COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Spring 2026.

Lecture 13

Logistics

- Problem Set 3 posted last night. Due next Friday 4/10 at 11:59pm.
- Given the timing of Midterm 2, Problem Set 4 might be 'split' into two halves, due before and after the midterm.
- Based on feedback, I am going to try to pick specific topics to do review and practice questions on during office hours. Today I will review eigenvectors and eigenvalues.
- You can still come to office hours with general questions unrelated to this topic, as usual.

Summary

Last Class:

$$\underset{\mathbb{R}^d}{\|x_i - x_j\|} = \underset{\mathbb{R}^k}{\|V^T x_i - V^T x_j\|}$$

- No-distortion embeddings for data lying in a k-dimensional subspace via an orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$ for that subspace.
- View as low-rank matrix factorization. Introduce concept of low-rank approximation.
- Idea of approximating a data matrix $\mathbf{X}$ with $\mathbf{X}\mathbf{V}\mathbf{V}^T$ when the data points lie close to the subspace spanned by $\mathbf{V}$'s columns.

Summary

Last Class:

- No-distortion embeddings for data lying in a k-dimensional subspace via an orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$ for that subspace.
- View as low-rank matrix factorization. Introduce concept of low-rank approximation.
- Idea of approximating a data matrix $\mathbf{X}$ with $\mathbf{X}\mathbf{V}\mathbf{V}^T$ when the data points lie close to the subspace spanned by $\mathbf{V}$'s columns.

This Class:

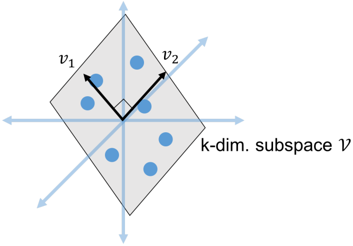
- 'Dual view' of low-rank approximation: data points that can be approximately reconstructed from a few basis vectors vs. linearly dependent features.
- Finding an optimal orthogonal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$ to minimize $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ when out data does not exactly lie in a low-dimensional subspace, via eigendecomposition.

Last Class: Embedding with Assumptions

Set Up: Assume that data points $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ lie in some k-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

$d = 3$

$k = 2$



Let $\vec{v}_1, \ldots, \vec{v}_k$ be an orthonormal basis for $\mathcal{V}$ and $\mathbf{V} \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns.

$\mathbb{R}^2$ $\mathbb{R}^3$

$$\|\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\vec{x}_j\|_2^2 = \|\vec{x}_i - \vec{x}_j\|_2^2.$$

Letting $\tilde{x}_i = \mathbf{V}^T\vec{x}_i$, we have a perfect embedding from $\mathcal{V}$ into $\mathbb{R}^k$.

Last Class: Low-Rank Factorization

Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie in a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as

$$\mathbf{X} = \mathbf{X}\mathbf{V}\mathbf{V}^T \text{ (implies } \textbf{rank}(\mathbf{X}) \leq k)$$

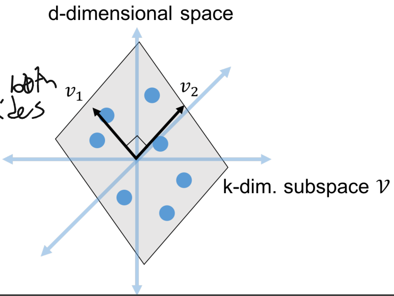
- $\mathbf{V}\mathbf{V}^T$ is a projection matrix, which projects the rows of $\mathbf{X}$ (the data points $\vec{x}_1, \ldots, \vec{x}_n$ onto the subspace $\mathcal{V}$.

$x_i^T V V^T = x_i^T$

$[x_i^T][VV^T] = [x_i]^T$

transpose both sides

$VV^T x_i = x_i$

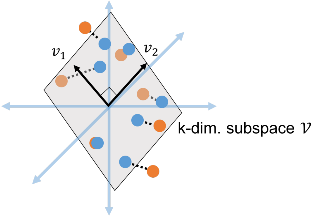


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation

Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T$$

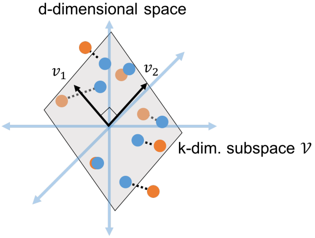


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation

Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T$$



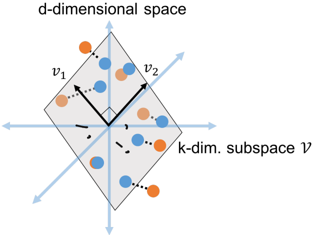
Note: $\mathbf{X}\mathbf{V}\mathbf{V}^T$ has rank k. It is a low-rank approximation of $\mathbf{X}$.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation

Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T$$



$$\|A\|_F^2 = \sum_{i,j} A_{ij}^2 = \sum_{i=1}^{n} \|a_i\|_2^2$$

Note: $\mathbf{X}\mathbf{V}\mathbf{V}^T$ has rank k. It is a low-rank approximation of $\mathbf{X}$.

$$\mathbf{X}\mathbf{V}\mathbf{V}^\mathsf{T} = \underset{\mathbf{B} \text{ with rows in } \mathcal{V}}{\arg\min} \|\mathbf{X} - \mathbf{B}\|_F^2 = \sum_{i,j} (\mathbf{X}_{i,j} - \mathbf{B}_{i,j})^2. = \sum_{i=1}^{n} \|x_i^T - x_i^T V V^T\|_2^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

This Class: Low-Rank Approximation

So Far: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T.$$

This is the closest approximation to $\mathbf{X}$ with rows in $\mathcal{V}$ (i.e., in the column span of $\mathbf{V}$).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

This Class: Low-Rank Approximation

So Far: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T.$$

This is the closest approximation to $\mathbf{X}$ with rows in $\mathcal{V}$ (i.e., in the column span of $\mathbf{V}$).

- Letting $\mathbf{V}\mathbf{V}^T\vec{x}_i, \mathbf{V}\mathbf{V}^T\vec{x}_j$ be the i^{th} and j^{th} projected data points,

$$\|\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}\mathbf{V}^T\vec{x}_j\|_2 = \|\mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_j\|_2. = \|\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\vec{x}_j\|_2.$$

$\mathbb{R}^d$ I $\mathbb{R}^k$

are in $\mathcal{V}$

no distortion embedding for points in subspace

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

This Class: Low-Rank Approximation

So Far: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T.$$

This is the closest approximation to $\mathbf{X}$ with rows in $\mathcal{V}$ (i.e., in the column span of $\mathbf{V}$).

- Letting $\mathbf{V}\mathbf{V}^T\vec{x}_i$, $\mathbf{V}\mathbf{V}^T\vec{x}_j$ be the i^{th} and j^{th} projected data points,

$$\|\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}\mathbf{V}^T\vec{x}_j\|_2 = \|\mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_j\|_2. = \|\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\vec{x}_j\|_2.$$

- I.e., we can use the rows of $\mathbf{X}\mathbf{V} \in \mathbb{R}^{n \times k}$ as a compressed approximate data set.

$x_1, \ldots x_n \in \mathbb{R}^d$, compression

$VV^Tx_1, \ldots VV^Tx_n \in \mathbb{R}^d \iff V^Tx_1 \ldots V^Tx_n \in \mathbb{R}^k$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

This Class: Low-Rank Approximation

So Far: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T.$$

This is the closest approximation to $\mathbf{X}$ with rows in $\mathcal{V}$ (i.e., in the column span of $\mathbf{V}$).

- Letting $\mathbf{V}\mathbf{V}^T\vec{x}_i$, $\mathbf{V}\mathbf{V}^T\vec{x}_j$ be the i^{th} and j^{th} projected data points,

$$\|\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}\mathbf{V}^T\vec{x}_j\|_2 = \|\mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_j\|_2. = \|\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\vec{x}_j\|_2.$$

- I.e., we can use the rows of $\mathbf{X}\mathbf{V} \in \mathbb{R}^{n \times k}$ as a compressed approximate data set.

Key question is how to find the subspace $\mathcal{V}$ and correspondingly $\mathbf{V}$.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Quick Exercise 1 : Show that $\mathbf{V}\mathbf{V}^T$ is idempotent. I.e., $(\mathbf{V}\mathbf{V}^T)(\mathbf{V}\mathbf{V}^T)\vec{y} = (\mathbf{V}\mathbf{V}^T)\vec{y}$ for any $\vec{y} \in \mathbb{R}^d$.

$$V(V^T V)V^T y = VV^T y \quad (V^T V = I)$$

Quick Exercise 2: Show that $\mathbf{V}\mathbf{V}^T(\mathbf{I} - \mathbf{V}\mathbf{V}^T) = \mathbf{0}$ (the projection is orthogonal to its complement).

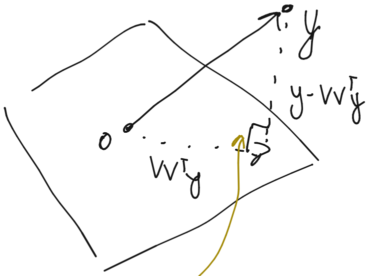
$$VV^T \cdot I - VV^T VV^T = VV^T - VV^T = 0$$

Pythagorean Theorem

Pythagorean Theorem: For any orthonormal $\mathbf{V} \in \mathbb{R}^{d \times k}$ and any $\vec{y} \in \mathbb{R}^d$,

$$\|\vec{y}\|_2^2 = \|(\mathbf{V}\mathbf{V}^T)\vec{y}\|_2^2 + \|\vec{y} - (\mathbf{V}\mathbf{V}^T)\vec{y}\|_2^2.$$

$\|(I-VV^T)y\|_2^2$ $\qquad$ $\|z\|_2^2 = z^T z$



$$\|y\|_2^2 = \|VV^Ty + (I-VV^T)y\|_2^2$$

$$(VV^Ty + (I-VV^T)y)^T(VV^Ty + (I-VV^T)y)$$

$$y^TVV^TVV^Ty + y^T(I-VV^T)VV^Ty + y^TVV^T(I-VV^T)y + y^T(I-VV^T)(I-VV^T)y$$

VV^Ty + (I-VV^T)y are orthogonal

$$y^TVV^T(I-VV^T)y = 0$$

$$\|y\|_2^2 = \|VV^Ty\|_2^2 + \|(I-VV^T)y\|_2^2$$

A Step Back: Why Low-Rank Approximation?

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

A Step Back: Why Low-Rank Approximation?

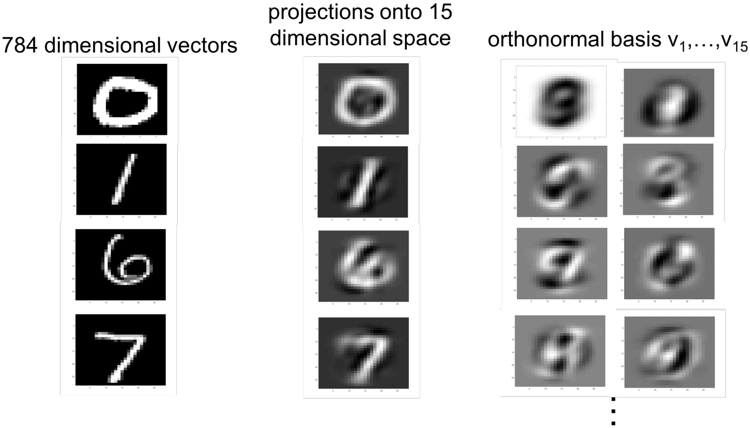
Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- The rows of $\mathbf{X}$ can be approximately reconstructed from a basis of k vectors.

A Step Back: Why Low-Rank Approximation?

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- The rows of $\mathbf{X}$ can be approximately reconstructed from a basis of k vectors.



Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- Equivalently, the columns of X are approx. spanned by k vectors.

$$X \approx XVV^T$$

rows of X are approx. spanned by rows of V^T

columns of X are approx. spanned by columns of XV

Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- Equivalently, the columns of $\mathbf{X}$ are approx. spanned by k vectors.

Linearly Dependent Variables:

	bedrooms	bathrooms	sq.ft.	floors	list price	sale price
home 1	2	2	1800	2	200,000	195,000
home 2	4	2.5	2700	1	300,000	310,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
home n	5	3.5	3600	3	450,000	450,000

Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- Equivalently, the columns of $\mathbf{X}$ are approx. spanned by k vectors.

Linearly Dependent Variables:

	bedrooms	bathrooms	sq.ft.	floors	list price	sale price
home 1	2	2	1800	2	200,000	195,000
home 2	4	2.5	2700	1	300,000	310,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
home n	5	3.5	3600	3	450,000	450,000

Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- Equivalently, the columns of $\mathbf{X}$ are approx. spanned by k vectors.

Linearly Dependent Variables:

	bedrooms	bathrooms	sq.ft.	floors	list price	sale price
home 1	2	2	1800	2	200,000	195,000
home 2	4	2.5	2700	1	300,000	310,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
home n	5	3.5	3600	3	450,000	450,000

Dual View of Low-Rank Approximation

Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a k-dimensional subspace?

- Equivalently, the columns of **X** are approx. spanned by k vectors.

Linearly Dependent Variables:

10000* bathrooms+ 10* (sq. ft.) ≈ list price

	bedrooms	bathrooms	sq.ft.	floors	list price	sale price
home 1	2	2	1800	2	200,000	195,000
home 2	4	2.5	2700	1	300,000	310,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
home n	5	3.5	3600	3	450,000	450,000

Best Fit Subspaces

Best Fit Subspace

If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Best Fit Subspace

If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

How do we find $\mathcal{V}$ (equivilantly $\mathbf{V}$)?

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Best Fit Subspace

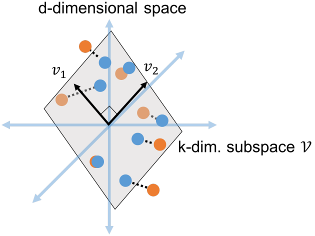
If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

How do we find $\mathcal{V}$ (equivilantly $\mathbf{V}$)?

$\|x_i\|_2^2 = \|V^T x_i\|_2^2 + \|x_i - VV^T x_i\|_2^2$

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\min} \|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2 = \sum_{i,j} (\mathbf{X}_{i,j} - (\mathbf{X}\mathbf{V}\mathbf{V}^T)_{i,j})^2 = \sum_{i=1}^{n} \|\vec{x}_i - \mathbf{V}\mathbf{V}^T \vec{x}_i\|_2^2$$

$\|x_i - VV^T x_i\|_2^2 = \|x_i\|_2^2 - \|V^T x_i\|_2^2$



$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

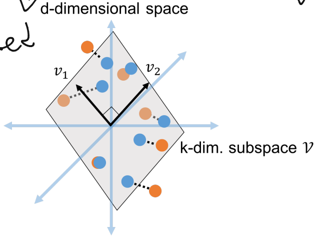
Best Fit Subspace

If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

How do we find $\mathcal{V}$ (equivilantly $\mathbf{V}$)?

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\min} \|\mathbf{X}\|_F^2 - \|\mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2 = \sum_{i=1}^{n} \|\vec{x}_i\|_2^2 - \|\mathbf{V}\mathbf{V}^T\vec{x}_i\|_2^2$$

fixed



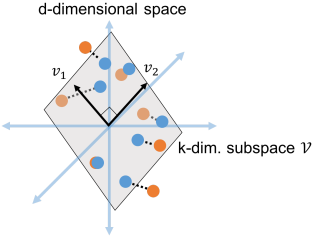
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Best Fit Subspace

If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

How do we find $\mathcal{V}$ (equivilantly $\mathbf{V}$)?

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\min} \|\mathbf{X}\|_F^2 - \|\mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2 = \sum_{i=1}^{n} \|\vec{x}_i\|_2^2 - \|\mathbf{V}\mathbf{V}^T\vec{x}_i\|_2^2$$



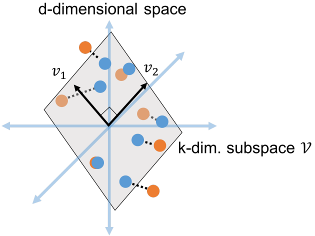
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Best Fit Subspace

If $\vec{x}_1, \ldots, \vec{x}_n$ are close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as $\mathbf{XVV}^T$. $\mathbf{XV}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

How do we find $\mathcal{V}$ (equivilantly $\mathbf{V}$)?

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{XVV}^T\|_F^2 = \sum_{i=1}^{n} \|\mathbf{VV}^T \vec{x}_i\|_2^2$$



$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

V minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

$$= \sum_{i=1}^{n} \|\mathbf{V}\mathbf{V}^T x_i\|_2^2$$

$$\left\| \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{bmatrix} V \end{bmatrix} \right\|_F^2 = \sum \text{squared row norms} = \sum \text{square column norms}$$

$$= \left\| \begin{bmatrix} Xv_1, Xv_2, \ldots Xv_k \end{bmatrix} \right\|_F^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

Surprisingly, can find the columns of $\mathbf{V}$, $\vec{v}_1, \ldots, \vec{v}_k$ greedily.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

Surprisingly, can find the columns of $\mathbf{V}$, $\vec{v}_1, \ldots, \vec{v}_k$ greedily.

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2 = 1}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

Surprisingly, can find the columns of $\mathbf{V}$, $\vec{v}_1, \ldots, \vec{v}_k$ greedily.

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2 = 1}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$$\vec{v}_2 = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_1 \rangle = 0}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$\|A\|_F^2 = \sum_{i=1} \sum_{j=1} A_{ij}^2$

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

Surprisingly, can find the columns of $\mathbf{V}$, $\vec{v}_1, \ldots, \vec{v}_k$ greedily.

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2 = 1}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$$\vec{v}_2 = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_1 \rangle = 0}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$$\cdots$$

$$\vec{v}_k = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_j \rangle = 0\ \forall j < k}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2.$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

$\|y\|_2^2 = y^T y$

$$\underset{\text{orthonormal } \mathbf{V} \in \mathbb{R}^{d \times k}}{\arg\max} \|\mathbf{X}\mathbf{V}\|_F^2 = \sum_{i=1}^{n} \|\mathbf{V}^T \vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|\mathbf{X}\vec{v}_j\|_2^2$$

Surprisingly, can find the columns of $\mathbf{V}$, $\vec{v}_1, \ldots, \vec{v}_k$ greedily.

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2 = 1}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2 \quad = v^T X^T X v$$

$$\vec{v}_2 = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_1 \rangle = 0}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2$$

$$\cdots$$

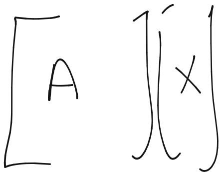
$$\vec{v}_k = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_j \rangle = 0\ \forall j < k}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2.$$

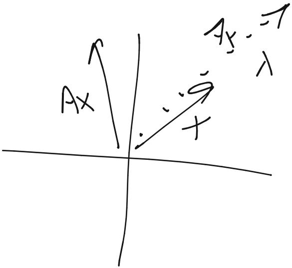
$\vec{v}_1, \ldots, \vec{v}_k$ are the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ by the *Courant-Fischer Principle.*

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Review of Eigenvectors and Eigendecomposition

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda\vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).





Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda\vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- That is, $\mathbf{A}$ just 'stretches' x.

Review of Eigenvectors and Eigendecomposition

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda\vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

$(X^T X)^T = X^T X$

- That is, $\mathbf{A}$ just 'stretches' x.
- If $\mathbf{A}$ is symmetric, can find d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns.

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda \vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- That is, $\mathbf{A}$ just 'stretches' x.
- If $\mathbf{A}$ is symmetric, can find d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns.

$$\mathbf{A}\mathbf{V} = \begin{bmatrix} | & | & | & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & | & | \end{bmatrix}$$

Review of Eigenvectors and Eigendecomposition

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda \vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- That is, $\mathbf{A}$ just 'stretches' x.
- If $\mathbf{A}$ is symmetric, can find d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns.

$$\mathbf{AV} = \begin{bmatrix} | & | & & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda\vec{v}_d \\ | & | & & | \end{bmatrix}$$

Review of Eigenvectors and Eigendecomposition

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda \vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- That is, $\mathbf{A}$ just 'stretches' x.
- If $\mathbf{A}$ is symmetric, can find d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns.

$$\mathbf{AV} = \begin{bmatrix} | & | & & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda\vec{v}_d \\ | & | & & | \end{bmatrix} = \mathbf{V\Lambda}$$

$$\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_d \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{bmatrix}$$

Review of Eigenvectors and Eigendecomposition

Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda \vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- That is, $\mathbf{A}$ just 'stretches' x.
- If $\mathbf{A}$ is symmetric, can find d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns.

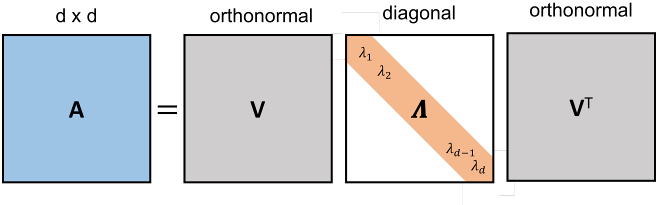
$$\mathbf{AV} = \begin{bmatrix} | & | & & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda\vec{v}_d \\ | & | & & | \end{bmatrix} = \mathbf{V\Lambda}$$

$AVV^T = V\Lambda V^T$

Yields eigendecomposition: $\mathbf{AVV}^T = \mathbf{A} = \mathbf{V\Lambda V}^T$.

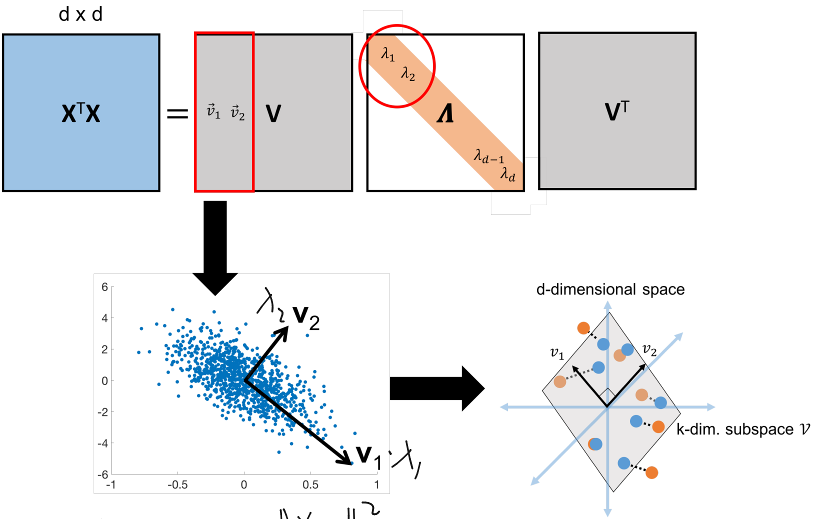
(eigendecomposition

Review of Eigenvectors and Eigendecomposition



Typically order the eigenvectors in decreasing order:
$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$.

Low-Rank Approximation via Eigendecomposition



$V_1 = \text{argmax} \; \|Xv_1\|_2^2$

Low-Rank Approximation via Eigendecomposition

Upshot: Letting $\mathbf{V}_k$ have columns $\vec{v}_1, \ldots, \vec{v}_k$ corresponding to the top k eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k$ is the orthogonal basis minimizing

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2.$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation via Eigendecomposition

Upshot: Letting $\mathbf{V}_k$ have columns $\vec{v}_1, \ldots, \vec{v}_k$ corresponding to the top k eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k$ is the orthogonal basis minimizing

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2.$$

MDS

SVD

This is principal component analysis (PCA). LSA

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation via Eigendecomposition

Upshot: Letting $\mathbf{V}_k$ have columns $\vec{v}_1, \ldots, \vec{v}_k$ corresponding to the top k eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k$ is the orthogonal basis minimizing

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2.$$

This is principal component analysis (PCA).

How accurate is this low-rank approximation?

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Low-Rank Approximation via Eigendecomposition

Upshot: Letting $\mathbf{V}_k$ have columns $\vec{v}_1, \ldots, \vec{v}_k$ corresponding to the top k eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k$ is the orthogonal basis minimizing

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2.$$

need algorithm to compute these

This is principal component analysis (PCA).

How accurate is this low-rank approximation? Can understand using eigenvalues of $\mathbf{X}^T\mathbf{X}$.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of X^TX (the top k principal components). Approximation error is:

$$\|X - XV_kV_k^T\|_F^2 = \|X\|_F^2 - \|XV_kV_k^T\|_F^2 = \|XV_k\|_F^2$$

(no distortion embedding)

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $X \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of X^TX, $V_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \|\mathbf{X}\|_F^2 - \|\mathbf{X}\mathbf{V}_k\|_F^2$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \|\mathbf{X}\|_F^2 - \|\mathbf{X}\mathbf{V}_k\|_F^2$$

$$\|y\|_2^2 = y^T y$$

- **Exercise:** For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^d \|\vec{a}_i\|_2^2 = \text{tr}(\mathbf{A}^T\mathbf{A})$ (sum of diagonal entries = sum eigenvalues).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \text{tr}(\mathbf{X}^T\mathbf{X}) - \text{tr}(\mathbf{V}_k^T\mathbf{X}^T\mathbf{X}\mathbf{V}_k)$$

- **Exercise:** For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^d \|\vec{a}_i\|_2^2 = \text{tr}(\mathbf{A}^T\mathbf{A})$ (sum of diagonal entries = sum eigenvalues).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \text{tr}(\mathbf{X}^T\mathbf{X}) - \text{tr}(\mathbf{V}_k^T\mathbf{X}^T\mathbf{X}\mathbf{V}_k)$$

$$= \sum_{i=1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X}) - \sum_{i=1}^{k} \vec{v}_i^T\mathbf{X}^T\mathbf{X}\vec{v}_i$$

– diagonal

$$\begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix} \begin{bmatrix} X^T X \end{bmatrix} \begin{bmatrix} v_1 \ldots v_k \end{bmatrix}$$

$\lambda_i \cdot v_i$

$v_i^T(\lambda_i v_i) = \lambda_i \cdot v_i^T v_i = \lambda_i (X^T X)$

- **Exercise:** For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^{d} \|\vec{a}_i\|_2^2 = \text{tr}(\mathbf{A}^T\mathbf{A})$ (sum of diagonal entries = sum eigenvalues).

> $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of X^TX (the top k principal components). Approximation error is:

$$\begin{aligned}\|X - XV_kV_k^T\|_F^2 &= \mathrm{tr}(X^TX) - \mathrm{tr}(V_k^TX^TXV_k)\\ &= \sum_{i=1}^{d} \lambda_i(X^TX) - \sum_{i=1}^{k} \vec{v}_i^TX^TX\vec{v}_i\\ &= \sum_{i=1}^{d} \lambda_i(X^TX) - \sum_{i=1}^{k} \lambda_i(X^TX)\end{aligned}$$

- **Exercise:** For any matrix A, $\|A\|_F^2 = \sum_{i=1}^d \|\vec{a}_i\|_2^2 = \mathrm{tr}(A^TA)$ (sum of diagonal entries = sum eigenvalues).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $X \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of X^TX, $V_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Let $\vec{v}_1, \ldots, \vec{v}_k$ be the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

$$\begin{aligned}\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 &= \mathrm{tr}(\mathbf{X}^T\mathbf{X}) - \mathrm{tr}(\mathbf{V}_k^T\mathbf{X}^T\mathbf{X}\mathbf{V}_k) \\ &= \sum_{i=1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X}) - \sum_{i=1}^{k} \vec{v}_i^T\mathbf{X}^T\mathbf{X}\vec{v}_i \\ &= \sum_{i=1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X}) - \sum_{i=1}^{k} \lambda_i(\mathbf{X}^T\mathbf{X}) = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})\end{aligned}$$

- **Exercise:** For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^{d} \|\vec{a}_i\|_2^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A})$ (sum of diagonal entries = sum eigenvalues).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n\times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d\times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

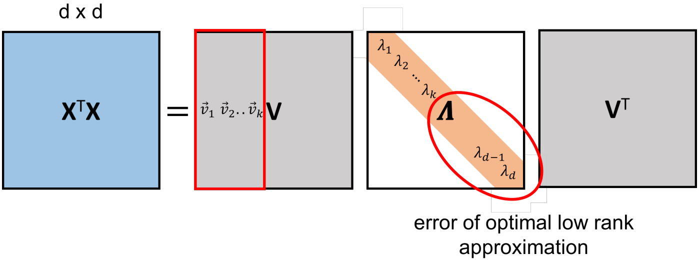
$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$

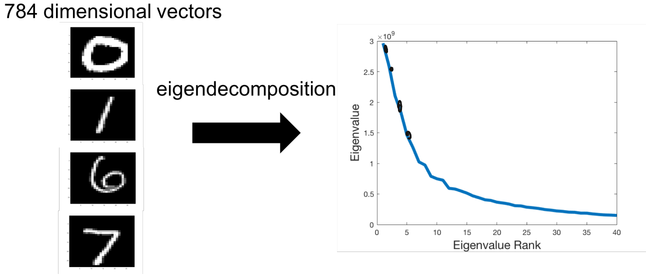


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$

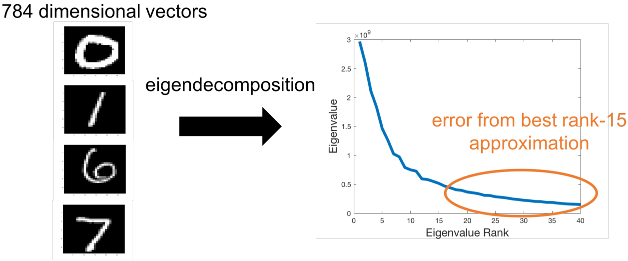


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$



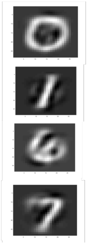
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

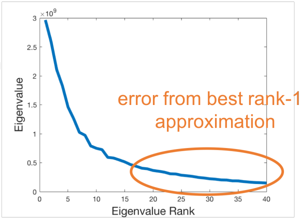
Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$

784 dimensional vectors



eigendecomposition

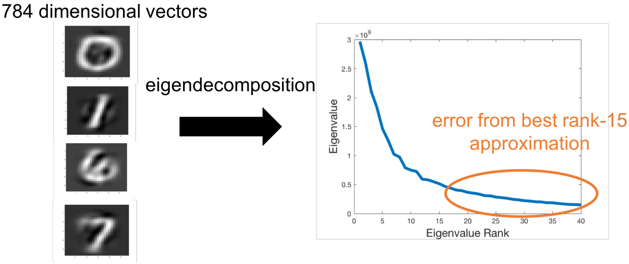


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

$$\|\mathbf{X} - \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T\|_F^2 = \sum_{i=k+1}^{d} \lambda_i(\mathbf{X}^T\mathbf{X})$$



- Choose k to balance accuracy/compression – often at an 'elbow'.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

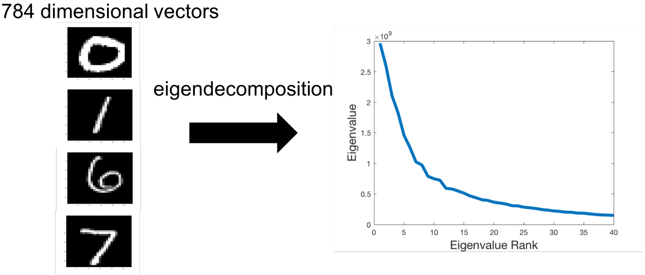
Spectrum Analysis

Plotting the spectrum of $\mathbf{X}^T\mathbf{X}$ (its eigenvalues) shows how compressible $\mathbf{X}$ is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

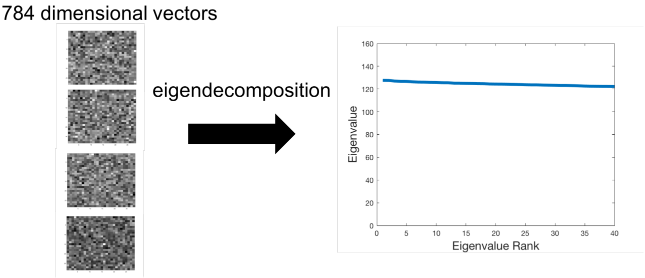
Plotting the spectrum of $\mathbf{X}^T\mathbf{X}$ (its eigenvalues) shows how compressible $\mathbf{X}$ is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).



$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

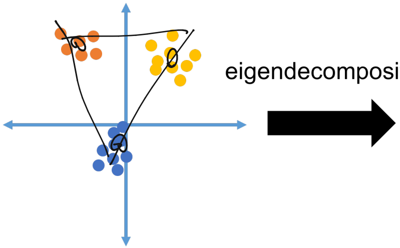
Plotting the spectrum of $\mathbf{X}^T\mathbf{X}$ (its eigenvalues) shows how compressible $\mathbf{X}$ is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).

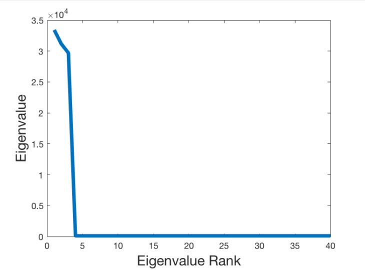


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

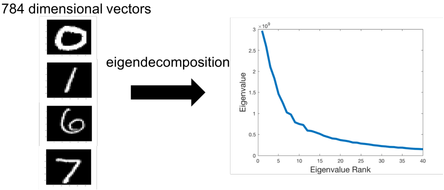
Plotting the spectrum of $\mathbf{X}^T\mathbf{X}$ (its eigenvalues) shows how compressible $\mathbf{X}$ is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).





$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis



Exercises:

1. Show that the eigenvalues of $\mathbf{X}^T\mathbf{X}$ are always positive. **Hint:** Use that $\lambda_j = \vec{v}_j^T\mathbf{X}^T\mathbf{X}\vec{v}_j$.
2. Show that for symmetric $\mathbf{A}$, the trace is the sum of eigenvalues: $\mathbf{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i(\mathbf{A})$. **Hint:** First prove the cyclic property of trace, that for any $\mathbf{MN}$, $\mathbf{tr}(\mathbf{MN}) = \mathbf{tr}(\mathbf{NM})$ and then apply this to $\mathbf{A}$'s eigendecomposition

Summary

- Many (most) datasets can be approximated via projection onto a low-dimensional subspace.
- Find this subspace via a maximization problem:

$$\max_{\text{orthonormal } \mathbf{V}} \|\mathbf{X}\mathbf{V}\|_F^2.$$

- Greedy solution via eigendecomposition of $\mathbf{X}^T\mathbf{X}$.
- Columns of $\mathbf{V}$ are the top eigenvectors of $\mathbf{X}^T\mathbf{X}$.
- Error of best low-rank approximation (compressibility of data) is determined by the tail of $\mathbf{X}^T\mathbf{X}$'s eigenvalue spectrum.