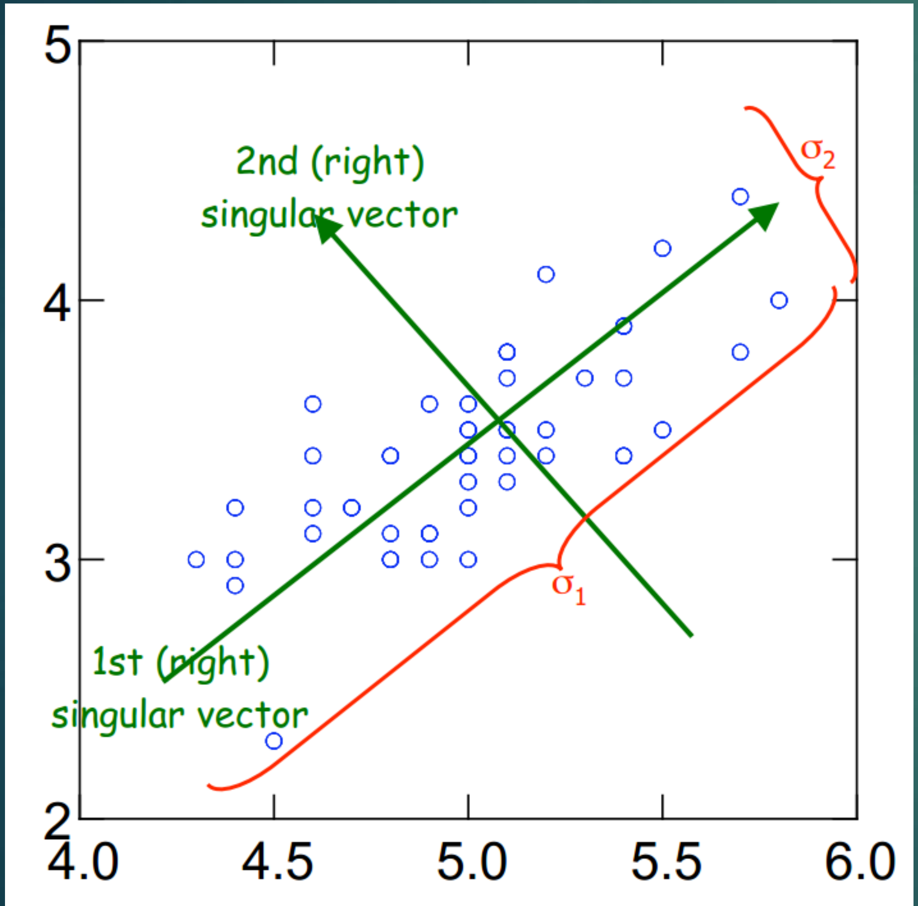
$\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are orthonormal, the left singular vectors of A, and

$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are orthonormal, the right singular vectors of A.

Singular Value Decomposition

- Closely related problems:
  - Eigenvalue decomposition $A \approx V \Lambda V^*$
  - Spanning columns or rows $A \approx C\ U\ R$
- Applications:
  - **Principal Component Analysis**: Form an empirical covariance matrix from some collection of statistical data. By computing the singular value decomposition of the matrix, you find the directions of maximal variance
  - **Finding spanning columns or rows**: Collect statistical data in a large matrix. By finding a set of spanning columns, you can identify some variables that "explain" the data. (Say a small collection of genes among a set of recorded genomes, or a small number of stocks in a portfolio)
  - **Relaxed solutions to k-means clustering**: Relaxed solutions can be found via the singular value decomposition
  - **PageRank**: primary eigenvector

Singular values, intuition



- Blue circles are m data points in $2D$
- The SVD of the $m \times 2$ matrix
  - V_1: 1st (right) singular vector: direction of maximal variance,
  - σ_1: how much of data variance is explained by the first singular vector
  - V_2: 2nd (right) singular vector: direction of maximal variance, after removing projection of the data along first singular vector.
  - σ_2: measures how much of the data variance is explained by the second singular vector

SVD - Interpretation

$M = U\Sigma V^*$ - example:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{array}{l} \qquad\qquad\qquad\qquad\qquad\qquad \mathrm{v1} \\ \begin{bmatrix} \boxed{0.58 \quad 0.58 \quad 0.58 \quad 0 \quad 0} \\ 0 \quad 0 \quad 0 \quad 0.71 \quad 0.71 \end{bmatrix} \end{array}$$

SVD - Interpretation

- $\mathbf{X} = \mathbf{U}\ S\ \mathbf{V}^T$ - example:

variance ('spread') on the v1 axis

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

SVD - Interpretation

$\boldsymbol{M} = \boldsymbol{U\Sigma V}^*$ - example:

- $U\Sigma$ gives the coordinates of the points in the projection axis

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

Dimensionality reduction

set the smallest eigenvalues to zero:

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

Dimensionality reduction

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 \\ 0.36 \\ 0.18 \\ 0.90 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{x} \begin{bmatrix} 9.64 \end{bmatrix} \text{x} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \end{bmatrix}$$

Dimensionality reduction

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Relation to Eigenvalue decomposition

- Eigenvalue decomposition: Can only be applied to certain classes of square matrices
- Given an SVD of a matrix M

$$M^*M = V\Sigma^*U^*\,U\Sigma V^* = V(\Sigma^*\Sigma)V^*$$
$$MM^* = U\Sigma V^*\,V\Sigma^*U^* = U(\Sigma\Sigma^*)U^*$$

- The columns of V are eigenvectors of M^*M
- The columns of U are eigenvectors of MM^*
- The non-zero elements of Σ are the square roots of the non-zero eigenvalues of MM^* or M^*M

Calculating inverses with SVD

- Let A be a $n \times n$ matrix
- Then, $U, \Sigma,$ and V are also $n \times n$
- U and V are orthogonal, so their inverses are equal to their transposes
- Σ is diagonal, so its inverse is the diagonal matrix whose elements are the inverses of the elements of Σ

$$A^{-1} = V \begin{pmatrix} 1/\sigma_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & 1/\sigma_n \end{pmatrix} U^T$$

Calculating inverses

- If one of the σ_i is zero or so small that its value is dominated by round-off error, then there is a problem
- The more of the σ_is that have this problem, the 'more singular' A is
- SVD gives a way of determining how singular A is
- The concept of 'how singular' A is, is linked with the condition number of A
- The condition number of A is the ratio of the largest singular value to its smallest singular value

Concepts you should know

- Null space of A → $x \mid Ax = 0$
- Range of A → $b \mid Ax = b$, for some vector x
- Rank of A → dimension of the range of A

- Singular Valued Decomposition constructs orthonormal bases for the range and null space of a matrix
- The columns of U which correspond to non-zero singular values of A are an orthonormal set of basis vectors for the range of A
- The columns of V which correspond to zero singular values form an orthonormal basis for the null space of A

Computing the SVD

- Reduce the matrix M to a bidiagonal matrix
  - Householder reflections
  - QR decomposition
- Compute the SVD of the bidiagonal matrix
  - Iterative methods

Randomized SVD

Goal: Given a $m \times n$ matrix A, for large m, n, we seek to compute a rank-k approximation, with $k \ll n$,

$$\underset{m \times n}{A} \approx \underset{m \times k}{U} \; \underset{k \times k}{\Sigma} \; \underset{k \times n}{V^*} = \sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*$$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$ are the (approximate) singular values of A
$\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ are orthonormal, the (approximate) left singular vectors of A, and
$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are orthonormal, the (approximate) right singular vectors of A.

Randomized SVD

1. Draw an $n \times k$ Gaussian random matrix, Ω
2. Form the $m \times k$ sample matrix $Y = A\Omega$
3. Form an $m \times k$ orthonormal matrix Q such that $Y = QR$
4. Form the $k \times n$ matrix $B = Q^* A$
5. Compute the SVD of the small matrix $B = \hat{U}\Sigma V^*$
6. Form the matrix $U = Q\hat{U}$

Computational Costs ?

2,4 → k –Matrix-Vector product
3,5,6 → dense operations on matrices $m \times k, k \times n$

Computational Costs

- If A can fit in RAM
  - Cost dominated by $2mnk$ flops required for steps 2,4
- If A cannot fit in RAM
  - Standard approaches suffer
  - Randomized SVD is successful as long as
    - Matrices of size $m \times k$ and $k \times n$ must fit in RAM
- Parallelization
  - Steps 2,4 permit k-way parallelization

Probabilistic Error Analysis

The error of the method is defined as

$$e_k = \|A - \hat{A}_k\|$$

e_k is a random variable whose theoretical minimum value is

$$\sigma_{k+1} = \min(\|A - A_k\| : A_k \text{ has rank } k)$$

Ideally, we would like e_k to be close to σ_{k+1} with high probability

Not true, the expectation of $\frac{e_k}{\sigma_{k+1}}$ is large and has very large variance

Oversampling …

Oversample a little. If p is a small integer (think $p = 5$), then we often can bound e_{k+p} by something close to σ_{k+1}

$$\mathbb{E}\|A - \hat{A}_{k+p}\| \leq \left(1 + \sqrt{\frac{k}{p-1}}\right)\sigma_{k+1} + \frac{e\sqrt{k+p}}{p}\left(\sum_{j=k+1}^{n} \sigma_j^2\right)^{\frac{1}{2}}$$