COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Fall 2024.

Lecture 15

Logistics

- We released Problem Set 3 on Friday. It is due Friday 11/8 at 11:59pm.
- Midterm grade have been posted. The average was good – around an 80%.
- Reach out to me if you are concerned about your midterm grade or your grade in the class overall.

Quiz Questions/Concerns

- Linear algebra review.
- Concerns about grades.
- Participation grade.

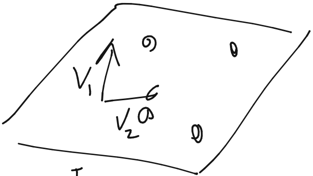
$\approx$ 78-80 $\approx$ B

- ec on problem sets

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. $X = CV^T = \underbrace{XV}_{C}V^T$
- Idea of approximating a data matrix $\mathbf{X}$ with $\mathbf{XVV}^T$ when the data points lie close to the subspace spanned by $\mathbf{V}$'s columns.



$$\|\underbrace{V^T x_i}_{\mathbb{R}^k} - V^T x_j\|_2 = \|\underbrace{x_i}_{\mathbb{R}^d} - x_j\|$$

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:

- 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.
- Measuring the error of the optimal low-rank approximation via covariance matrix eigenvalues.

Low-Rank Factorization

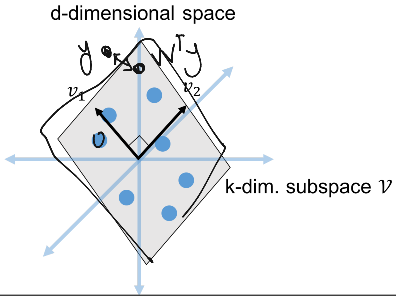
Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie in a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as

$$\begin{array}{c|c} & k \\ d & \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{bmatrix} \end{array}$$

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

VV^T: projection matrix onto subspace

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

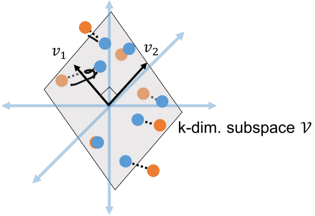


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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:

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

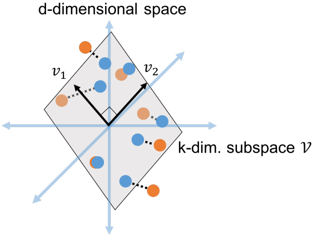


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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:

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



$$\begin{array}{ccc} & k & d \\ n & \left[\, XV \,\right] & \left[\, V^T \,\right] k \end{array}$$

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:

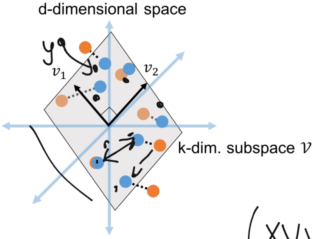
$(AB)^T = B^T A^T$

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

$d \left[VV^T \right]_{d} \left[y \right]_{d \times 1} = \left[VV^T y \right]_{d \times 1}$

finding subspace V

$\underset{k \times 1}{V^T y} \approx \underset{d \times 1}{y}$



$\begin{bmatrix} [\ x_i^T\ ] \\ X \\ \\ \end{bmatrix} \left[VV^T \right] = [x_i^T V V^T]$

$(XVV^T)^T = (VV^T)^T X^T = VV^T X^T$

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

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

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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

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 $\vec{v}_1, \ldots, \vec{v}_k$.

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,

$$\overset{d\times 1}{\|\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}\mathbf{V}^T\vec{x}_j\|_2} = \|\mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\mathbf{V}\mathbf{V}^T\vec{x}_j\|_2. = \overset{k\times 1}{\|\mathbf{V}^T\vec{x}_i - \mathbf{V}^T\vec{x}_j\|_2}.$$

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $\vec{v}_1, \ldots, \vec{v}_k$.

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.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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

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 $\mathbf{V}$).

$$\|V^T x - V^T y\| \neq \|x - y\|$$

- 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,

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

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

$$\underset{\text{original}}{n\overset{d}{\begin{bmatrix} X \end{bmatrix}}} \to n\overset{d}{\begin{bmatrix} XVV^T \end{bmatrix}} \to n\overset{k}{\begin{bmatrix} XV \end{bmatrix}}\ \text{compressed}$$

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 $\vec{y} \in \mathbb{R}^d$.

$V\underbrace{V^T V}_{I} V^T y = \underline{VV^T y}$. V is orthonormal so $V^T V = I_{k,k}$

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

$(VV^T y)^T$

$y^T (V^T)^T V^T = y^T VV^T$

$(VV^T)^T = VV^T$

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

$= VV^T - VV^T = 0$

k-dim subspace V

↳ set of all vectors $x = c_1 \vec{v}_1 + c_2 \vec{v}_2 + .. c_k \vec{v}_k$

for any coefficients $c_1, .. c_k \in \mathbb{R}$

Pythagorean Theorem

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

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

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

$\|a+b\|_2^2$

$\langle a+b, a+b\rangle$

$\langle a,a\rangle + \langle b,b\rangle$

$+ 2\langle a,b\rangle$

$\|a\|_2^2 + \|b\|_2^2$

$+ 2a^Tb$

$$VV^Ta = VV^Tb \quad \not\Downarrow \quad a = b$$

$$\underbrace{\|\overset{a}{VV^Ty} + \overset{b}{(I-VV^T)y}}_{y}\|_2^2 = \overset{\|a\|_2^2}{\|VV^Ty\|_2^2} + \overset{\|b\|_2^2}{\|(I-VV^T)y\|_2^2} + \overset{2a^Tb}{2y^TVV^T(I-VV^T)y} \to 0$$

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

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 $\vec{v}_1, \ldots, \vec{v}_k$.

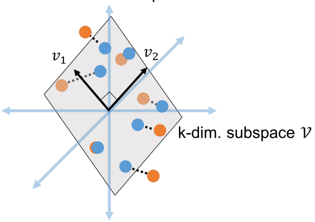
Best Fit Subspace

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 $\mathbf{V}$)?

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

$$\begin{bmatrix} X \end{bmatrix} - \begin{bmatrix} XVV^T \end{bmatrix}$$



$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $\vec{v}_1, \ldots, \vec{v}_k$.

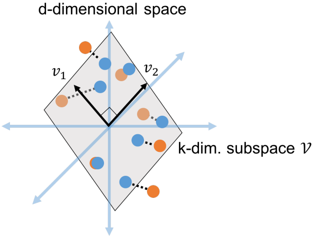
Best Fit Subspace

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 $\mathcal{V}$.

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

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 $\vec{v}_1, \ldots, \vec{v}_k$.

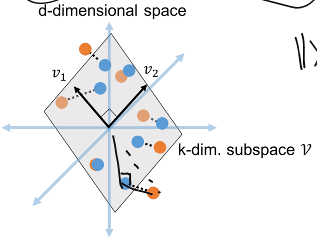
Best Fit Subspace

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 $\mathbf{X}\mathbf{V}\mathbf{V}^T$. $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$.

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

$\|X - XVV^T\|_F^2$ $\qquad$ $\|x_i - VV^T x_i\|_2^2$

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



$\|X - VV^TX\|_F^2 = \sum_{i=1}^n \|x_i - VV^T x_i\|_2^2$

pythagorem

$\sum_{i=1}^n \|x_i\|_2^2 - \|VV^T x_i\|_2^2$

$\underbrace{\sum_{i=1}^n \|x_i\|_2^2}_{\|X\|_F^2} - \underbrace{\sum_{i=1}^n \|VV^T x_i\|_2^2}_{\|XVV^T\|_F^2}$

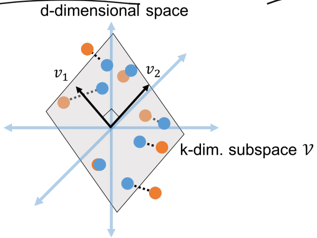
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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 $\mathbf{V}$)?

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



$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $\vec{v}_1, \ldots, \vec{v}_k$.

Solution via Eigendecomposition

V minimizing $\|X - XVV^T\|_F^2$ is given by:

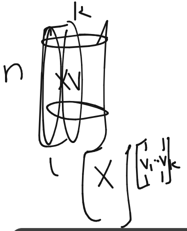
$$\underset{\text{orthonormal } V \in \mathbb{R}^{d\times k}}{\arg\max} \|XV\|_F^2 = \sum_{i=1}^{n} \|V^T\vec{x}_i\|_2^2 = \sum_{j=1}^{k} \|X\vec{v}_j\|_2^2$$

$$\underset{V}{\arg\max}\ \|XVV^T\|_F^2 = \sum_{i=1}^{n} \|VV^Tx_i\|_2^2 = \sum_{i=1}^{n} \|V^Tx_i\|_2^2$$

$$\|VV^Tx_i\|_2^2 = x_i^T V \underbrace{V^T V}_{I} V^T x_i$$

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

$$\|V^T x_i\|_2^2$$



$$XV = \begin{bmatrix} | & | & \\ Xv_1 & Xv_2 & \cdots \\ | & | & \end{bmatrix}$$

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

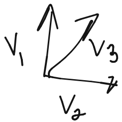
$\mathbf{V}$ minimizing $\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2$ is given by:

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

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:

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

$$\begin{bmatrix} X \end{bmatrix} \begin{bmatrix} v \end{bmatrix} = \overset{n\times 1}{\begin{bmatrix} Xv \end{bmatrix}}$$

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

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2 = 1}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2 \quad = v_1^T X^T X v_1$$

$$\vec{v}_2 = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_1 \rangle = 0}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2 \quad = v_2^T X^T X v_2$$

$$\cdots$$

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

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $\vec{v}_1, \ldots, \vec{v}_k$.

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.

eigenvectors of X^TX

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

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

$$\cdots$$

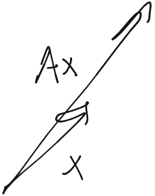
$$\vec{v}_k = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_j \rangle = 0\ \forall j < k}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2.$$

$\vec{v}_1, \ldots, \vec{v}_k$ 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 $\vec{v}_1, \ldots, \vec{v}_k$.

$_d\left[\,^{n}X^T\right]\left[X^{d}\right]_n = {}_d\left[X^T X\right]^{d}$

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 $\vec{x}$).



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.

Review of Eigenvectors and Eigendecomposition

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

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.
- 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.

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.

$$\mathbf{A}\mathbf{V} = \begin{array}{c} \begin{array}{cccc} \lambda_1 \cdot V_1 & \lambda_2 V_2 & \cdots & \lambda_d V_d \end{array} \\ \begin{bmatrix} | & | & | & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & | & | \end{bmatrix} \end{array}$$

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.

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

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.

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

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

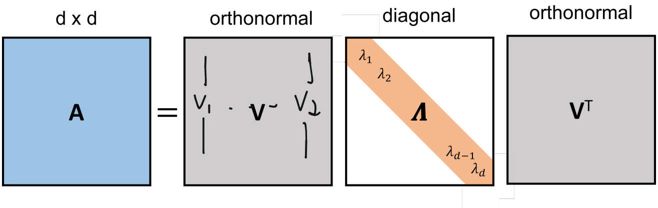
Eigenvector: $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ if $\mathbf{A}\vec{x} = \lambda \vec{x}$ for some scalar λ (the eigenvalue corresponding to $\vec{x}$).

- 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.

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

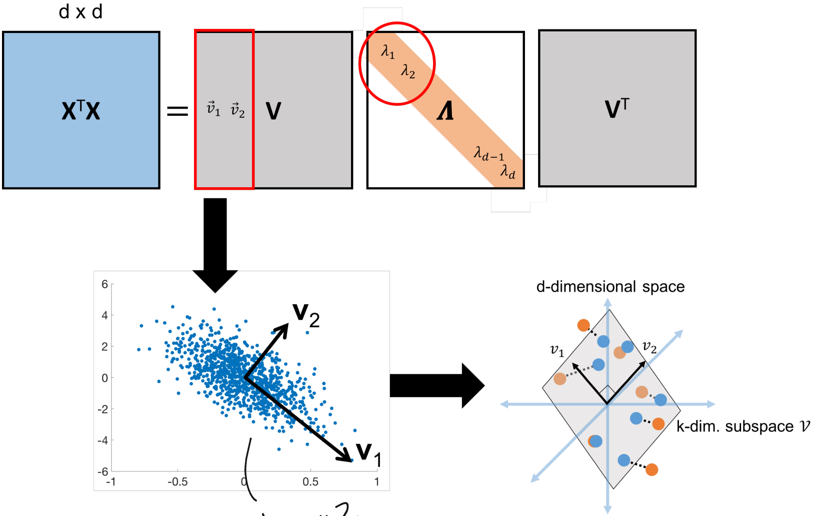
Yields eigendecomposition: $\mathbf{AVV}^T = \mathbf{A} \underbrace{= \mathbf{V\Lambda V}^T}_{\text{eigendecomp.}}.$

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



$\|Xv_1\|_2^2$

Low-Rank Approximation via Eigendecomposition

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

$$\|X - X V_k V_k^T\|_F^2.$$

L optimal low rank approximation

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $X^T X$, $V_k \in \mathbb{R}^{d \times k}$: 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).

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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

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

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 $\vec{v}_1, \ldots, \vec{v}_k$.