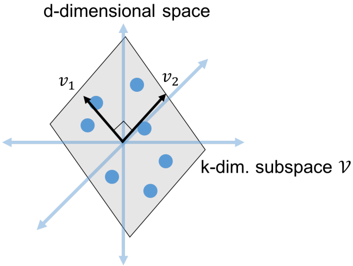
COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 14

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 $\mathbb{R}^d$.



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.

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

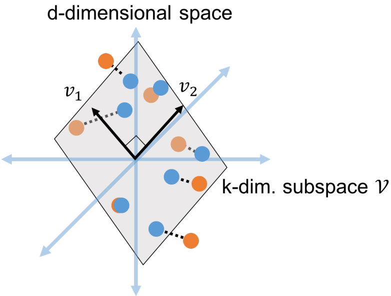
Letting $\tilde{x}_i = \mathbf{V}^T \vec{x}_i$, we have a perfect embedding from $\mathcal{V}$ into $\mathbb{R}^k$.

LAST CLASS: PROJECTION VIEW

Warm-Up: 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

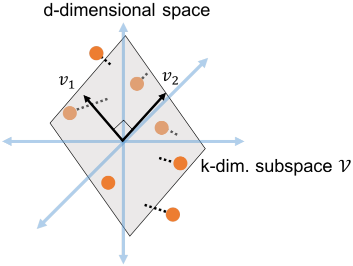
$$\mathbf{X} = \mathbf{X}\mathbf{V}\mathbf{V}^T = \mathbf{C}\mathbf{V}^T$$

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



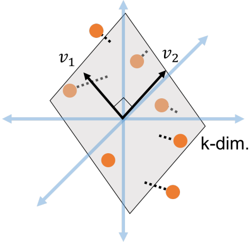
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

Assume data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to some k-dimensional subspace $\mathcal{V}$:



EMBEDDING WITH ASSUMPTIONS

Assume data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to some k-dimensional subspace $\mathcal{V}$:

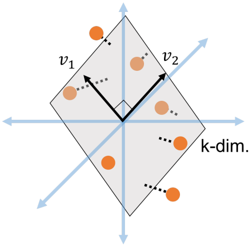


Letting $\vec{v}_1, \ldots, \vec{v}_k$ be an orthonormal basis for $\mathcal{V}$ and $\mathbf{V} \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns, $\mathbf{V}^T \vec{x}_i \in \mathbb{R}^k$ is still a good embedding for $x_i \in \mathbb{R}^d$ and $\mathbf{XVV}^T$ is still a good approximation for $\mathbf{X}$ in the sense:

$$\mathbf{XVV^T} = \underset{\mathbf{B} \text{ with rows in } \mathcal{V}}{\arg\min} \|\mathbf{X} - \mathbf{B}\|_F^2.$$

EMBEDDING WITH ASSUMPTIONS

Assume data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to some k-dimensional subspace $\mathcal{V}$:



Letting $\vec{v}_1, \ldots, \vec{v}_k$ 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, $\mathbf{V}^T \vec{x}_i \in \mathbb{R}^k$ is still a good embedding for $x_i \in \mathbb{R}^d$ and $\mathbf{XVV}^T$ is still a good approximation for $\mathbf{X}$ in the sense:

$$\mathbf{XVV^T} = \underset{\mathbf{B} \text{ with rows in } \mathcal{V}}{\arg\min} \|\mathbf{X} - \mathbf{B}\|_F^2.$$

Will show first show this and then investigate how do we find $\mathcal{V}$ and $\mathbf{V}$?

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

$$(\mathbf{V}\mathbf{V}^T)(\mathbf{V}\mathbf{V}^T)\vec{y} = (\mathbf{V}\mathbf{V}^T)\vec{y}$$

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

$$(\mathbf{V}\mathbf{V}^T)(\mathbf{V}\mathbf{V}^T)\vec{y} = (\mathbf{V}\mathbf{V}^T)\vec{y}$$

Exercise 2: The projection is orthogonal to its complement: For any $\vec{y} \in \mathbb{R}^d$, $\langle \mathbf{V}\mathbf{V}^T\vec{y}, \vec{y} - \mathbf{V}\mathbf{V}^T\vec{y} \rangle = 0$

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

$$(\mathbf{V}\mathbf{V}^T)(\mathbf{V}\mathbf{V}^T)\vec{y} = (\mathbf{V}\mathbf{V}^T)\vec{y}$$

Exercise 2: The projection is orthogonal to its complement: For any $\vec{y} \in \mathbb{R}^d$, $\langle \mathbf{V}\mathbf{V}^T\vec{y}, \vec{y} - \mathbf{V}\mathbf{V}^T\vec{y} \rangle = 0$

Implies the Pythagorean Theorem: Show that for any $\vec{y} \in \mathbb{R}^d$,

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

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

$$(\mathbf{V}\mathbf{V}^T)(\mathbf{V}\mathbf{V}^T)\vec{y} = (\mathbf{V}\mathbf{V}^T)\vec{y}$$

Exercise 2: The projection is orthogonal to its complement: For any $\vec{y} \in \mathbb{R}^d$, $\langle \mathbf{V}\mathbf{V}^T\vec{y}, \vec{y} - \mathbf{V}\mathbf{V}^T\vec{y}\rangle = 0$

Implies the Pythagorean Theorem: Show that for any $\vec{y} \in \mathbb{R}^d$,

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

Follows since $\vec{y} = (\vec{y} - (\mathbf{V}\mathbf{V}^T)\vec{y}) + (\mathbf{V}\mathbf{V}^T)\vec{y}$ and

$$\|\vec{a} + \vec{b}\|_2^2 = \|\vec{a}\|_2^2 + \|\vec{b}\|_2^2 + 2\langle \vec{a}, \vec{b}\rangle \ .$$

PROJECTION VECTOR IS CLOSEST POINT IN SUBSPACE

Let $\mathbf{V} \in \mathbb{R}^{n \times k}$ have orthonormal columns and let $\vec{y} \in \mathbb{R}^n$. Then the Pythagorean Theorem proves that $\mathbf{V}\mathbf{V}^T\vec{y}$ is the closest vector to $\vec{y}$ that can be expressed as a linear combination of the columns of $\mathbf{V}$

Let $\mathbf{V} \in \mathbb{R}^{n \times k}$ have orthonormal columns and let $\vec{y} \in \mathbb{R}^n$. Then the Pythagorean Theorem proves that $\mathbf{V}\mathbf{V}^T\vec{y}$ is the closest vector to $\vec{y}$ that can be expressed as a linear combination of the columns of $\mathbf{V}$

- Apply Pythagorus to $\vec{y} - \vec{z}$ for arbitrary $\vec{z} \in \mathbb{R}^n$:

$$\begin{aligned}\|\vec{y} - \vec{z}\|_2^2 &= \|\mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 + \|\vec{y} - \vec{z} - \mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 \\ &= \|\mathbf{V}\mathbf{V}^T\vec{y} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2 + \|\vec{y} - \mathbf{V}\mathbf{V}^T\vec{y} + \vec{z} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2.\end{aligned}$$

PROJECTION VECTOR IS CLOSEST POINT IN SUBSPACE

Let $\mathbf{V} \in \mathbb{R}^{n \times k}$ have orthonormal columns and let $\vec{y} \in \mathbb{R}^n$. Then the Pythagorean Theorem proves that $\mathbf{V}\mathbf{V}^T\vec{y}$ is the closest vector to $\vec{y}$ that can be expressed as a linear combination of the columns of $\mathbf{V}$

- Apply Pythagorus to $\vec{y} - \vec{z}$ for arbitrary $\vec{z} \in \mathbb{R}^n$:

$$\begin{aligned} \|\vec{y} - \vec{z}\|_2^2 &= \|\mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 + \|\vec{y} - \vec{z} - \mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 \\ &= \|\mathbf{V}\mathbf{V}^T\vec{y} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2 + \|\vec{y} - \mathbf{V}\mathbf{V}^T\vec{y} + \vec{z} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2. \end{aligned}$$

- If $\vec{z} = \mathbf{V}\vec{c}$ for some $\vec{c} \in \mathbb{R}^k$, then $\mathbf{V}\mathbf{V}^T\vec{z} = \mathbf{V}\mathbf{V}^T\mathbf{V}\vec{c} = \mathbf{V}\vec{c} = \vec{z}$ and the above simplifies to

$$\|\mathbf{V}\mathbf{V}^T\vec{y} - \vec{z}\|_2^2 + \|\vec{y} - \mathbf{V}\mathbf{V}^T\vec{y}\|_2^2$$

PROJECTION VECTOR IS CLOSEST POINT IN SUBSPACE

Let $\mathbf{V} \in \mathbb{R}^{n\times k}$ have orthonormal columns and let $\vec{y} \in \mathbb{R}^n$. Then the Pythagorean Theorem proves that $\mathbf{V}\mathbf{V}^T\vec{y}$ is the closest vector to $\vec{y}$ that can be expressed as a linear combination of the columns of $\mathbf{V}$

- Apply Pythagorus to $\vec{y} - \vec{z}$ for arbitrary $\vec{z} \in \mathbb{R}^n$:

$$\begin{aligned}\|\vec{y} - \vec{z}\|_2^2 &= \|\mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 + \|\vec{y} - \vec{z} - \mathbf{V}\mathbf{V}^T(\vec{y} - \vec{z})\|_2^2 \\ &= \|\mathbf{V}\mathbf{V}^T\vec{y} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2 + \|\vec{y} - \mathbf{V}\mathbf{V}^T\vec{y} + \vec{z} - \mathbf{V}\mathbf{V}^T\vec{z}\|_2^2.\end{aligned}$$

- If $\vec{z} = \mathbf{V}\vec{c}$ for some $\vec{c} \in \mathbb{R}^k$, then $\mathbf{V}\mathbf{V}^T\vec{z} = \mathbf{V}\mathbf{V}^T\mathbf{V}\vec{c} = \mathbf{V}\vec{c} = \vec{z}$ and the above simplifies to

$$\|\mathbf{V}\mathbf{V}^T\vec{y} - \vec{z}\|_2^2 + \|\vec{y} - \mathbf{V}\mathbf{V}^T\vec{y}\|_2^2$$

- To minimize this, set $\vec{z} = \mathbf{V}\mathbf{V}^T\vec{y}$.

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 $\mathbf{XVV}^T$ and $\mathbf{XV}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$. How do we find $\mathcal{V}$ (equivalently $\mathbf{V}$)?

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 $\mathbf{X}\mathbf{V}\mathbf{V}^T$ and $\mathbf{X}\mathbf{V}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$. How do we find $\mathcal{V}$ (equivalently $\mathbf{V}$)?

$$
\begin{aligned}
\|\mathbf{X} - \mathbf{X}\mathbf{V}\mathbf{V}^T\|_F^2 &= \|\mathbf{X}^T - \mathbf{V}\mathbf{V}^T\mathbf{X}^T\|_F^2 \\
&= \sum_{i=1}^{n} \|\vec{x}_i - \mathbf{V}\mathbf{V}^T\vec{x}_i\|_2^2 \\
&= \sum_{i=1}^{n} \|\vec{x}_i\|_2^2 - \|\mathbf{V}\mathbf{V}^T\vec{x}_i\|_2^2
\end{aligned}
$$

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 $\mathbf{XVV}^T$ and $\mathbf{XV}$ gives optimal embedding of $\mathbf{X}$ in $\mathcal{V}$. How do we find $\mathcal{V}$ (equivalently $\mathbf{V}$)?

$$\begin{aligned} \|\mathbf{X} - \mathbf{XVV}^T\|_F^2 &= \|\mathbf{X}^T - \mathbf{VV}^T\mathbf{X}^T\|_F^2 \\ &= \sum_{i=1}^n \|\vec{x}_i - \mathbf{VV}^T\vec{x}_i\|_2^2 \\ &= \sum_{i=1}^n \|\vec{x}_i\|_2^2 - \|\mathbf{VV}^T\vec{x}_i\|_2^2 \end{aligned}$$

So minimizing $\|\mathbf{X} - \mathbf{XVV}^T\|_F^2$ is the same as maximizing

$$\sum_i \|\mathbf{VV}^T\vec{x}_i\|_2^2 = \sum_i \vec{x}_i^T\mathbf{VV}^T\mathbf{VV}^T\vec{x}_i = \sum_i \|\mathbf{V}^T\vec{x}_i\|_2^2$$

SOLUTION VIA EIGENDECOMPOSITION

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

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

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

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

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

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

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:

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

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

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

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

SOLUTION VIA EIGENDECOMPOSITION

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

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

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

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

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

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

SOLUTION VIA EIGENDECOMPOSITION

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

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

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

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

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

$$\cdots$$

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

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

SOLUTION VIA EIGENDECOMPOSITION

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

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

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

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

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

$$\cdots$$

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

These are exactly the top k eigenvectors 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, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthogonal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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.

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, it has d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns and $\mathbf{\Lambda}$ be the diagonal matrix with corresponding eigenvalues on the diagonal.

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, it has d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns and $\mathbf{\Lambda}$ be the diagonal matrix with corresponding eigenvalues on the diagonal.

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

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, **A** just 'stretches' x.
- If **A** is symmetric, it has d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns and $\boldsymbol{\Lambda}$ be the diagonal matrix with corresponding eigenvalues on the diagonal.

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

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, it has d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns and $\mathbf{\Lambda}$ be the diagonal matrix with corresponding eigenvalues on the diagonal.

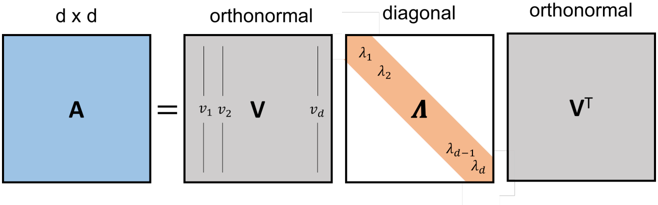
$$\mathbf{AV} = \begin{bmatrix} | & | & | & | \\ \mathbf{A}\vec{v}_1 & \mathbf{A}\vec{v}_2 & \cdots & \mathbf{A}\vec{v}_d \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | & | \\ \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda\vec{v}_d \\ | & | & | & | \end{bmatrix} = \mathbf{V\Lambda}$$

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, it has d orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ have these vectors as columns and $\mathbf{\Lambda}$ be the diagonal matrix with corresponding eigenvalues on the diagonal.

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

Yields eigendecomposition: $\mathbf{AVV}^T = \mathbf{A} = \mathbf{V\Lambda V}^T$ where the first inequality follows since rows of $\mathbf{A}$ are in span of the eigenvectors.



Typically order the eigenvectors in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$$

Courant-Fischer Principal: For symmetric $\mathbf{A}$, the eigenvectors are given via the greedy optimization:

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2=1}{\arg\max} \vec{v}^T \mathbf{A} \vec{v}.$$

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

$$\cdots$$

$$\vec{v}_d = \underset{\vec{v} \text{ with } \|v\|_2=1,\ \langle \vec{v}, \vec{v}_j \rangle = 0\ \forall j < d}{\arg\max} \vec{v}^T \mathbf{A} \vec{v}.$$

COURANT-FISCHER PRINCIPAL

Courant-Fischer Principal: For symmetric $\mathbf{A}$, the eigenvectors are given via the greedy optimization:

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2=1}{\arg\max} \; \vec{v}^T \mathbf{A} \vec{v}.$$

$$\vec{v}_2 = \underset{\vec{v} \text{ with } \|v\|_2=1, \; \langle \vec{v}, \vec{v}_1 \rangle = 0}{\arg\max} \; \vec{v}^T \mathbf{A} \vec{v}.$$

$$\cdots$$

$$\vec{v}_d = \underset{\vec{v} \text{ with } \|v\|_2=1, \; \langle \vec{v}, \vec{v}_j \rangle = 0 \;\; \forall j < d}{\arg\max} \; \vec{v}^T \mathbf{A} \vec{v}.$$

- $\vec{v}_j^T \mathbf{A} \vec{v}_j = \lambda_j \cdot \vec{v}_j^T \vec{v}_j = \lambda_j$, the j^{th} largest eigenvalue.

Courant-Fischer Principal: For symmetric $\mathbf{A}$, the eigenvectors are given via the greedy optimization:

$$\vec{v}_1 = \underset{\vec{v} \text{ with } \|v\|_2=1}{\arg\max} \ \vec{v}^T \mathbf{A} \vec{v}.$$

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

$$\cdots$$

$$\vec{v}_d = \underset{\vec{v} \text{ with } \|v\|_2=1, \ \langle \vec{v}, \vec{v}_j \rangle = 0 \ \forall j < d}{\arg\max} \ \vec{v}^T \mathbf{A} \vec{v}.$$

- $\vec{v}_j^T \mathbf{A} \vec{v}_j = \lambda_j \cdot \vec{v}_j^T \vec{v}_j = \lambda_j$, the j^{th} largest eigenvalue.
- The first k eigenvectors of $\mathbf{X}^T\mathbf{X}$ (corresponding to the largest k eigenvalues) are exactly the directions of greatest "variance" in $\mathbf{X}$ that we use for low-rank approximation. We'll talk more about this next time.