

# AAIL100: Introduction to Data Science

Ragesh Jaiswal, IITD

## Best Fit Subspaces and Singular Value Decomposition (SVD)

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- Question: How do we define **best fit line**?

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- Question: How do we define **best fit line**?
  - A line that minimises the sum of squared distance of the  $n$  points to the line.

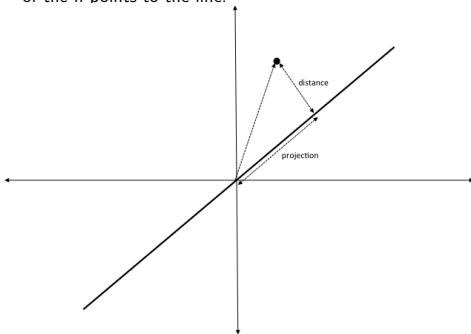
# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- Question: How do we define **best fit line**?
  - A line that minimises the sum of squared distance of the  $n$  points to the line.
  - Claim: The best fit line maximises the sum of projections squared of the  $n$  points to the line.



# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- The best fit line through the origin is one that minimises the sum of squared distance of the  $n$  points to the line.
- Let  $\mathbf{v}$  denote a unit vector ( $d \times 1$  matrix) in the direction of the best fit line.
- Claim: The sum of squared lengths of projections of the points onto  $\mathbf{v}$  is  $\|A\mathbf{v}\|^2$ .

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- The best fit line through the origin is one that minimises the sum of squared distance of the  $n$  points to the line.
- Let  $\mathbf{v}$  denote a unit vector ( $d \times 1$  matrix) in the direction of the best fit line.
- Claim: The sum of squared lengths of projections of the points onto  $\mathbf{v}$  is  $\|A\mathbf{v}\|^2$ .
- So, the best fit line is defined by unit vector  $\mathbf{v}$  that maximises  $\|A\mathbf{v}\|$ .
- This is the **first singular vector** of the matrix  $A$ . So, the first singular vector is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- The best fit line through the origin is one that minimises the sum of squared distance of the  $n$  points to the line.
- Let  $\mathbf{v}$  denote a unit vector ( $d \times 1$  matrix) in the direction of the best fit line.
- Claim: The sum of squared lengths of projections of the points onto  $\mathbf{v}$  is  $\|A\mathbf{v}\|^2$ .
- So, the best fit line is defined by unit vector  $\mathbf{v}$  that maximises  $\|A\mathbf{v}\|$ .
- This is the **first singular vector** of the matrix  $A$ . So, the first singular vector is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

- The value  $\sigma_1 = \|A\mathbf{v}_1\|$  is called the first singular value of  $A$ .

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit line** through the origin for the given  $n$  points.

- The first singular vector is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\|$$

- The value  $\sigma_1 = \|\mathbf{A}\mathbf{v}_1\|$  is called the first singular value of  $A$ .
- So,  $\sigma_1^2$  is equal to the sum of squared length of projections.
- Note that if all the data points are “close” to a line through the origin, then the first singular vector gives such a line.
- Question: if the data points are close to a plane (and in general close to a  $k$ -dimensional subspace), then how do we find such a plane?

# Best Fit Subspaces and SVD

## Best fit line

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit plane** through the origin for the given  $n$  points.

- Let  $\mathbf{v}_1$  denote the first singular vector of  $A$ .
- Idea: Find a unit vector  $\mathbf{v}$  perpendicular to  $\mathbf{v}_1$  that maximises  $\|A\mathbf{v}\|$ . Output the plane through the origin defined by vectors  $\mathbf{v}_1$  and  $\mathbf{v}$ .
- Claim: The plane defined above indeed maximises sum of squared distances of all the points.
- The second singular vector is defined as:

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|A\mathbf{v}\|.$$

- The value  $\sigma_2 = \|A\mathbf{v}_2\|$  is called the second singular value of  $A$ .

# Best Fit Subspaces and SVD

## Best fit plane

### Problem

Given an  $n \times d$  matrix  $A$ , where we interpret the rows of the matrix as points in  $\mathbb{R}^d$ , find a **best fit plane** through the origin for the given  $n$  points.

- Let  $\mathbf{v}_1$  denote the first singular vector of  $A$ .
- The second singular vector is defined as:

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|A\mathbf{v}\|.$$

- The value  $\sigma_2 = \|A\mathbf{v}_2\|$  is called the second singular value of  $A$ .

### Theorem

*For any matrix  $A$ , the plane spanned by  $\mathbf{v}_1$  and  $\mathbf{v}_2$  is the best fit plane.*

# Best Fit Subspaces and SVD

## Best fit plane

- The first singular vector is defined as:  $\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\|$ .
- The second singular vector is defined as:  
 $\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|\mathbf{A}\mathbf{v}\|$ .

### Theorem

*For any matrix  $A$ , the plane spanned by  $\mathbf{v}_1$  and  $\mathbf{v}_2$  is the best fit plane.*

### Proof sketch

- Let  $W$  denote the best fit plane for  $A$ .
- Claim 1: There exists an orthonormal basis  $(\mathbf{w}_1, \mathbf{w}_2)$  of  $W$  such that  $\mathbf{w}_2$  is perpendicular to  $\mathbf{v}_1$ .
- Claim 2:  $\|\mathbf{A}\mathbf{w}_1\|^2 \leq \|\mathbf{A}\mathbf{v}_1\|^2$ .
- Claim 3:  $\|\mathbf{A}\mathbf{w}_2\|^2 \leq \|\mathbf{A}\mathbf{v}_2\|^2$ .
- This gives  $\|\mathbf{A}\mathbf{w}_1\|^2 + \|\mathbf{A}\mathbf{w}_2\|^2 \leq \|\mathbf{A}\mathbf{v}_1\|^2 + \|\mathbf{A}\mathbf{v}_2\|^2$ . □

# Best Fit Subspaces and SVD

## Best fit subspace

- The first singular vector and first singular value is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_1 = \|\mathbf{A}\mathbf{v}_1\|$$

- The second singular vector and second singular value is defined as:

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_2 = \|\mathbf{A}\mathbf{v}_2\|.$$

- The third singular vector and third singular value is defined as:

$$\mathbf{v}_3 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \mathbf{v}_2} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_3 = \|\mathbf{A}\mathbf{v}_3\|.$$

- ...and so on.

- Let  $r$  be the smallest positive integer such that:

$\max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_r} \|\mathbf{A}\mathbf{v}\| = 0$ . Then  $A$  has  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ .

### Theorem

Let  $A$  be any  $n \times d$  matrix with  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ . For  $1 \leq k \leq r$ , let  $V_k$  be the subspace spanned by  $\mathbf{v}_1, \dots, \mathbf{v}_k$ . For each  $k$ ,  $V_k$  is the best-fit  $k$ -dimensional subspace for  $A$ .

# Best Fit Subspaces and SVD

## Best fit subspace

- The first singular vector and first singular value is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_1 = \|\mathbf{A}\mathbf{v}_1\|$$

- The second singular vector and second singular value is defined as:

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_2 = \|\mathbf{A}\mathbf{v}_2\|.$$

- The third singular vector and third singular value is defined as:

$$\mathbf{v}_3 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \mathbf{v}_2} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_3 = \|\mathbf{A}\mathbf{v}_3\|.$$

- ...and so on.
- Let  $r$  be the smallest positive integer such that:  
 $\max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_r} \|\mathbf{A}\mathbf{v}\| = 0$ . Then  $A$  has  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ .
- The vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  are more specifically called the **right singular vectors**.

# Best Fit Subspaces and SVD

## Best fit subspace

- The first singular vector and first singular value is defined as:

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_1 = \|\mathbf{A}\mathbf{v}_1\|$$

- The second singular vector and second singular value is defined as:

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_2 = \|\mathbf{A}\mathbf{v}_2\|.$$

- The third singular vector and third singular value is defined as:

$$\mathbf{v}_3 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \mathbf{v}_2} \|\mathbf{A}\mathbf{v}\| \quad \text{and} \quad \sigma_3 = \|\mathbf{A}\mathbf{v}_3\|.$$

- ...and so on.
- Let  $r$  be the smallest positive integer such that:  
 $\max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_r} \|\mathbf{A}\mathbf{v}\| = 0$ . Then  $A$  has  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ .
- The vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  are more specifically called the **right singular vectors**.
- For any singular vector  $\mathbf{v}_i$ ,  $\sigma_i = \|\mathbf{A}\mathbf{v}_i\|$  may be interpreted as the **component** of the matrix  $A$  along  $\mathbf{v}_i$ .
- Given this interpretation, the *"the components should add up to give the whole content of A"*.

# Best Fit Subspaces and SVD

## Frobenius Norm

- Let  $r$  be the smallest positive integer such that:  
 $\max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_r} \|\mathbf{A}\mathbf{v}\| = 0$ . Then  $A$  has  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ .
- The vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  are more specifically called the **right singular vectors**.
- For any singular vector  $\mathbf{v}_i$ ,  $\sigma_i = \|\mathbf{A}\mathbf{v}_i\|$  may be interpreted as the **component** of the matrix  $A$  along  $\mathbf{v}_i$ .
- Given this interpretation, the “*the components should add up to give the whole content of A*”.
- For any row  $a_j$  in the matrix  $A$ , we can write  $\|a_j\|^2 = \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2$ .  
This further gives:

$$\sum_{j=1}^n \|a_j\|^2 = \sum_{j=1}^n \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2 = \sum_{i=1}^r \|\mathbf{A}\mathbf{v}_i\|^2 = \sum_{i=1}^r \sigma_i^2.$$

# Best Fit Subspaces and SVD

## Frobenius Norm

- Let  $r$  be the smallest positive integer such that:  
 $\max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_r} \|\mathbf{A}\mathbf{v}\| = 0$ . Then  $A$  has  $r$  singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ .
- The vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  are more specifically called the **right singular vectors**.
- For any singular vector  $\mathbf{v}_i$ ,  $\sigma_i = \|\mathbf{A}\mathbf{v}_i\|$  may be interpreted as the **component** of the matrix  $A$  along  $\mathbf{v}_i$ .
- Given this interpretation, the “*the components should add up to give the whole content of A*”.
- For any row  $a_j$  in the matrix  $A$ , we can write  $\|a_j\|^2 = \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2$ . This further gives:

$$\sum_{j=1}^n \|a_j\|^2 = \sum_{j=1}^n \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2 = \sum_{i=1}^r \|\mathbf{A}\mathbf{v}_i\|^2 = \sum_{i=1}^r \sigma_i^2.$$

- The LHS of the above equation may be interpreted as “*content of the matrix*” defines the **Frobenius Norm** of the matrix  $A$ .

### Definition (Frobenius Norm)

The Frobenius norm of a given  $n \times d$  matrix  $A$ , denoted by  $\|A\|_F$ , is

defined as:  $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^d A_{i,j}^2}$ .

# Best Fit Subspaces and SVD

## Frobenius Norm

- For any row  $a_j$  in the matrix  $A$ , we can write  $\|a_j\|^2 = \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2$ . This further gives:

$$\sum_{j=1}^n \|a_j\|^2 = \sum_{j=1}^n \sum_{i=1}^r (a_j \cdot \mathbf{v}_i)^2 = \sum_{i=1}^r \|A\mathbf{v}_i\|^2 = \sum_{i=1}^r \sigma_i^2.$$

- The LHS of the above equation may be interpreted as “*content of the matrix*” defines the **Frobenius Norm** of the matrix  $A$ .

### Definition (Frobenius Norm)

The Frobenius norm of a given  $n \times d$  matrix  $A$ , denoted by  $\|A\|_F$ , is defined as:  $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^d A_{i,j}^2}$ .

### Theorem

*For any matrix  $A$ , the sum of squares of the right singular values equals the square of the Frobenius norm of the matrix.*

# Singular Value Decomposition (SVD)

## Left singular vectors

- Let  $\mathbf{v}_1, \dots, \mathbf{v}_r$  be the right singular vectors and  $\sigma_1, \dots, \sigma_r$  be the corresponding singular values of matrix  $A$ .
- The **left singular vectors** are defined as  $\mathbf{u}_i = \frac{1}{\sigma_i} A \mathbf{v}_i$ .
- $\sigma_i \mathbf{u}_i$  may be interpreted as a vector whose components are the projections of the rows of  $A$  onto  $\mathbf{v}_i$ .

# Singular Value Decomposition (SVD)

## Left singular vectors

- Let  $\mathbf{v}_1, \dots, \mathbf{v}_r$  be the right singular vectors and  $\sigma_1, \dots, \sigma_r$  be the corresponding eigenvalues of matrix  $A$ .
- The **left singular vectors** are defined as  $\mathbf{u}_i = \frac{1}{\sigma_i} A \mathbf{v}_i$ .
- $\sigma_i \mathbf{u}_i$  may be interpreted as a vector whose components are the projections of the rows of  $A$  onto  $\mathbf{v}_i$ .
- $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$  is a rank one matrix whose rows can be interpreted as component of rows of  $A$  along  $\mathbf{v}_i$ .
- Given this, the following decomposition of  $A$  into rank one matrices should make sense (we will prove this):  $A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ .

### Theorem

*Let  $A$  be any  $n \times d$  matrix with right singular vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ , left-singular vectors  $\mathbf{u}_1, \dots, \mathbf{u}_r$ , and corresponding singular values  $\sigma_1, \dots, \sigma_r$ . Then  $A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ .*

End