Big Data Analysis
(MA60306)

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 8
January 19, 2023

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

$\rightarrow$ Consider each row of A as a point in d-dimensional space i.e. a data point has d features

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

- → Consider each row of A as a point in d-dimensional space i.e. a data point has d features
- → The problem: find the best-fitting k-dimensional subspace that is close to all the given data points

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

- → Consider each row of A as a point in d-dimensional space i.e. a data point has d features
- → The problem: find the best-fitting k-dimensional subspace that is close to all the given data points
- → "best" meaning - minimizing the sum of the squares of the perpendicular distances of the points to the subspace

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

- → Consider each row of A as a point in d-dimensional space i.e. a data point has d features
- → The problem: find the best-fitting k-dimensional subspace that is close to all the given data points
- → "best" meaning - minimizing the sum of the squares of the perpendicular distances of the points to the subspace
- → Thus the is to maximize the sum of squares of the lengths of the projections of the points onto the subspace

Singular value decomposition

Regression problem Can we construct a plane which contains a given set of data points!

Let A be an $n \times d$ data matrix.

- $\rightarrow$ Consider each row of A as a point in d-dimensional space i.e. a data point has d features
- $\rightarrow$ The problem: find the best-fitting k-dimensional subspace that is close to all the given data points
- $\rightarrow$ "best" meaning - minimizing the sum of the squares of the perpendicular distances of the points to the subspace
- $\rightarrow$ Thus the is to maximize the sum of squares of the lengths of the projections of the points onto the subspace
- $\rightarrow$ What should be the value of k?

Singular value decomposition

Consider a data point $\mathbf{a_i}^T = (a_{i1}, a_{i2}, \ldots, a_{id})$, which represents a vector (line) through the origin.
Then

$$a_{i1}^2 + a_{i2}^2 + \ldots + a_{id}^2 = (\text{length of projection})^2 + (\text{distance of point to line})^2$$

i.e.

$$(\text{distance of point to line})^2 = a_{i1}^2 + a_{i2}^2 + \ldots + a_{id}^2 - (\text{length of projection})^2$$

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Now

$$\begin{aligned} \|A\mathbf{v}\|_2^2 &= |\mathbf{a_1}^T\mathbf{v}|^2 + \ldots + |\mathbf{a_n}^T\mathbf{v}|^2 \\ &= \text{sum of squres of projection lengths} \end{aligned}$$

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Now

$$\begin{aligned} \|A\mathbf{v}\|_2^2 &= |\mathbf{a_1}^T\mathbf{v}|^2 + \ldots + |\mathbf{a_n}^T\mathbf{v}|^2 \\ &= \text{sum of squres of projection lengths} \end{aligned}$$

Since best-fit line is the one that maximizes $\|A\mathbf{v}\|_2^2$ (equivalent to saying minimizing the sum of the squared distances of the points to the line), define

$$\mathbf{v_1} = \arg \max_{\|\mathbf{v}\|_2=1} \|A\mathbf{v}\|_2$$

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Now

$$\begin{aligned} \|A\mathbf{v}\|_2^2 &= |\mathbf{a_1}^T\mathbf{v}|^2 + \ldots + |\mathbf{a_n}^T\mathbf{v}|^2 \\ &= \text{sum of squres of projection lengths} \end{aligned}$$

Since best-fit line is the one that maximizes $\|A\mathbf{v}\|_2^2$ (equivalent to saying minimizing the sum of the squared distances of the points to the line), define

$$\mathbf{v_1} = \arg\max_{\|\mathbf{v}\|_2=1} \|A\mathbf{v}\|_2$$

$\rightarrow$ Let us call $\mathbf{v_1}$ as the first singular vector

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Now

$$\begin{aligned} \|A\mathbf{v}\|_2^2 &= |\mathbf{a_1}^T\mathbf{v}|^2 + \ldots + |\mathbf{a_n}^T\mathbf{v}|^2 \\ &= \text{sum of squres of projection lengths} \end{aligned}$$

Since best-fit line is the one that maximizes $\|A\mathbf{v}\|_2^2$ (equivalent to saying minimizing the sum of the squared distances of the points to the line), define

$$\mathbf{v_1} = \arg\max_{\|\mathbf{v}\|_2=1} \|A\mathbf{v}\|_2$$

→ Let us call $\mathbf{v_1}$ as the first singular vector

→ The value $\sigma_1(A) = \|A\mathbf{v_1}\|_2$ is called the first singular value of A.

Singular value decomposition

Given the data matrix A with rows of A as the data points, and for a unit vector $\mathbf{v}$, the length of projection of $\mathbf{a_i}^T$, the ith row of A, onto $\mathbf{v}$ is $|\mathbf{a_i}^T\mathbf{v}|$

Now

$$\begin{aligned} \|A\mathbf{v}\|_2^2 &= |\mathbf{a_1}^T\mathbf{v}|^2 + \ldots + |\mathbf{a_n}^T\mathbf{v}|^2 \\ &= \text{sum of squres of projection lengths} \end{aligned}$$

Since best-fit line is the one that maximizes $\|A\mathbf{v}\|_2^2$ (equivalent to saying minimizing the sum of the squared distances of the points to the line), define

$$\mathbf{v_1} = \arg\max_{\|\mathbf{v}\|_2=1} \|A\mathbf{v}\|_2$$

- → Let us call $\mathbf{v_1}$ as the first singular vector
- → The value $\sigma_1(A) = \|A\mathbf{v_1}\|_2$ is called the first singular value of A. Then

$$\sigma_1^2 = \sum_{i=1}^{n} (\mathbf{a_i}^T\mathbf{v_1})^2$$

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

→ If the data points lie on a line or close to a line then $\mathbf{v_1}$ is the desired line

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

→ If the data points lie on a line or close to a line then $\mathbf{v_1}$ is the desired line

→ If the data points are not close to a line but a 2-dimensional subspace then we should look for a 2-d subspace that contains $\mathbf{v_1}$

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

- $\rightarrow$ If the data points lie on a line or close to a line then $\mathbf{v_1}$ is the desired line
- $\rightarrow$ If the data points are not close to a line but a 2-dimensional subspace then we should look for a 2-d subspace that contains $\mathbf{v_1}$
- $\rightarrow$ For every 2-d subspace containing $\mathbf{v_1}$, the sum of squared lengths of the projections onto the subspace equals the sum of squared projections onto $\mathbf{v_1}$ + the sum of squared projections along a vector perpendicular to $\mathbf{v_1}$ in the subspace

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

- $\rightarrow$ If the data points lie on a line or close to a line then $\mathbf{v_1}$ is the desired line
- $\rightarrow$ If the data points are not close to a line but a 2-dimensional subspace then we should look for a 2-d subspace that contains $\mathbf{v_1}$
- $\rightarrow$ For every 2-d subspace containing $\mathbf{v_1}$, the sum of squared lengths of the projections onto the subspace equals the sum of squared projections onto $\mathbf{v_1}$ + the sum of squared projections along a vector perpendicular to $\mathbf{v_1}$ in the subspace
- $\rightarrow$ Thus, instead of looking for the best 2-d subspace containing v_1, look for a unit vector v_2 perpendicular to $\mathbf{v_1}$ that maximizes $\|A\mathbf{v}\|_2^2$ among all such unit vectors.

Singular value decomposition

Then

$$\sigma_1^2 = \text{the sum of the squared lengths of the projections of the points onto the line } \mathbf{v_1}$$

- $\rightarrow$ If the data points lie on a line or close to a line then $\mathbf{v_1}$ is the desired line
- $\rightarrow$ If the data points are not close to a line but a 2-dimensional subspace then we should look for a 2-d subspace that contains $\mathbf{v_1}$
- $\rightarrow$ For every 2-d subspace containing $\mathbf{v_1}$, the sum of squared lengths of the projections onto the subspace equals the sum of squared projections onto $\mathbf{v_1}$ $+$ the sum of squared projections along a vector perpendicular to $\mathbf{v_1}$ in the subspace
- $\rightarrow$ Thus, instead of looking for the best 2-d subspace containing v_1, look for a unit vector v_2 perpendicular to $\mathbf{v_1}$ that maximizes $\|A\mathbf{v}\|_2^2$ among all such unit vectors.
- $\rightarrow$ And the process goes on.... using the greedy strategy

Singular value decomposition

The second singular vector, $\mathbf{v_2}$ is given by

$$\mathbf{v_2} = \arg \max_{\mathbf{v} \perp \mathbf{v_1}, \|\mathbf{v}\|_2 = 1} \|A\mathbf{v}\|_2$$

Singular value decomposition

The second singular vector, $\mathbf{v_2}$ is given by

$$\mathbf{v_2} = \arg \max_{\mathbf{v} \perp \mathbf{v_1}, \|\mathbf{v}\|_2 = 1} \|A\mathbf{v}\|_2$$

$$\sigma_2(A) = \|A\mathbf{v_2}\|_2$$

is called the second singular value of A.

Singular value decomposition

The second singular vector, $\mathbf{v_2}$ is given by

$$\mathbf{v_2} = \arg \max_{\mathbf{v} \perp \mathbf{v_1}, \|\mathbf{v}\|_2 = \mathbf{1}} \|A\mathbf{v}\|_2$$

$$\sigma_2(A) = \|A\mathbf{v_2}\|_2$$

is called the second singular value of A.

The process continues, and finally, we can obtain singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ such that

$$\max_{\mathbf{v} \perp \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}, \|\mathbf{v}\|_2 = 1} \|A\mathbf{v}\|_2 = 0$$

Singular value decomposition

The second singular vector, $\mathbf{v_2}$ is given by

$$\mathbf{v_2} = \arg \max_{\mathbf{v} \perp \mathbf{v_1}, \|\mathbf{v}\|_2 = \mathbf{1}} \|A\mathbf{v}\|_2$$

$$\sigma_2(A) = \|A\mathbf{v_2}\|_2$$

is called the second singular value of A.

The process continues, and finally, we can obtain singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ such that

$$\max_{\mathbf{v} \perp \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}, \|\mathbf{v}\|_2 = 1} \|A\mathbf{v}\|_2 = 0$$

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

→ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

- $\rightarrow$ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W
- $\rightarrow$ Choose $(\mathbf{w_1}, \mathbf{w_2})$ such that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

- $\rightarrow$ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W
- $\rightarrow$ Choose $(\mathbf{w_1}, \mathbf{w_2})$ such that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$
- $\rightarrow$ Choose $\mathbf{w_2}$ as the unit vector in W perpendicular to the projection of $\mathbf{v_1}$ onto W

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

- $\rightarrow$ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W
- $\rightarrow$ Choose $(\mathbf{w_1}, \mathbf{w_2})$ such that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$
- $\rightarrow$ Choose $\mathbf{w_2}$ as the unit vector in W perpendicular to the projection of $\mathbf{v_1}$ onto W
- $\rightarrow$ Since $\mathbf{v_1}$ maximizes $\|A\mathbf{v}\|_2^2$, $\|A\mathbf{w_1}\|_2^2 \leq \|A\mathbf{v_1}\|_2^2$

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

- $\rightarrow$ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W
- $\rightarrow$ Choose $(\mathbf{w_1}, \mathbf{w_2})$ such that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$
- $\rightarrow$ Choose $\mathbf{w_2}$ as the unit vector in W perpendicular to the projection of $\mathbf{v_1}$ onto W
- $\rightarrow$ Since $\mathbf{v_1}$ maximizes $\|A\mathbf{v}\|_2^2$, $\|A\mathbf{w_1}\|_2^2 \leq \|A\mathbf{v_1}\|_2^2$
- $\rightarrow$ Thus $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2 \leq \|A\mathbf{v_1}\|_2^2 + \|A\mathbf{v_2}\|_2^2$

Singular value decomposition

Theorem Let A be an $n \times d$ matrix with singular vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. For $1 \leq k \leq r$, let V_k be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$. For each k, V_k is the best-fit k-dimensional subspace for A.

Proof For $k = 1$ the statement if obviously true. For $k = 2$, let W be the best fit 2-d subspace for A.

- $\rightarrow$ For any orthonormal basis $(\mathbf{w_1}, \mathbf{w_2})$ of W, $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2$ is the sum of squared lengths of the projections of the rows of A onto W
- $\rightarrow$ Choose $(\mathbf{w_1}, \mathbf{w_2})$ such that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$
- $\rightarrow$ Choose $\mathbf{w_2}$ as the unit vector in W perpendicular to the projection of $\mathbf{v_1}$ onto W
- $\rightarrow$ Since $\mathbf{v_1}$ maximizes $\|A\mathbf{v}\|_2^2$, $\|A\mathbf{w_1}\|_2^2 \leq \|A\mathbf{v_1}\|_2^2$
- $\rightarrow$ Thus $\|A\mathbf{w_1}\|_2^2 + \|A\mathbf{w_2}\|_2^2 \leq \|A\mathbf{v_1}\|_2^2 + \|A\mathbf{v_2}\|_2^2$
- $\rightarrow$ Then use induction hypothesis

Singular value decomposition

Observation Note that for any row $\mathbf{a_j}^T$ of A,

$$\sum_{i=1}^{r}(\mathbf{a_j^T v_i})^2 = \|\mathbf{a_j}\|_2^2$$

Singular value decomposition

Observation Note that for any row $\mathbf{a_j}^T$ of A,

$$\sum_{i=1}^{r}(\mathbf{a_j^T v_i})^2 = \|\mathbf{a_j}\|_2^2$$

Then

$$\sum_{j=1}^{n}\|\mathbf{a_j}\|_2^2 = \sum_{j=1}^{n}\sum_{i=1}^{r}(\mathbf{a_j}^T\mathbf{v_i})^2 = \sum_{i=1}^{r}\|A\mathbf{v_i}\|_2^2 = \sum_{i=1}^{r}\sigma_i^2(A)$$

Singular value decomposition

Observation Note that for any row $\mathbf{a_j}^T$ of A,

$$\sum_{i=1}^{r}(\mathbf{a_j^T v_i})^2 = \|\mathbf{a_j}\|_2^2$$

Then

$$\sum_{j=1}^{n}\|\mathbf{a_j}\|_2^2 = \sum_{j=1}^{n}\sum_{i=1}^{r}(\mathbf{a_j}^T\mathbf{v_i})^2 = \sum_{i=1}^{r}\|A\mathbf{v_i}\|_2^2 = \sum_{i=1}^{r}\sigma_i^2(A)$$

Also, $\|A\|_F = \sqrt{\sum_{j,k} a_{jk}^2} = \sqrt{\sum_{j=1}^{n}\|\mathbf{a_j}\|_2^2} = \sqrt{\sum_{i=1}^{r}\sigma_i^2(A)}$

Singular value decomposition

Observation Note that for any row ${\mathbf{a_j}}^T$ of A,

$$\sum_{i=1}^{r}(\mathbf{a_j^T v_i})^2 = \|\mathbf{a_j}\|_2^2$$

Then

$$\sum_{j=1}^{n}\|\mathbf{a_j}\|_2^2 = \sum_{j=1}^{n}\sum_{i=1}^{r}({\mathbf{a_j}}^T\mathbf{v_i})^2 = \sum_{i=1}^{r}\|A\mathbf{v_i}\|_2^2 = \sum_{i=1}^{r}\sigma_i^2(A)$$

Also, $\|A\|_F = \sqrt{\sum_{j,k} a_{jk}^2} = \sqrt{\sum_{j=1}^{n}\|\mathbf{a_j}\|_2^2} = \sqrt{\sum_{i=1}^{r}\sigma_i^2(A)}$

Conclusion The sum of squares of the singular values of A is the square of the "whole content of A", i.e., the sum of squares of all the entries