COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Fall 2024.

Lecture 16

Logistics

- Problem Set 3 due next Friday 11/8 at 11:59pm.
- There is no class next Tuesday due to election day. But I will hold my regular office hours from 2:30-3:30pm. Location TBD.
- Get any midterm regrade requests in by tomorrow.

Summary

Last 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 the data does not exactly lie in a low-dimensional subspace.
- Solution by taking the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (this is PCA/optimal low-rank approximation)
- Greedy optimization problem and connection to Courant-Fischer principal.

Summary

Last 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 the data does not exactly lie in a low-dimensional subspace.
- Solution by taking the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (this is PCA/optimal low-rank approximation)
- Greedy optimization problem and connection to Courant-Fischer principal.

This Class:

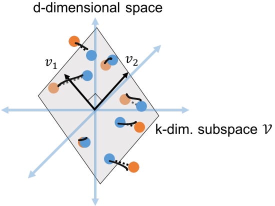
- Wrap up optimal low-rank approximation.
- Measuring the error of the low-rank approximation via covariance matrix eigenvalues.
- General linear algebra review.

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 $\mathcal{V}$.

We can find $\mathbf{V}$ by solving the optimization problem:

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



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

"covariance matrix"

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

$$d\left[\, X^T \,\right]_n \; n\left[\, X \,\right]^d = \left[\, X^TX \,\right]^d_d$$

$$(X^TX)_{ij} = \langle x_i, x_j \rangle$$

$$(X^TX)v = \lambda v$$

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

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

$$\cdots$$

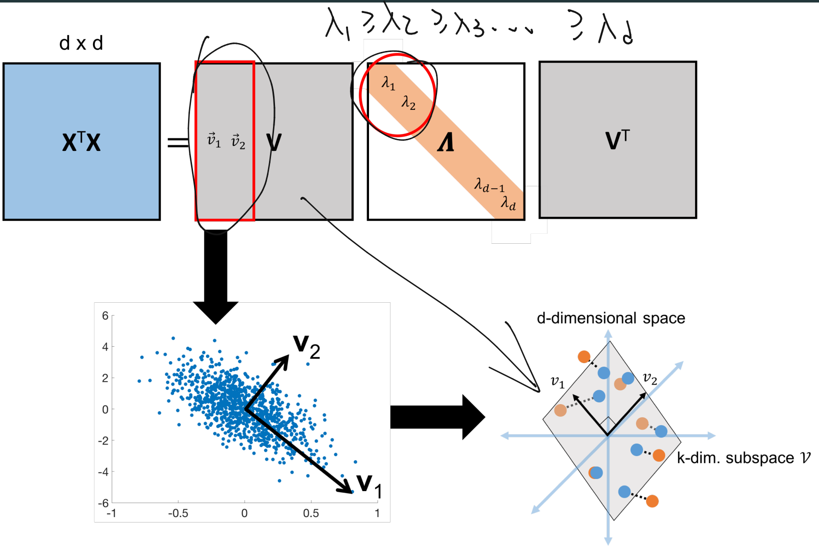
$$\vec{v}_k = \underset{\vec{v} \text{ with } \|v\|_2 = 1,\ \langle \vec{v}, \vec{v}_j \rangle = 0\ \forall j < k}{\arg\max} \|\mathbf{X}\vec{v}\|_2^2.$$

$$\sum_{i=1}^{k} \|Xv_i\|_2^2$$

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

Low-Rank Approximation via Eigendecomposition



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

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

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

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 $\mathbf{X}^T\mathbf{X}$ (the top k principal components). Approximation error is:

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

Pythagorean theorem

$$\|x_i - V_kV_k^Tx_i\|_2^2 = \|x_i\|_2^2 - \|V_kV_k^Tx_i\|_2^2$$

"no distortion embedding"

$$\|V_kV_k^Tx_i\|_2^2 = \|V_k^Tx_i\|_2^2$$

$$\|X - XV_kV_k^T\|_F^2 = \|(X - XV_kV_k^T)^T\|_F^2 = tr\left((X - XV_kV_k^T)(X - XV_kV_k^T)^T\right)$$

$$= tr\left(XX^T - 2XV_kV_k^TX^T + XV_kV_k^TV_kV_k^TX^T\right)$$

($V_k^TV_k = I$)

$$tr(XX^T - XV_kV_k^TX^T) = tr(XX^T) - tr(XV_kV_k^TX^T) = \|X\|_F^2 - \|XV_k\|_F^2$$

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

$$tr(A+B) = \sum_{i=1} (A_{ii} + B_{ii}) = \sum_{i=1} A_{ii} + \sum_{i=1} B_{ii} = tr(A) + tr(B)$$

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:

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

ith column of A

$\|A\|_F^2$

$$tr(A^TA) = \sum_{i=1}^{d} (A^TA)_{ii} = \sum_{i=1}^{d} \|a_i\|_2^2 = \sum_{i=1}^{d}\sum_{j=1}^{n} A_{ji}^2 = \|A\|_F^2$$

- **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:

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

if $V\Lambda V^T$ is eigendecomp of A.

$\mathrm{tr}(A) = \mathrm{tr}(V\Lambda V^T) = \mathrm{tr}(V^TV\Lambda) = \mathrm{tr}(\Lambda) = \sum_{i=1}^{d} \lambda_i$ (cyclic prop.)

Exercise: Cyclic property of trace: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

- **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). → symmetric

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

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

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

- **Exercise:** For any matrix $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^{d} \|\vec{a}_i\|_2^2 = \operatorname{tr}(\mathbf{A}^T\mathbf{A})$ (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:

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

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

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

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

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

X^TX is positive semidefinite, $\lambda_i(X^TX) \geq 0 \ \forall i$

$\vec{v}_i^T X^T X \vec{v}_i = \lambda_i(X^TX) = \|X v_i\|_2^2 \geq 0$

- **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

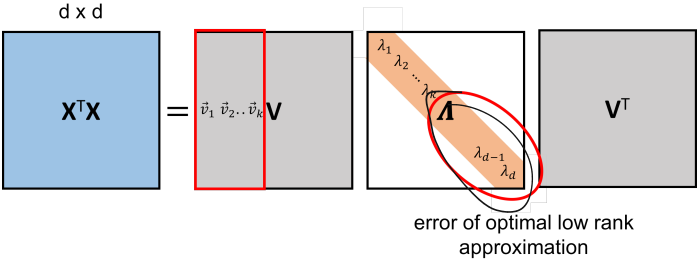
Claim: The error in approximating $\mathbf{X}$ with the best rank k approximation (projecting onto the top k eigenvectors of $\mathbf{X}^T\mathbf{X}$) is:

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

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

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:

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

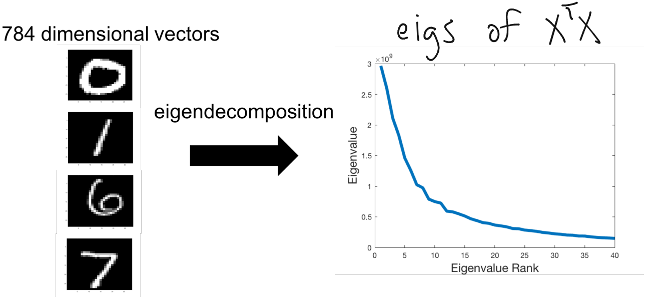


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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:

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



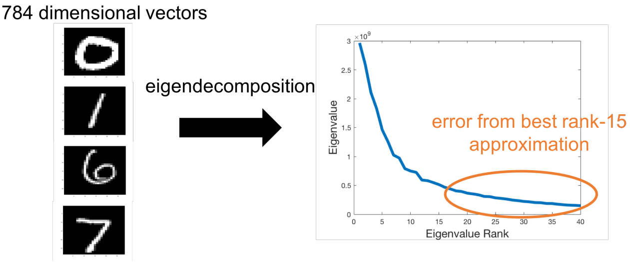
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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:

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

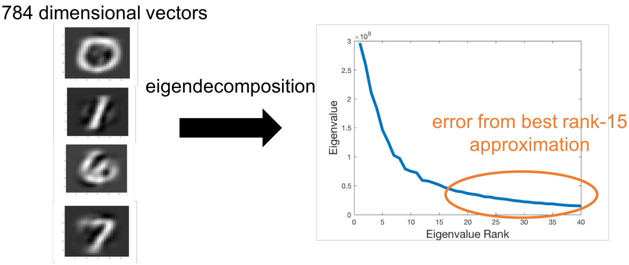
$\|X\|_F^2$



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

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:

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

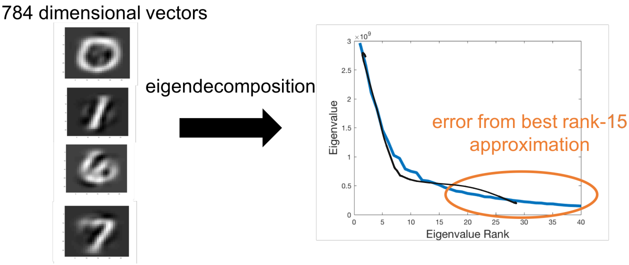


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{V}_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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:

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



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

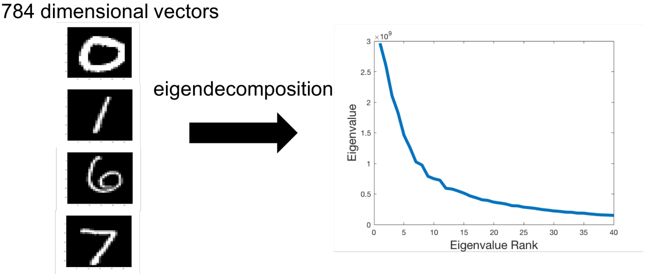
Spectrum Analysis

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

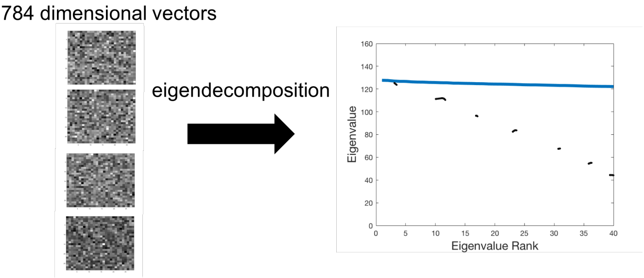
Plotting the spectrum of $\mathbf{X}^T\mathbf{X}$ (its eigenvalues) shows how compressible $\mathbf{X}$ is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).



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

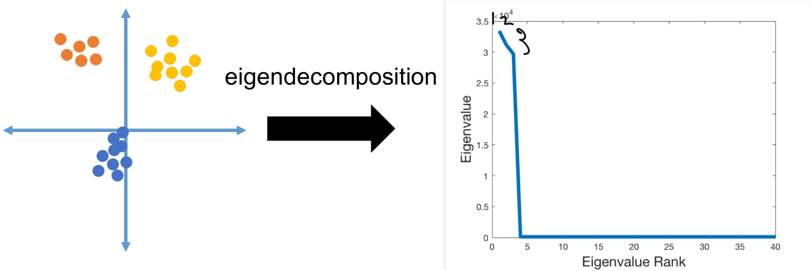
Plotting the spectrum of X^TX (its eigenvalues) shows how compressible X is using low-rank approximation (i.e., how close $\vec{x}_1, \ldots, \vec{x}_n$ are to a low-dimensional subspace).



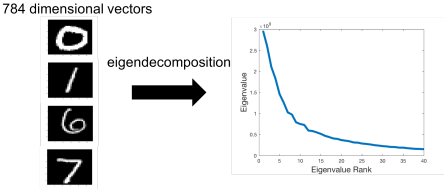
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $X \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: top eigenvectors of X^TX, $V_k \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Spectrum Analysis

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



Exercises:

non-negative ≥ 0

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:

$$\|X - XVV^T\| \Leftrightarrow \max_{\text{orthonormal } V} \|XV\|_F^2.$$

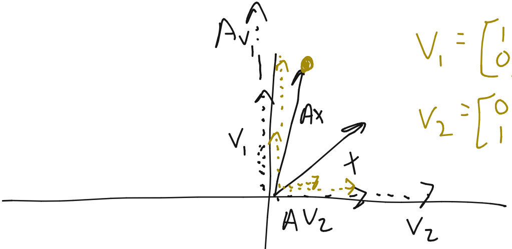
- Greedy solution via eigendecomposition of $X^T X$.
- Columns of V are the top eigenvectors of $X^T X$.
- Error of best low-rank approximation (compressibility of data) is determined by the tail of $X^T X$'s eigenvalue spectrum.

Linear Algebra Review

Eigenvectors / Eigenvalues

$$\underset{2\times 2}{A} = \underset{2\times 2}{V} \Lambda V^T = \begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

$$\underset{2\times 2}{\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} c_1 v_1 \end{bmatrix}} = \underset{2\times 1}{\begin{bmatrix} v_1^T c_1 v_1 \\ v_2^T c_1 v_1 \end{bmatrix}} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}$$



$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\lambda_1 = 2$

$v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\lambda_2 = 0.3$

$$Ax = A(c_1 v_1 + c_2 v_2) = V \Lambda V^T (c_1 v_1 + c_2 v_2)$$

$$= V\Lambda \left(\begin{bmatrix} c_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ c_2 \end{bmatrix} \right) = V\Lambda \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = V \begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \end{bmatrix} = \lambda_1 c_1 \cdot v_1 + \lambda_2 c_2 v_2$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

$$V_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \lambda_1 = 3$$

$$V_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \lambda_2 = 2$$

$$V_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \lambda_3 = 1$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

all vectors are eigenvectors of I

Linear Algebra Review

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$$

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \lambda_1 = 3$$

$$tr(A) = 3 = \sum_{i=1}^{3} \lambda_i$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}_A \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix}$$

$$v_2 = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \quad \lambda_2 = 0$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$v_3 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \quad \lambda_3 = 0$$

$$rank(A) = 1$$

$$\|A - A v_1 v_1^T\|_F^2 = 0 = \sum_{i=2}^{3} \lambda_i$$

$$\det(A - \lambda I) = \prod (\lambda_i - \lambda)$$

$$\lambda_1 - \lambda$$

$$\lambda_2 - \lambda$$

$$\vdots$$

$$\lambda_d - \lambda$$

Linear Algebra Review

Linear Algebra Review

Linear Algebra Review

Linear Algebra Review

Linear Algebra Review

Linear Algebra Review