Big Data Analysis
(MA60306)

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 9
January 20, 2023

Singular value decomposition

Define

$$\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$$

i.e. $\sigma_i \mathbf{u_i}$ is a vector whose coordinates correspond to the projections of the rows of A onto $\mathbf{v_i}$

Singular value decomposition

Define

$$\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$$

i.e. $\sigma_i \mathbf{u_i}$ is a vector whose coordinates correspond to the projections of the rows of A onto $\mathbf{v_i}$

Claim $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$

Singular value decomposition

Define

$$\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$$

i.e. $\sigma_i \mathbf{u_i}$ is a vector whose coordinates correspond to the projections of the rows of A onto $\mathbf{v_i}$

Claim $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$
Proof Obvious!!

Singular value decomposition

Define

$$\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$$

i.e. $\sigma_i \mathbf{u_i}$ is a vector whose coordinates correspond to the projections of the rows of A onto $\mathbf{v_i}$

Claim $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$
Proof Obvious!!

SVD Then we have

$$\begin{bmatrix} \\ n \times d \\ \\ \end{bmatrix} = A = U\Sigma V^T = \begin{bmatrix} \\ n \times r \\ \\ \end{bmatrix} \begin{bmatrix} r \times r \\ \\ \end{bmatrix} \begin{bmatrix} r \times d \\ \\ \end{bmatrix}$$

where $\mathbf{u_i}$ is the ith column of U, $\mathbf{v_i}^T$ is the ith row of V^T, and Σ is a diagonal matrix with σ_i as the ith entry on its diagonal.

Singular value decomposition

Define

$$\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$$

i.e. $\sigma_i \mathbf{u_i}$ is a vector whose coordinates correspond to the projections of the rows of A onto $\mathbf{v_i}$

Claim $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$
Proof Obvious!!

SVD Then we have

$$\begin{bmatrix} \\ n \times d \\ \\ \end{bmatrix} = A = U\Sigma V^T = \begin{bmatrix} \\ n \times r \\ \\ \end{bmatrix} \begin{bmatrix} r \times r \\ \\ \end{bmatrix} \begin{bmatrix} r \times d \\ \\ \end{bmatrix}$$

where $\mathbf{u_i}$ is the ith column of U, $\mathbf{v_i}^T$ is the ith row of V^T, and Σ is a diagonal matrix with σ_i as the ith entry on its diagonal. $\mathbf{v_i}$s are called *right* singular vectors and $\mathbf{u_i}$s are called *left* singular vectors

Question Is this decomposition unique?

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i}\mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i}\mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

$\rightarrow$ Suppose B minimizes $\|A - B\|_F^2$ among all rank k or less matrices

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

→ Suppose B minimizes $\|A - B\|_F^2$ among all rank k or less matrices

→ Let V be the row space of B

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

- → Suppose B minimizes $\|A - B\|_F^2$ among all rank k or less matrices
- → Let V be the row space of B
- → Then each row of B is the projection of the corresponding row of A onto V

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

- → Suppose B minimizes $\|A - B\|_F^2$ among all rank k or less matrices
- → Let V be the row space of B
- → Then each row of B is the projection of the corresponding row of A onto V
- → Then $\|A - B\|_F^2$ is the sum of squared distances of rows of A to V

Singular value decomposition

Let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T, 1 \leq k \leq r$$

Then $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$, the matrix whose rows are the projections of the rows of A onto V_k

Theorem For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

- → Suppose B minimizes $\|A - B\|_F^2$ among all rank k or less matrices
- → Let V be the row space of B
- → Then each row of B is the projection of the corresponding row of A onto V
- → Then $\|A - B\|_F^2$ is the sum of squared distances of rows of A to V
- → Since A_k minimizes the sum of squared distance of rows of A to any k-dimensional subspace, $\|A - A_k\|_F \leq \|A - B\|_F$

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

$\rightarrow$ It takes $O(nd)$ to compute $A\mathbf{x}$

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

- $\rightarrow$ It takes $O(nd)$ to compute $A\mathbf{x}$
- $\rightarrow$ Suppose we approximate A by A_k, and then approximate $A\mathbf{x}$ by $A_k\mathbf{x}$, which can be done in $O(kd + kn)$, where $k \ll \min\{d, n\}$

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

- $\rightarrow$ It takes $O(nd)$ to compute $A\mathbf{x}$
- $\rightarrow$ Suppose we approximate A by A_k, and then approximate $A\mathbf{x}$ by $A_k\mathbf{x}$, which can be done in $O(kd + kn)$, where $k \ll \min\{d, n\}$
- $\rightarrow$ How much will be the error?

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

- $\rightarrow$ It takes $O(nd)$ to compute $A\mathbf{x}$
- $\rightarrow$ Suppose we approximate A by A_k, and then approximate $A\mathbf{x}$ by $A_k\mathbf{x}$, which can be done in $O(kd + kn)$, where $k \ll \min\{d, n\}$
- $\rightarrow$ How much will be the error?

$$\max_{\|x\|_2=1} \|(A - A_k)x = \|A - A_k\|_2$$

Singular value decomposition

Observation Suppose we want to compute $A\mathbf{x}$, where A is an $n \times d$ large matrix

- $\rightarrow$ It takes $O(nd)$ to compute $A\mathbf{x}$
- $\rightarrow$ Suppose we approximate A by A_k, and then approximate $A\mathbf{x}$ by $A_k\mathbf{x}$, which can be done in $O(kd + kn)$, where $k \ll \min\{d, n\}$
- $\rightarrow$ How much will be the error?

$$\max_{\|x\|_2=1} \|(A - A_k)x = \|A - A_k\|_2$$

Homework $\|A\|_2 = \sigma_1(A)$, called the *spectral* norm of A!!

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

$\rightarrow$ Without loss of generality assume that $\mathbf{u_i}^T \mathbf{u_j} = \delta > 0$

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

→ Without loss of generality assume that $\mathbf{u_i}^T \mathbf{u_j} = \delta > 0$

→ For $\epsilon > 0$, let

$$\mathbf{v_i'} = \frac{\mathbf{v_i} + \epsilon \mathbf{v_j}}{\|\mathbf{v_i} + \epsilon \mathbf{v_j}\|_2}$$

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

→ Without loss of generality assume that $\mathbf{u_i}^T\mathbf{u_j} = \delta > 0$

→ For $\epsilon > 0$, let

$$\mathbf{v_i'} = \frac{\mathbf{v_i} + \epsilon\mathbf{v_j}}{\|\mathbf{v_i} + \epsilon\mathbf{v_j}\|_2}$$

→ Then $A\mathbf{v_i'} = \frac{\sigma_i\mathbf{u_i} + \epsilon\sigma_j\mathbf{u_j}}{\sqrt{1+\epsilon^2}}$ which has length at least as large as its component along $\mathbf{u_i}$ given by

$$\mathbf{u_i}^T\left(\frac{\sigma_i\mathbf{u_i} + \epsilon\sigma_j\mathbf{u_j}}{\sqrt{1+\epsilon^2}}\right) \quad > \quad (\sigma_i + \epsilon\sigma_j\delta)\left(1 - \frac{\epsilon^2}{2}\right)$$

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

→ Without loss of generality assume that ${\mathbf{u_i}}^T \mathbf{u_j} = \delta > 0$

→ For $\epsilon > 0$, let

$$\mathbf{v_i'} = \frac{\mathbf{v_i} + \epsilon \mathbf{v_j}}{\|\mathbf{v_i} + \epsilon \mathbf{v_j}\|_2}$$

→ Then $A\mathbf{v_i'} = \frac{\sigma_i \mathbf{u_i} + \epsilon \sigma_j \mathbf{u_j}}{\sqrt{1+\epsilon^2}}$ which has length at least as large as its component along $\mathbf{u_i}$ given by

$$\begin{aligned} {\mathbf{u_i}}^T \left(\frac{\sigma_i \mathbf{u_i} + \epsilon \sigma_j \mathbf{u_j}}{\sqrt{1+\epsilon^2}} \right) &> (\sigma_i + \epsilon \sigma_j \delta) \left(1 - \frac{\epsilon^2}{2} \right) \\ &= \sigma_i - \frac{\epsilon^2}{2}\sigma_i + \epsilon \sigma_j \delta - \frac{\epsilon^3}{2}\sigma_j \delta \end{aligned}$$

Singular value decomposition

Theorem The left singular vectors are pairwise orthogonal

Proof Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$, $j > i$

→ Without loss of generality assume that $\mathbf{u_i}^T\mathbf{u_j} = \delta > 0$

→ For $\epsilon > 0$, let

$$\mathbf{v_i'} = \frac{\mathbf{v_i} + \epsilon\mathbf{v_j}}{\|\mathbf{v_i} + \epsilon\mathbf{v_j}\|_2}$$

→ Then $A\mathbf{v_i'} = \frac{\sigma_i\mathbf{u_i} + \epsilon\sigma_j\mathbf{u_j}}{\sqrt{1+\epsilon^2}}$ which has length at least as large as its component along $\mathbf{u_i}$ given by

$$\begin{aligned}
\mathbf{u_i}^T\left(\frac{\sigma_i\mathbf{u_i} + \epsilon\sigma_j\mathbf{u_j}}{\sqrt{1+\epsilon^2}}\right) &> (\sigma_i + \epsilon\sigma_j\delta)\left(1 - \frac{\epsilon^2}{2}\right) \\
&= \sigma_i - \frac{\epsilon^2}{2}\sigma_i + \epsilon\sigma_j\delta - \frac{\epsilon^3}{2}\sigma_j\delta \\
&\approx \sigma_i
\end{aligned}$$

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ.

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ. However this is a contradiction!!

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ. However this is a contradiction!!

Lemma $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ. However this is a contradiction!!

Lemma $\|A - A_k\|_2^2 = \sigma_{k+1}^2$
Proof Let $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ be the SVD of A, and $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$.

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ. However this is a contradiction!!

Lemma $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

Proof Let $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i}\mathbf{v_i}^T$ be the SVD of A, and $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i}\mathbf{v_i}^T$. Then

$$A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i}\mathbf{v_i}^T.$$

Singular value decomposition

Note that the above inequality is true for sufficiently small ϵ. However this is a contradiction!!

Lemma $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

Proof Let $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ be the SVD of A, and $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. Then

$$A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T.$$

Let $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}, \quad \underbrace{\mathbf{v_{r+1}}, \ldots, \mathbf{v_d}}_{\text{extended portion}} \quad \}$ be an ONB of $\mathbb{R}^d$. Then for the top singular vector $\mathbf{v}$ of $A - A_k$, writing $\mathbf{v} = \sum_{j=1}^{d} c_j \mathbf{v_j}$, we have

$$\|(A - A_k)\mathbf{v}\|_2 = \left\| \sum_{j=1}^{r} c_i \sigma_i \mathbf{u_i} \right\|_2 = \sqrt{\sum_{i=k+1}^{r} c_i^2 \sigma_i^2}$$

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?
Since $\sum_{i=1}^{r} c_i^2 = 1$, set $c_{k+1} = 1$ and rest of the $c_i = 0$. This completes the proof

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?
Since $\sum_{i=1}^{r} c_i^2 = 1$, set $c_{k+1} = 1$ and rest of the $c_i = 0$. This completes the proof
Question What is the conclusion from the above lemma?

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?
Since $\sum_{i=1}^{r} c_i^2 = 1$, set $c_{k+1} = 1$ and rest of the $c_i = 0$. This completes the proof
Question What is the conclusion from the above lemma?

Theorem For any matrix B of rank at most k,

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?

Since $\sum_{i=1}^{r} c_i^2 = 1$, set $c_{k+1} = 1$ and rest of the $c_i = 0$. This completes the proof

Question What is the conclusion from the above lemma?

Theorem For any matrix B of rank at most k,

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

Proof Homework!!

Singular value decomposition

Question Which vector $\mathbf{v}$ gives the maximum value of $\|(A - A_k)\mathbf{v}\|_2$ in the above expression?
Since $\sum_{i=1}^{r} c_i^2 = 1$, set $c_{k+1} = 1$ and rest of the $c_i = 0$. This completes the proof
Question What is the conclusion from the above lemma?

Theorem For any matrix B of rank at most k,

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

Proof Homework!!
Analog of eigenvalues and eigenvectors

$$A\mathbf{v_i} = \sigma_i \mathbf{u_i} \text{ and } A^T\mathbf{u_i} = \sigma_i \mathbf{v_i}$$