More on PCA

Yingzhen Li

Department of Computing
Imperial College London

@liyzhen2
yingzhen.li@imperial.ac.uk

Nov 5, 2021

Recap

PCA for dimensionality reduction:

- Data: $\mathcal{D} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}, \boldsymbol{x}_n \in \mathbb{R}^{D \times 1}$ s.t. $\text{mean}(\boldsymbol{x}_n) = \mathbf{0}$
- Find projections in a **lower-dimensional** space:

$$\boldsymbol{x}_n \approx \tilde{\boldsymbol{x}}_n := \sum_{j=1}^{M} z_{nj} \mathbf{b}_j, \quad z_{nj} := \mathbf{b}_j^\top \boldsymbol{x}_n$$

using an **orthonormal basis**

$$\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_M], \quad \mathbf{b}_m \in \mathbb{R}^{D \times 1}, \quad M < D$$

Recap

PCA for dimensionality reduction:

- Data: $\mathcal{D} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}, \boldsymbol{x}_n \in \mathbb{R}^{D\times 1}$ s.t. $\text{mean}(\boldsymbol{x}_n) = \mathbf{0}$
- Find projections in a **lower-dimensional** space:

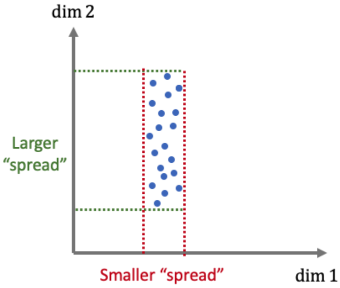
$$\boldsymbol{x}_n \approx \tilde{\boldsymbol{x}}_n := \sum_{j=1}^{M} z_{nj}\mathbf{b}_j, \quad z_{nj} := \mathbf{b}_j^\top \boldsymbol{x}_n$$

using an **orthonormal basis**

$$\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_M], \quad \mathbf{b}_m \in \mathbb{R}^{D\times 1}, \quad M < D$$

PCA: maximum variance perspective

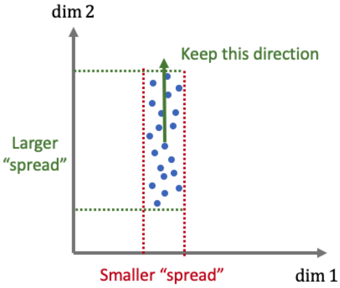
The “maximum variance” intuition of PCA:
project onto directions where the datapoints “vary the most”



“Spread” is defined as the variance along a given direction

PCA: maximum variance perspective

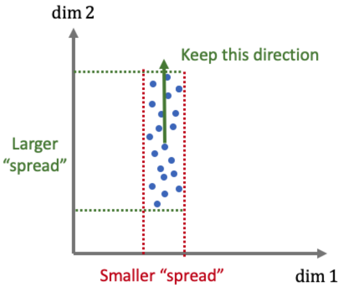
The “maximum variance” intuition of PCA:
project onto directions where the datapoints “vary the most”

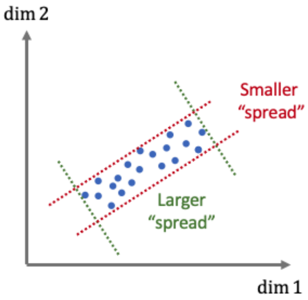


“Spread” is defined as the variance along a given direction

PCA: maximum variance perspective

The “maximum variance” intuition of PCA:
project onto directions where the datapoints “vary the most”

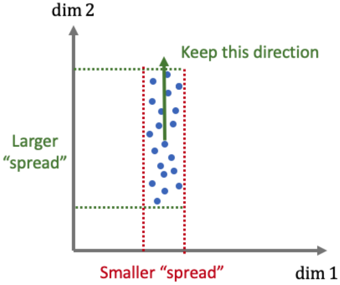


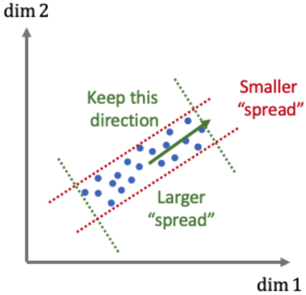


“Spread” is defined as the variance along a given direction

PCA: maximum variance perspective

The “maximum variance” intuition of PCA:
project onto directions where the datapoints “vary the most”





“Spread” is defined as the variance along a given direction

PCA: maximum variance perspective

Recall the problem set-up:

- Data: $\mathcal{D} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^{D \times 1}$ s.t. $\text{mean}(\boldsymbol{x}_n) = \mathbf{0}$
- Find projections in a **lower-dimensional** space:

$$\boldsymbol{z}_n := \mathbf{B}^\top \boldsymbol{x}_n \quad \Leftrightarrow \quad z_{nj} := \mathbf{b}_j^\top \boldsymbol{x}_n$$

  using an **orthonormal basis**

$$\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_M], \quad \mathbf{b}_m \in \mathbb{R}^{D \times 1}, \quad M < D$$

- Solve for $\mathbf{b}_1$ such that

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] \quad \text{is maximised}$$

PCA: maximum variance perspective

Solve for $\mathbf{b}_1$ such that

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] \quad \text{is maximised, subject to } ||\mathbf{b}_1||_2 = 1$$

PCA: maximum variance perspective

Solve for $\mathbf{b}_1$ such that

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] \quad \text{is maximised, subject to } ||\mathbf{b}_1||_2 = 1$$

- Variance after projection (recall that $\boldsymbol{x}_n$ has mean zero):

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] := \frac{1}{N}\sum_{n=1}^{N}(\mathbf{b}_1^\top \boldsymbol{x}_n)^2 = \mathbf{b}_1^\top (\underbrace{\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top}_{=\mathbf{S}=\mathbf{Q}\Lambda\mathbf{Q}^\top})\mathbf{b}_1$$

$$= \mathbf{b}_1^\top \mathbf{Q}\Lambda\mathbf{Q}^\top \mathbf{b}_1 = \sum_{d=1}^{D} \lambda_d \beta_{1d}^2$$

PCA: maximum variance perspective

Solve for $\mathbf{b}_1$ such that

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] \quad \text{is maximised, subject to } ||\mathbf{b}_1||_2 = 1$$

- Variance after projection (recall that $\boldsymbol{x}_n$ has mean zero):

$$\mathbb{V}[\mathbf{b}_1^\top \boldsymbol{x}_n] := \frac{1}{N}\sum_{n=1}^{N}(\mathbf{b}_1^\top \boldsymbol{x}_n)^2 = \mathbf{b}_1^\top \underbrace{\left(\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^\top\right)}_{=\mathbf{S}=\mathbf{Q}\Lambda\mathbf{Q}^\top}\mathbf{b}_1$$

$$= \mathbf{b}_1^\top \mathbf{Q}\Lambda\mathbf{Q}^\top \mathbf{b}_1 = \sum_{d=1}^{D}\lambda_d \beta_{1d}^2$$

- **Solution: $\mathbf{b}_1 = q_1$ (the eigenvector with the largest eigenvalue)**

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- Compute the "remainder" of projection:

$$\hat{\boldsymbol{x}}_n = \boldsymbol{x}_n - \sum_{j=1}^{m-1} \boldsymbol{z}_{nj}\mathbf{b}_j = \boldsymbol{x}_n - \sum_{j=1}^{m-1} (\mathbf{b}_j^\top \boldsymbol{x}_n)\mathbf{b}_j$$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- Compute the "remainder" of projection:

$$\hat{\boldsymbol{x}}_n = \boldsymbol{x}_n - \sum_{j=1}^{m-1} \boldsymbol{z}_{nj}\mathbf{b}_j = \boldsymbol{x}_n - \sum_{j=1}^{m-1} (\mathbf{b}_j^\top \boldsymbol{x}_n)\mathbf{b}_j$$

- maximise $\mathbb{V}[\mathbf{b}_m^\top \hat{\boldsymbol{x}}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- Compute the "remainder" of projection:

$$\hat{\boldsymbol{x}}_n = \boldsymbol{x}_n - \sum_{j=1}^{m-1} z_{nj}\mathbf{b}_j = \boldsymbol{x}_n - \sum_{j=1}^{m-1} (\mathbf{b}_j^\top \boldsymbol{x}_n)\mathbf{b}_j$$

- maximise $\mathbb{V}[\mathbf{b}_m^\top \hat{\boldsymbol{x}}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Variance after projection (recall that $\boldsymbol{x}_n$ has mean zero):

$$\mathbb{V}[\mathbf{b}_m^\top \hat{\boldsymbol{x}}_n] := \frac{1}{N}\sum_{n=1}^{N}(\mathbf{b}_m^\top \boldsymbol{x}_n - \sum_{j=1}^{m-1}(\mathbf{b}_j^\top \boldsymbol{x}_n)\underbrace{\mathbf{b}_j^\top \mathbf{b}_m}_{=0})^2 = \mathbf{b}_m^\top(\underbrace{\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^\top}_{=\mathbf{S}})\mathbf{b}_m$$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^D \beta_{mj}\boldsymbol{q}_j$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^{D} \beta_{mj} \boldsymbol{q}_j$

   2b. $||\mathbf{b}_m||_2 = 1 \quad \Rightarrow \quad \sum_{j=m}^{D} \beta_{mj}^2 = 1$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^{D} \beta_{mj}\boldsymbol{q}_j$

   2b. $||\mathbf{b}_m||_2 = 1 \quad \Rightarrow \quad \sum_{j=m}^{D} \beta_{mj}^2 = 1$

   2c. Solve for maximum of $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$ w.r.t. β_{mj}:

$$\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n] = (\sum_{j=m}^{D} \beta_{mj}\boldsymbol{q}_j)^\top \mathbf{S} (\sum_{j=m}^{D} \beta_{mj}\boldsymbol{q}_j)$$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^{D} \beta_{mj} \boldsymbol{q}_j$

   2b. $||\mathbf{b}_m||_2 = 1 \quad \Rightarrow \quad \sum_{j=m}^{D} \beta_{mj}^2 = 1$

   2c. Solve for maximum of $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$ w.r.t. β_{mj}:

$$\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n] = \sum_{d=m}^{D} \lambda_d \beta_{md}^2$$

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^{D} \beta_{mj} \boldsymbol{q}_j$

   2b. $||\mathbf{b}_m||_2 = 1 \quad \Rightarrow \quad \sum_{j=m}^{D} \beta_{mj}^2 = 1$

   2c. Solve for maximum of $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$ w.r.t. β_{mj}:

   Solution: $\mathbf{b}_m = \boldsymbol{q}_m$ **(i.e.** $\beta_{mm} = 1, \beta_{mj} = 0, j > m$**)**

PCA: maximum variance perspective

Iteratively solve for the rest of the directions $\mathbf{b}_2, ...\mathbf{b}_M$:

For $m = 2, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

Recall that $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$:

1. $\mathbf{b}_1 = \boldsymbol{q}_1$
2. For $m = 2, ..., M$, assume $\mathbf{b}_j = \boldsymbol{q}_j, j = 1, ..., m-1$

   2a. $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1 \quad \Rightarrow \quad \mathbf{b}_m = \sum_{j=m}^{D} \beta_{mj} \boldsymbol{q}_j$

   2b. $||\mathbf{b}_m||_2 = 1 \quad \Rightarrow \quad \sum_{j=m}^{D} \beta_{mj}^2 = 1$

   2c. Solve for maximum of $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$ w.r.t. β_{mj}:

   Solution: $\mathbf{b}_m = \boldsymbol{q}_m$ (i.e. $\beta_{mm} = 1, \beta_{mj} = 0, j > m$)

3. **"Proof by induction": $\mathbf{b}_m = \boldsymbol{q}_m, m = 1, ..., M$**

PCA: maximum variance perspective

For $m = 1, ..., M$:

- maximise $\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$, subject to $||\mathbf{b}_m||_2 = 1$, $\mathbf{b}_m \perp \mathbf{b}_j, j = 1, ..., m-1$

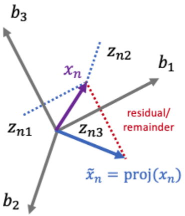
Solutions: $\mathbf{b}_m = \boldsymbol{q}_m$ for $m = 1, ..., M$

$\Rightarrow$ Projecting $\boldsymbol{x}_n$ to a subspace

$$span(\{\boldsymbol{q}_m\}_{m=1}^M) = span(\{\boldsymbol{q}_j\}_{j=M+1}^D)^\perp$$

$$\boldsymbol{x}_n = \underbrace{\sum_{j=1}^{M} z_{nj}\boldsymbol{q}_j}_{:=\tilde{\boldsymbol{x}}_n} + \underbrace{\sum_{j=M+1}^{D} z_{nj}\mathbf{b}_j}_{\text{dropped}}, \quad \mathbf{b}_i \perp \boldsymbol{q}_j$$

$$\tilde{\boldsymbol{x}}_n \in span(\{\boldsymbol{q}_m\}_{m=1}^M)$$



$b_i \perp b_j, ||b_i||_2 = 1$

$span(\{b_1, b_2, b_3\}) = R^3$

PCA: comparing both views

- Minimum reconstruction error view:

$$\mathbf{B}^*_{full} = \{\mathbf{b}_1, ..., \mathbf{b}_M, \boldsymbol{q}_{M+1}, ..., \boldsymbol{q}_D\}, \quad \mathbf{b}_i \perp \mathbf{b}_j, \mathbf{b}_i \perp \boldsymbol{q}_j$$

- Maximum variance view:

$$\mathbf{B}^*_{full} = \{\boldsymbol{q}_1, ..., \boldsymbol{q}_M, \mathbf{b}_{M+1}, ..., \mathbf{b}_D\}, \quad \mathbf{b}_i \perp \mathbf{b}_j, \mathbf{b}_i \perp \boldsymbol{q}_j$$

PCA: comparing both views

- Minimum reconstruction error view:

$$\mathbf{B}^*_{full} = \{\mathbf{b}_1, ..., \mathbf{b}_M, \boldsymbol{q}_{M+1}, ..., \boldsymbol{q}_D\}, \quad \mathbf{b}_i \perp \mathbf{b}_j, \mathbf{b}_i \perp \boldsymbol{q}_j$$

- Maximum variance view:

$$\mathbf{B}^*_{full} = \{\boldsymbol{q}_1, ..., \boldsymbol{q}_M, \mathbf{b}_{M+1}, ..., \mathbf{b}_D\}, \quad \mathbf{b}_i \perp \mathbf{b}_j, \mathbf{b}_i \perp \boldsymbol{q}_j$$

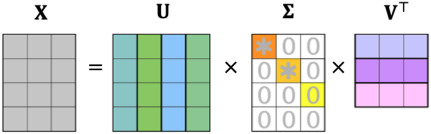
- By convention we often use $\mathbf{B}^*_{full} = \mathbf{Q}$

Singular value decomposition

Key computations for PCA: eigen decomposition of the variance

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top = \frac{1}{N}\mathbf{X}\mathbf{X}^\top, \quad \mathbf{X} := [\boldsymbol{x}_1, ..., \boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$$

Finding eigendecomposition of **S** via SVD of **X**!



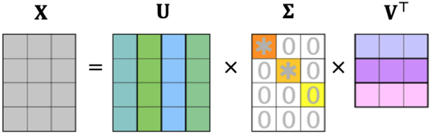
(× represents matrix product)

Singular value decomposition

Key computations for PCA: eigen decomposition of the variance

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top = \frac{1}{N}\mathbf{X}\mathbf{X}^\top, \quad \mathbf{X} := [\boldsymbol{x}_1, ..., \boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$$

Finding eigendecomposition of **S** via SVD of **X**!

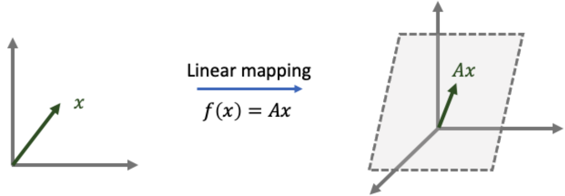


(× represents matrix product)

- $\mathbf{U} = [\boldsymbol{u}_1, ..., \boldsymbol{u}_D] \in \mathbb{R}^{D\times D}$: orthonormal basis of $\mathbb{R}^D$
- $\mathbf{V} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_N] \in \mathbb{R}^{N\times N}$: orthonormal basis of $\mathbb{R}^N$
- Σ contains a diagonal block with singular values of **X**: $\text{diag}(\sigma_1, ..., \sigma_r), \sigma_i > 0$:
- $r = \text{rank}(\mathbf{X}) \leqslant min(D, N)$: rank of **X**

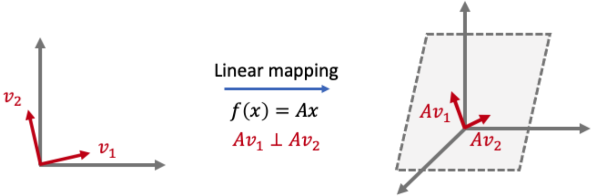
Singular value decomposition

Intuition of SVD: Assume $\mathbf{A} \in \mathbb{R}^{N \times D}$



Singular value decomposition

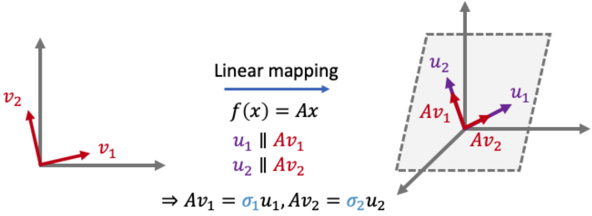
Intuition of SVD: Assume $\mathbf{A} \in \mathbb{R}^{N \times D}$



- There exist an orthonormal basis $\{v_d\}_{d=1}^D$ of $\mathbb{R}^D$ such that $\langle \mathbf{A}v_i, \mathbf{A}v_j \rangle = 0, \forall i \neq j$

Singular value decomposition

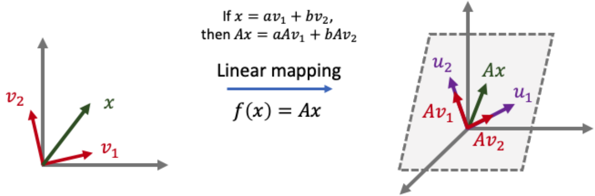
Intuition of SVD: Assume $\mathbf{A} \in \mathbb{R}^{N \times D}$



- There exist an orthonormal basis $\{\boldsymbol{v}_d\}_{d=1}^D$ of $\mathbb{R}^D$ such that $\langle \mathbf{A}\boldsymbol{v}_i, \mathbf{A}\boldsymbol{v}_j \rangle = 0, \forall i \neq j$
- By rescaling $\{\mathbf{A}\boldsymbol{v}_i\}$ vectors and adding more orthogonal vectors, we can construct another orthonormal basis $\{\boldsymbol{u}_n\}_{n=1}^N$ of $\mathbb{R}^N$

Singular value decomposition

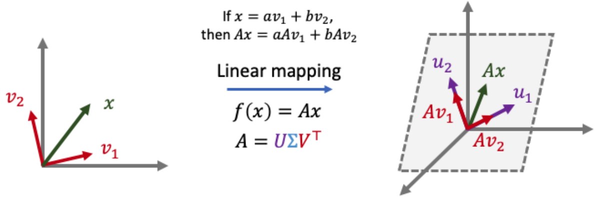
Intuition of SVD: Assume $\mathbf{A} \in \mathbb{R}^{N \times D}$



- There exist an orthonormal basis $\{\boldsymbol{v}_d\}_{d=1}^{D}$ of $\mathbb{R}^D$ such that $\langle \mathbf{A}\boldsymbol{v}_i, \mathbf{A}\boldsymbol{v}_j \rangle = 0, \forall i \neq j$
- By rescaling $\{\mathbf{A}\boldsymbol{v}_i\}$ vectors and adding more orthogonal vectors, we can construct another orthonormal basis $\{\boldsymbol{u}_n\}_{n=1}^{N}$ of $\mathbb{R}^N$
- Then $\mathbf{A}\boldsymbol{x}$ is constructed as a linear combination of $\{\mathbf{A}\boldsymbol{v}_i\}$ vectors

Singular value decomposition

Intuition of SVD: Assume $\mathbf{A} \in \mathbb{R}^{N \times D}$



Computing $\mathbf{A}\boldsymbol{x}$ with $\boldsymbol{x} \in \mathbb{R}^D$ represented using $\{\boldsymbol{e}_d\}_{d=1}^D$ basis:

- $\mathbf{V}^\top \boldsymbol{x}$: compute the coordinates of $\boldsymbol{x}$ in $\{\boldsymbol{v}_d\}_{d=1}^D$ basis
- $\Sigma\mathbf{V}^\top \boldsymbol{x}$: stretch/shrink the vector
- $\mathbf{U}\Sigma\mathbf{V}^\top \boldsymbol{x}$: map to the output space $\mathbb{R}^N$ whose basis is $\{\boldsymbol{u}_n\}_{n=1}^N$, and compute the coordinates back in $\{\boldsymbol{e}_n\}_{n=1}^N$ basis

Singular value decomposition

A "decomposition" of $\mathbf{S}$ derived from SVD:

$$\mathbf{S} = \frac{1}{N}\mathbf{X}\mathbf{X}^\top = \frac{1}{N}\mathbf{U}\Sigma\underbrace{\mathbf{V}^\top\mathbf{V}}_{=\mathbf{I}_{N\times N}}\Sigma^\top\mathbf{U}^\top = \mathbf{U}\text{diag}(\frac{\sigma_1^2}{N}, ..., \frac{\sigma_r^2}{N}, \underbrace{0, ..., 0}_{D-r})\mathbf{U}^\top$$

Singular value decomposition

A "decomposition" of $\mathbf{S}$ derived from SVD:

$$\mathbf{S} = \frac{1}{N}\mathbf{X}\mathbf{X}^\top = \frac{1}{N}\mathbf{U}\Sigma \underbrace{\mathbf{V}^\top\mathbf{V}}_{=\mathbf{I}_{N\times N}} \Sigma^\top\mathbf{U}^\top = \mathbf{U}\mathrm{diag}(\frac{\sigma_1^2}{N}, ..., \frac{\sigma_r^2}{N}, \underbrace{0, ..., 0}_{D-r})\mathbf{U}^\top$$

Compare eigen decomposition and the decomposition from SVD:

$$\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top \quad \text{v.s} \quad \mathbf{S} = \mathbf{U}\mathrm{diag}(\frac{\sigma_1^2}{N}, ..., \frac{\sigma_r^2}{N}, \underbrace{0, ..., 0}_{D-r})\mathbf{U}^\top$$

Singular value decomposition

A "decomposition" of $\mathbf{S}$ derived from SVD:

$$\mathbf{S} = \frac{1}{N}\mathbf{X}\mathbf{X}^\top = \frac{1}{N}\mathbf{U}\Sigma \underbrace{\mathbf{V}^\top\mathbf{V}}_{=\mathbf{I}_{N\times N}} \Sigma^\top\mathbf{U}^\top = \mathbf{U}\text{diag}(\frac{\sigma_1^2}{N}, ..., \frac{\sigma_r^2}{N}, \underbrace{0, ..., 0}_{D-r})\mathbf{U}^\top$$

Compare eigen decomposition and the decomposition from SVD:

$$\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top \quad \text{v.s} \quad \mathbf{S} = \mathbf{U}\text{diag}(\frac{\sigma_1^2}{N}, ..., \frac{\sigma_r^2}{N}, \underbrace{0, ..., 0}_{D-r})\mathbf{U}^\top$$

- Returning eigen decomposition of $\mathbf{X}\mathbf{X}^\top$ via SVD of $\mathbf{X}$
- PCA requires the **U basis vectors with largest singular values**

Some home exercises

Some home exercises

- Read & understand eigenvalue decomposition & SVD
- Derive PCA solutions

Further aspects of PCA:

- constrained optimisation (later lectures)
- probabilistic view (MML book section 10.7)