Singular Value Decomposition

Nicole S. Immorlica and Vahab S. Mirrokni

MIT

Outline

- Motivation for SVD
- Definition of SVD
- Existence and Uniqueness
- Properties of SVDs
- Power Method
- Sampling Algorithm

Recall: Eigenvalues and Eigenvectors

Given:

Matrix $A \in \Re^{m \times n}$

Find:

eigenvectors $v_1, \ldots, v_m \in \Re^n$

eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$

such that $Av_i = \lambda_i v_i$

As the λ *satisfy* $\det(A - \lambda I) = 0$, *they are unique.*

If A *is symmetric, all eigenvalues are real and there is an orthonormal base of eigenvectors:* $V^T A V = D$

Motivation for SVD

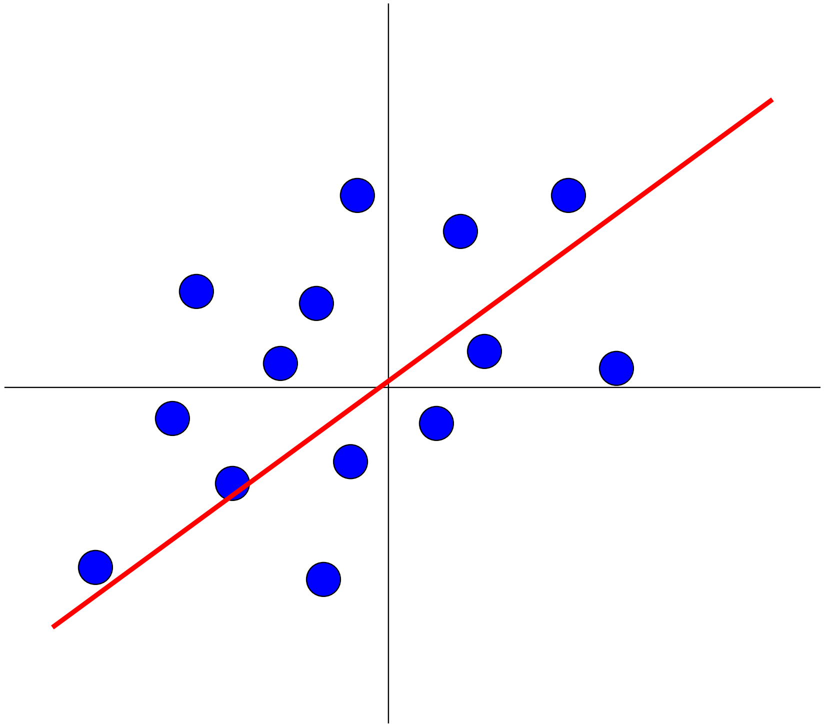
Given:

A set of m points in $\Re^n$

Find:

A low dim subspace that "approximates" the points

Motivation for SVD



The best line is given by the SVD of the matrix of the points!

Motivation for SVD

Another motivation: Google!

Definition of SVD

For any $A \in \Re^{m \times n}$, there exist

$U \in \Re^{m \times m}$, $V \in \Re^{n \times n}$, and $\Sigma \in \Re^{m \times n}$

such that

- $A = U \Sigma V^T$
- U and V are orthonormal
- $\Sigma_{m \times n}$ is diagonal

Definition of SVD

Special case: A is real, symmetric $n \times n$ matrix with real orthonormal eigenvectors B and real eigenvalues D. Then

$$B^T A B = D \Rightarrow A = B D B^T.$$

- $U = V \equiv B$
- $\Sigma \equiv D$

SVD of a sym matrix is its eigenvalue decomposition!

Definition of SVD

Example: compute $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Definition of SVD

The singular values tell us how much matrix acts in each direction.

Existence and Uniqueness

Vector Norms: $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$
the "size" of x

Matrix p-norms: $||A||_p = \sup_{||x||_p=1} ||Ax||_p$
the biggest "effect" of A

Frobenius F-norm $||A||_F^2 = \sum_{i=1}^{m}(\sum_{j=1}^{n} |a_{ij}|^2)$
the "size" of A

Fact The 2-norm and F-norm are invariant with respect to orthogonal transformations.

Existence and Uniqueness

Theorem

The SVD of $A \in \Re^{m \times n}$ exists.

Proof

Construct the singular vectors inductively.

- Take any $v \in \Re^n$, $||v||_2 = 1$
- Define $u \in \Re^n$, $||u||_2 = 1$, such that $Av = ||A||_2 u$
- Extend v and u to orthonormal bases V and U

Then

$$A_1 \equiv U^T A V = \begin{bmatrix} ||A||_2 & w^T \\ 0 & B \end{bmatrix}.$$

Proof (continued)

$$A_1 \equiv U^T A V = \begin{bmatrix} ||A||_2 & w^T \\ 0 & B \end{bmatrix}$$

implies

$$\left\| A_1 \begin{pmatrix} ||A||_2 \\ w \end{pmatrix} \right\|_2^2 \geq (||A||_2^2 + w^T w)^2$$

and so

$$||A||_2^2 = ||A_1||_2^2 \geq ||A||_2^2 + w^T w$$

implies $w = 0$.

Theorem

The singular values Σ of A are unique.

Proof

Notice

$$A = U\Sigma V^T \Rightarrow AA^T = U\Sigma^2 U^T$$

and so σ_i^2's are eigenvalues of AA^T, which are unique.

FACT: We can find the SVD of A by finding the eigenvectors of AA^T.

Properties of SVDs

- Singular values: $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$
- $\text{rank}(A) = r$
- $||A||_F^2 = \sum_{i=1}^{p} \sigma_i^2$
- $||A||_2 = \sigma_1$
- $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = U_r \Sigma_r V_r^T$

The SVD finds the best-fit subspace.

Given:

matrix $A \in \Re^{m \times n}$

Find:

nearest (in 2-norm or F-norm) matrix B of rank at most k

ANSWER: $B = A_k \equiv \sum_{i=1}^{k} \sigma_i u_i v_i^T$

Properties of SVDs

Theorem: The matrix B of rank at most k that minimizes $||A - B||_2$ is

$$A_k \equiv \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

and $||A - B||_2 = \sigma_{k+1}$.

Proof: First we show A_k has the desired properties.

- $U^T A_k V = \text{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) \Rightarrow \text{rank}(A_k) = k$.
- $||A - A_k||_2 = ||U^T(A - A_k)V||_2 = ||\text{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_p)||_2 = \sigma_{k+1}$.

Proof (continued): Suppose B is of rank k.

- $\exists$ orthonormal $x_1, \ldots, x_{n-k}$ that span the null space of B
- $\exists z \in \text{span}\{x_1, \ldots, x_{n-k}\} \cap \text{span}\{v_1, \ldots, v_{k+1}\}$, $||z||_2 = 1$

From here, we see

$$\begin{aligned}
||A - B||_2^2 &\geq ||(A - B)z||_2^2 \\
&= ||Az||_2^2 \\
&= ||\sum_{i=1}^{k+1} \sigma_i u_i (v_i^T z)||_2^2 \\
&= \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2 \\
&\geq \sigma_{k+1}^2
\end{aligned}$$

Power Method

We show how to compute the largest eigenvalue.

- **Idea:** $x = \sum_{i=1}^{n} c_i v_i \Rightarrow A^k x = \sum_{i=1}^{n} c_i \lambda_i^k v_i \simeq c_1 \lambda_1^k v_1$
- Take a random x and compute $A^k x = A(A(A \ldots (A(x)) \ldots))$

take a random $v_{n\times 1}^{(0)}$

for $k = 1, 2, \ldots$

$z^{(k)} = Av^{(k-1)}$

$\lambda^{(k)} = z_i^{(k)}$ where $|z_i^{(k)}| = ||z^{(k)}||_\infty$

$v^{(k)} = z^{(k)}/\lambda^{(k)}$

- v_0 should have a component in the direction of v_1.
- The running time depends on $\frac{|\lambda_1|}{|\lambda_2|}$.

Sampling Algorithm

Select $s = O(k)$ columns of A to create matrix S

Find $S^T S$

Find the top k eigenvectors p_i of $S^T S$

Return p_i and $q_i = Sp_i/|Sp_i|$ as singular vectors of A

FACT: SVD of S is very close to the SVD of A whp.