23. Singular Value Decomposition (SVD) and eigen-decomposition

Last time

- Least squares
- Optimization
- Solution by linear algebra
- Pseudo-inverse

Goals for today

- Action of a matrix
- Eigenvector decomposition
- Singular Value Decomposition (SVD)

Action of a matrix

- Consider arbitrary $(m \times n)$ matrix A

Action of a matrix

- Consider arbitrary $(m \times n)$ matrix A
- Study its action on unit sphere $\mathbb{S}_n$ in $\mathbb{R}^n$
- $\mathbb{S}_n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}$

Action of a matrix

- Consider arbitrary $(m \times n)$ matrix A
- Study its action on unit sphere $\mathbb{S}_n$ in $\mathbb{R}^n$
- $\mathbb{S}_n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}$

- Visualization suggests $A(\mathbb{S}_n)$ is **ellipsoid** (hyper-ellipse)
- Stretched and rotated sphere

Action of a matrix

- Consider arbitrary $(m \times n)$ matrix A
- Study its action on unit sphere $\mathbb{S}_n$ in $\mathbb{R}^n$
- $\mathbb{S}_n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}$

- Visualization suggests $A(\mathbb{S}_n)$ is **ellipsoid** (hyper-ellipse)
- Stretched and rotated sphere

- Key quantities:
  - Lengths σ_i of semi-axes – amount of stretch
  - Directions $\mathbf{u}_i$ of stretches

Action of a matrix II

- Let's try to understand how this comes about:

Action of a matrix II

- Let's try to understand how this comes about:
- Suppose $\mathbf{x} \in \mathbb{S}_n$; define $\mathbf{y} := \mathrm{A}\mathbf{x}$ to be its image

Action of a matrix II

- Let's try to understand how this comes about:
- Suppose $\mathbf{x} \in \mathbb{S}_n$; define $\mathbf{y} := \mathsf{A}\mathbf{x}$ to be its image
- Suppose that A square + invertible for now

Action of a matrix II

- Let's try to understand how this comes about:
- Suppose $\mathbf{x} \in \mathbb{S}_n$; define $\mathbf{y} := \mathsf{A}\mathbf{x}$ to be its image
- Suppose that A square + invertible for now
- Then $\|\mathsf{A}^{-1}\mathbf{y}\|^2 = 1$

Action of a matrix II

- Let's try to understand how this comes about:
- Suppose $\mathbf{x} \in \mathbb{S}_n$; define $\mathbf{y} := \mathsf{A}\mathbf{x}$ to be its image
- Suppose that A square + invertible for now
- Then $\|\mathsf{A}^{-1}\mathbf{y}\|^2 = 1$
- Hence

$$\mathbf{y}^\top \mathsf{S}\mathbf{y} = 1$$

where $\mathsf{S} := (\mathsf{A}^{-1})^\top(\mathsf{A}^{-1})$ is symmetric + positive definite

Eigenvalues

- Recall: square $m \times m$ matrix M has **eigenvector v** $\neq$ **0** with **eigenvalue** $\lambda \in \mathbb{C}$ if

$$\mathrm{M}\mathbf{v} = \lambda \mathbf{v}$$

Eigenvalues

- Recall: square $m \times m$ matrix M has **eigenvector v** $\neq$ **0** with **eigenvalue** $\lambda \in \mathbb{C}$ if

$$\mathrm{M}\mathbf{v} = \lambda \mathbf{v}$$

- Geometrical meaning: Direction $\hat{\mathbf{v}}$ is **invariant** (unchanged) under A
- *Stretched* (and possibly reflected) by factor λ

Spectral theorem

- **Spectrum** of matrix: set of eigenvalues

Spectral theorem

- **Spectrum** of matrix: set of eigenvalues
- Eigenvalues satsify $\det(\mathsf{A} - \lambda \mathsf{I}) = 0$ – **characteristic polynomial**
- Fundamental thm of algebra $\implies$ there are m eigenvalues
- For a real, **symmetric** matrix S, *exercise*:
  - eigenvalues λ_i are **real**
  - eigenvectors $\mathbf{v}_i$ with *distinct* λ_i are **orthogonal**

Spectral theorem

- **Spectrum** of matrix: set of eigenvalues

- Eigenvalues satsify $\det(\mathsf{A} - \lambda \mathsf{I}) = 0$ – **characteristic polynomial**
- Fundamental thm of algebra $\implies$ there are m eigenvalues
- For a real, **symmetric** matrix S, *exercise*:
  - eigenvalues λ_i are **real**
  - eigenvectors $\mathbf{v}_i$ with *distinct* λ_i are **orthogonal**
- "Hence" **spectral theorem**: real symmetric matrices have a basis of orthonormal eigenvectors

Spectral theorem

- **Spectrum** of matrix: set of eigenvalues

- Eigenvalues satsify $\det(\mathsf{A} - \lambda \mathsf{I}) = 0$ – **characteristic polynomial**
- Fundamental thm of algebra $\implies$ there are m eigenvalues
- For a real, **symmetric** matrix S, *exercise*:
  - eigenvalues λ_i are **real**
  - eigenvectors $\mathbf{v}_i$ with *distinct* λ_i are **orthogonal**
- "Hence" **spectral theorem**: real symmetric matrices have a basis of orthonormal eigenvectors

Eigen-decomposition

- Have $\mathsf{SQ} = \mathsf{Q}\Lambda$, where

$$\mathsf{Q} := \left(\mathbf{v}_1 \middle| \mathbf{v}_2 \middle| \cdots \middle| \mathbf{v}_n\right)$$

$$\Lambda := \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

Eigen-decomposition

- Have $\mathsf{S}\mathsf{Q} = \mathsf{Q}\Lambda$, where

$$\mathsf{Q} := \left(\mathbf{v}_1 | \mathbf{v}_2 | \cdots | \mathbf{v}_n\right)$$

$$\Lambda := \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

- Hence $\mathsf{S} = \mathsf{Q}\Lambda\mathsf{Q}^\top$

Action of a matrix III

- So $\mathbf{y}^\top \mathrm{Q}\Lambda\mathrm{Q}^\top\mathbf{y} = 1$

Action of a matrix III

- So $\mathbf{y}^\top \mathrm{Q}\Lambda\mathrm{Q}^\top \mathbf{y} = 1$
- Hence $\mathbf{z}^\top \Lambda \mathbf{z} = 1$

  where $\mathbf{z} := \mathrm{Q}^\top y$

Action of a matrix III

- So $\mathbf{y}^\top \mathrm{Q}\Lambda\mathrm{Q}^\top\mathbf{y} = 1$
- Hence $\mathbf{z}^\top\Lambda\mathbf{z} = 1$

  where $\mathbf{z} := \mathrm{Q}^\top y$
- So $\lambda_1 z_1^2 + \cdots + \lambda_n z_n^2 = 1$

Action of a matrix III

- So $\mathbf{y}^\top \mathsf{Q}\Lambda\mathsf{Q}^\top\mathbf{y} = 1$
- Hence $\mathbf{z}^\top\Lambda\mathbf{z} = 1$

  where $\mathbf{z} := \mathsf{Q}^\top y$
- So $\lambda_1 z_1^2 + \cdots + \lambda_n z_n^2 = 1$

- Can show $(\mathsf{A}^\top\mathsf{A})^{-1}$ has eigenvalues $\lambda_i > 0$ (exercise)

Action of a matrix III

- So $\mathbf{y}^\top \mathsf{Q} \Lambda \mathsf{Q}^\top \mathbf{y} = 1$
- Hence $\mathbf{z}^\top \Lambda \mathbf{z} = 1$

  where $\mathbf{z} := \mathsf{Q}^\top y$
- So $\lambda_1 z_1^2 + \cdots + \lambda_n z_n^2 = 1$

- Can show $(\mathsf{A}^\top \mathsf{A})^{-1}$ has eigenvalues $\lambda_i > 0$ (exercise)
- Hence $\mathbf{z}$ lies on ellipse

  with semi-axis lengths $\sigma_i := \sqrt{\lambda_i}$
- $\mathbf{y}$ lies on rotated ellipse

Singular values and singular vectors

- σ_i are **singular values** of A [terrible name]

Singular values and singular vectors

- σ_i are **singular values** of A [terrible name]
- Directions $\mathbf{u}_i$, columns of Q, are **left singular vectors**
  - directions in which ellipse is stretched
- $\mathbf{v}_i$ such that A$\mathbf{v}_i = \sigma_i \mathbf{u}_i$ are **right singular vectors**

Singular-value decomposition (SVD)

- *Any* $(m \times n)$ matrix A has an SVD:

$$\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^\top$$

  - columns of U are $\mathbf{u}_i$
  - columns V are $\mathbf{v}_i$
  - Σ is diagonal matrix with σ_i on diagonal

Singular-value decomposition (SVD)

- *Any* $(m \times n)$ matrix A has an SVD:

$$\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^\top$$

  - columns of U are $\mathbf{u}_i$
  - columns V are $\mathbf{v}_i$
  - Σ is diagonal matrix with σ_i on diagonal
- Geometrical interpretation:

Calculating eigenvalues

- As usual, still don't know how to actually *calculate* anything!

Calculating eigenvalues

- As usual, still don't know how to actually *calculate* anything!
- Need to calculate eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of symmetric matrix S

Calculating eigenvalues

- As usual, still don't know how to actually *calculate* anything!
- Need to calculate eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of symmetric matrix S
- Note: there are direct methods for the singular values σ_i that avoid g $\mathsf{A}^\top \mathsf{A}$

Power iteration method

- Simple method to get eigenvector with largest eigenvalue, **leading** eigenvector

Power iteration method

- Simple method to get eigenvector with largest eigenvalue, **leading** eigenvector
- Start with any initial $\mathbf{v} \neq \mathbf{0}$

Power iteration method

- Simple method to get eigenvector with largest eigenvalue, **leading** eigenvector
- Start with any initial $\mathbf{v} \neq \mathbf{0}$
- Apply A repeatedly to form $\mathrm{A}^n\mathbf{v}$
- In general will explode, so normalize
- Converges to leading eigenvector!

Power iteration method

- Simple method to get eigenvector with largest eigenvalue, **leading** eigenvector
- Start with any initial $\mathbf{v} \neq \mathbf{0}$
- Apply A repeatedly to form $\mathsf{A}^n\mathbf{v}$
- In general will explode, so normalize
- Converges to leading eigenvector!
- To see this expand $\mathbf{v}$ in basis of eigenvectors

Simultaneous power method

- Suppose start with several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$
- If iterate each separately, all will converge to leading eigenvector

Simultaneous power method

- Suppose start with several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$
- If iterate each separately, all will converge to leading eigenvector
- Solution: Orthogonalize using Gram–Schmidt

Simultaneous power method

- Suppose start with several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$
- If iterate each separately, all will converge to leading eigenvector

- Solution: Orthogonalize using Gram–Schmidt
- Converge to leading k eigenvectors

Simultaneous power method

- Suppose start with several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$
- If iterate each separately, all will converge to leading eigenvector
- Solution: Orthogonalize using Gram–Schmidt
- Converge to leading k eigenvectors
- This is part of QR algorithm:
  1. Factorize $\mathsf{QR} = \mathsf{A}$
  2. Set $\mathsf{A} \leftarrow \mathsf{RQ}$
  3. Repeat

Simultaneous power method

- Suppose start with several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$
- If iterate each separately, all will converge to leading eigenvector

- Solution: Orthogonalize using Gram–Schmidt
- Converge to leading k eigenvectors
- This is part of QR algorithm:
  1. Factorize $\mathsf{QR} = \mathsf{A}$
  2. Set $\mathsf{A} \leftarrow \mathsf{RQ}$
  3. Repeat
- A converges to diagonal matrix of eigenvalues!

Rank of a matrix

- Number of non-zero singular values = column rank = rank

Rank of a matrix

- Number of non-zero singular values = column rank = rank
- SVD gives **best approximation** to A by matrix of rank ka:

$$\mathsf{A}_k := \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Rank of a matrix

- Number of non-zero singular values = column rank = rank
- SVD gives **best approximation** to A by matrix of rank ka:

$$\mathsf{A}_k := \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

- Use to find low-rank approximation of collection of data – PCA (Principal Component Analysis in statistics)
- E.g. image compression

Summary

- SVD gives matrix as rotation + stretch + rotation
- Closely related to eigendecomposition of $A^\top A$

Summary

- SVD gives matrix as rotation + stretch + rotation
- Closely related to eigendecomposition of $A^\top A$
- Applications to data summarization compression