COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor
Lecture 15
Set Up: Assume that data points $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d$ lie in some $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

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

\[
\| V^T \tilde{x}_i - V^T \tilde{x}_j \|_2^2 = \| \tilde{x}_i - \tilde{x}_j \|_2^2.
\]

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

\[
\mathbf{X} = \mathbf{XVV}^T = \mathbf{CV}^T
\]

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

$$X = XVV^T = CV^T$$

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

$d$-dimensional space

k-dim. subspace $\mathcal{V}$
**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

$$\mathbf{X} = \mathbf{XVV}^T = \mathbf{CV}^T$$

- $\mathbf{VV}^T$ 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 that data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ lie close to any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$. 

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

$$XV^T = \arg \min_B \|X - B\|_F$$ 

Will show above in homework. Today's focus: How do we find $V$ and $V$?
Assume that data points \( \vec{x}_1, \ldots, \vec{x}_n \) lie close to any \( k \)-dimensional subspace \( \mathcal{V} \) of \( \mathbb{R}^d \).

Will show above in homework. Today’s focus: How do we find \( \mathcal{V} \) and \( \mathbf{V} \)?
Assume that data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

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 $\vec{x}_i \in \mathbb{R}^d$ and $\mathbf{XVV}^T$ is still a good approximation for $\mathbf{X}$:

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

Will show above in homework. Today’s focus: How do we find $\mathcal{V}$ and $\mathbf{V}$?
Question: Why might we expect $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ to lie close to a $k$-dimensional subspace?
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 $X$ can be approximately reconstructed from a basis of $k$ vectors.
**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.
Question: Why might we expect \( \bar{x}_1, \ldots, \bar{x}_n \in \mathbb{R}^d \) 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 $\mathbf{X}$ are approx. spanned by $k$ vectors.
**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:**

<table>
<thead>
<tr>
<th></th>
<th>bedrooms</th>
<th>bathrooms</th>
<th>sq.ft.</th>
<th>floors</th>
<th>list price</th>
<th>sale price</th>
</tr>
</thead>
<tbody>
<tr>
<td>home 1</td>
<td>2</td>
<td>2</td>
<td>1800</td>
<td>2</td>
<td>200,000</td>
<td>195,000</td>
</tr>
<tr>
<td>home 2</td>
<td>4</td>
<td>2.5</td>
<td>2700</td>
<td>1</td>
<td>300,000</td>
<td>310,000</td>
</tr>
<tr>
<td>home n</td>
<td>5</td>
<td>3.5</td>
<td>3600</td>
<td>3</td>
<td>450,000</td>
<td>450,000</td>
</tr>
</tbody>
</table>
**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:**

<table>
<thead>
<tr>
<th>Home 1</th>
<th>Home 2</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>bedrooms</td>
<td>bathrooms</td>
<td>sq.ft.</td>
<td>floors</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1800</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2.5</td>
<td>2700</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Home n</td>
<td>bedrooms</td>
<td>bathrooms</td>
<td>sq.ft.</td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
<td>3600</td>
<td>3</td>
</tr>
</tbody>
</table>
**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:**

<table>
<thead>
<tr>
<th></th>
<th>bedrooms</th>
<th>bathrooms</th>
<th>sq.ft.</th>
<th>floors</th>
<th>list price</th>
<th>sale price</th>
</tr>
</thead>
<tbody>
<tr>
<td>home 1</td>
<td>2</td>
<td>2</td>
<td>1800</td>
<td>2</td>
<td>200,000</td>
<td>195,000</td>
</tr>
<tr>
<td>home 2</td>
<td>4</td>
<td>2.5</td>
<td>2700</td>
<td>1</td>
<td>300,000</td>
<td>310,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>home n</td>
<td>5</td>
<td>3.5</td>
<td>3600</td>
<td>3</td>
<td>450,000</td>
<td>450,000</td>
</tr>
</tbody>
</table>
**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:

<table>
<thead>
<tr>
<th></th>
<th>bedrooms</th>
<th>bathrooms</th>
<th>sq.ft.</th>
<th>floors</th>
<th>list price</th>
<th>sale price</th>
</tr>
</thead>
<tbody>
<tr>
<td>home 1</td>
<td>2</td>
<td>2</td>
<td>1800</td>
<td>2</td>
<td>200,000</td>
<td>195,000</td>
</tr>
<tr>
<td>home 2</td>
<td>4</td>
<td>2.5</td>
<td>2700</td>
<td>1</td>
<td>300,000</td>
<td>310,000</td>
</tr>
<tr>
<td>home n</td>
<td>5</td>
<td>3.5</td>
<td>3600</td>
<td>3</td>
<td>450,000</td>
<td>450,000</td>
</tr>
</tbody>
</table>

10000* bathrooms + 10* (sq. ft.) \( \approx \) list price
Quick Exercise 1: Show that $VV^T$ is idempotent. I.e.,
$$(VV^T)(VV^T)\vec{y} = (VV^T)\vec{y}$$ for any $\vec{y} \in \mathbb{R}^d$. 
Quick Exercise 1: Show that $\mathbf{VV}^T$ is idempotent. I.e.,

$$(\mathbf{VV}^T)(\mathbf{VV}^T)\mathbf{y} = (\mathbf{VV}^T)\mathbf{y}$$

for any $\mathbf{y} \in \mathbb{R}^d$.

Quick Exercise 2: The projection is orthogonal to its complement: For any $\mathbf{y} \in \mathbb{R}^d$, 

$$\langle \mathbf{VV}^T \mathbf{y}, (\mathbf{I} - \mathbf{V V}^T)\mathbf{y} \rangle = 0$$
Quick Exercise 1: Show that $\mathbf{VV}^T$ is idempotent. I.e.,

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

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

Quick Exercise 2: The projection is orthogonal to its complement: For any $\vec{y} \in \mathbb{R}^d$,

$$\langle \mathbf{VV}^T \vec{y}, (\mathbf{I} - \mathbf{VV}^T)\vec{y} \rangle = 0$$

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

$$\|\vec{y}\|_2^2 = \|\mathbf{VV}^T\vec{y}\|_2^2 + \|\vec{y} - (\mathbf{VV}^T)\vec{y}\|_2^2.$$
**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 $\vec{y} \in \mathbb{R}^d$.

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

**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

$$\|\vec{a} + \vec{b}\|_2^2 = \|\vec{a}\|_2^2 + \|\vec{b}\|_2^2 + 2\langle \vec{a}, \vec{b} \rangle.$$
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 $\mathbf{V}$)?
If \( \vec{x}_1, \ldots, \vec{x}_n \) are close to a \( k \)-dimensional subspace \( \mathcal{V} \) with orthonormal basis \( V \in \mathbb{R}^{d \times k} \), the data matrix can be approximated as \( XVV^T \) and \( XV \) gives optimal embedding of \( X \) in \( \mathcal{V} \). How do we find \( \mathcal{V} \) (equivalently \( V \))?

\[
\|X - XVV^T\|_F^2 = \sum_{i,j}(x_{i,j} - (XVV^T)_{i,j})^2
\]

\[
= \sum_{i=1}^{n}\|\vec{x}_i - vv^T \vec{x}_i\|_2^2
\]

\[
= \sum_{i=1}^{n}\|\vec{x}_i\|_2^2 - \|vv^T \vec{x}_i\|_2^2
\]
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 \( \mathbf{V} \))?

\[
\| \mathbf{X} - \mathbf{X} \mathbf{V} \mathbf{V}^T \|_F^2 = \sum_{i,j} (\mathbf{x}_{i,j} - (\mathbf{X} \mathbf{V} \mathbf{V}^T)_{i,j})^2 = \sum_{i=1}^{n} \| \vec{x}_i - \mathbf{V} \mathbf{V}^T \vec{x}_i \|_2^2 = \sum_{i=1}^{n} \| \vec{x}_i \|_2^2 - \| \mathbf{V} \mathbf{V}^T \vec{x}_i \|_2^2
\]

So minimizing \( \| \mathbf{X} - \mathbf{X} \mathbf{V} \mathbf{V}^T \|_F^2 \) is the same as maximizing

\[
\sum_i \| \mathbf{V} \mathbf{V}^T \vec{x}_i \|_2^2 = \sum_i \vec{x}_i^T \mathbf{V} \mathbf{V}^T \mathbf{V} \mathbf{V}^T \vec{x}_i = \sum_i \| \mathbf{V}^T \vec{x}_i \|_2^2
\]
V minimizing \( \|X - XVV^T\|_F^2 \) is given by:

\[
\arg \max_{\text{orthonormal } V \in \mathbb{R}^{d \times k}} \sum_{i=1}^{n} \left\| V^T \vec{x}_i \right\|_2^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} \langle \vec{v}_j, \vec{x}_i \rangle^2
\]

\( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \): data points, \( X \in \mathbb{R}^{n \times d} \): 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 \( \vec{v}_1, \ldots, \vec{v}_k \).
V minimizing $\|X - XVV^T\|^2_F$ is given by:

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

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

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

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

Surprisingly, can find the columns of \( \mathbf{V} \), \( \tilde{v}_1, \ldots, \tilde{v}_k \) greedily:

\[
\tilde{v}_1 = \arg \max_{\tilde{v} \text{ with } \| \tilde{v} \|_2 = 1} \| \mathbf{X} \tilde{v} \|^2_2.
\]

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

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

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

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

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

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 \( \| \mathbf{X} - \mathbf{X} \mathbf{VV}^T \|_F^2 \) is given by:

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

Surprisingly, can find the columns of \( \mathbf{V}, \tilde{v}_1, \ldots, \tilde{v}_k \) greedily.

\[
\tilde{v}_1 = \arg \max_{\tilde{v} \text{ with } \| \tilde{v} \|_2 = 1} \tilde{v}^T \mathbf{X}^T \mathbf{X} \tilde{v}.
\]

\[
\tilde{v}_2 = \arg \max_{\tilde{v} \text{ with } \| \tilde{v} \|_2 = 1, \langle \tilde{v}, \tilde{v}_1 \rangle = 0} \tilde{v}^T \mathbf{X}^T \mathbf{X} \tilde{v}.
\]

\( \tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d \): data points, \( \mathbf{X} \in \mathbb{R}^{n \times d} \): data matrix, \( \tilde{v}_1, \ldots, \tilde{v}_k \in \mathbb{R}^d \): orthogonal basis for subspace \( \mathcal{V} \). \( \mathbf{V} \in \mathbb{R}^{d \times k} \): matrix with columns \( \tilde{v}_1, \ldots, \tilde{v}_k \).
**SOLUTION VIA EIGENDECOMPOSITION**

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

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

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

$$\bar{v}_1 = \text{arg max}_{\bar{v} \text{ with } \|\bar{v}\|_2 = 1} \bar{v}^T X^T X \bar{v}.$$  

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

$$\vdots$$

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

---

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

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

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

Surprisingly, can find the columns of \( \mathbf{V} \), \( \tilde{v}_1, \ldots, \tilde{v}_k \) greedily.

\[
\tilde{v}_1 = \arg \max_{\mathbf{\bar{v}} \text{ with } \| \mathbf{\bar{v}} \|_2 = 1} \mathbf{\bar{v}}^T \mathbf{X}^T \mathbf{X} \mathbf{\bar{v}}.
\]

\[
\tilde{v}_2 = \arg \max_{\mathbf{\bar{v}} \text{ with } \| \mathbf{\bar{v}} \|_2 = 1, \langle \mathbf{\bar{v}}, \tilde{v}_1 \rangle = 0} \mathbf{\bar{v}}^T \mathbf{X}^T \mathbf{X} \mathbf{\bar{v}}.
\]

\[\ldots\]

\[
\tilde{v}_k = \arg \max_{\mathbf{\bar{v}} \text{ with } \| \mathbf{\bar{v}} \|_2 = 1, \langle \mathbf{\bar{v}}, \tilde{v}_j \rangle = 0 \forall j < k} \mathbf{\bar{v}}^T \mathbf{X}^T \mathbf{X} \mathbf{\bar{v}}.
\]

These are exactly the top \( k \) eigenvectors of \( \mathbf{X}^T \mathbf{X} \).

\( \tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d \): data points, \( \mathbf{X} \in \mathbb{R}^{n \times d} \): data matrix, \( \tilde{v}_1, \ldots, \tilde{v}_k \in \mathbb{R}^d \): orthogonal basis for subspace \( \mathcal{V} \). \( \mathbf{V} \in \mathbb{R}^{d \times k} \): matrix with columns \( \tilde{v}_1, \ldots, \tilde{v}_k \).
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 \( \lambda \) (the eigenvalue corresponding to \( \vec{x} \)).
**Eigenvalue:** $\vec{x} \in \mathbb{R}^d$ is an eigenvector of a matrix $A \in \mathbb{R}^{d \times d}$ if $A\vec{x} = \lambda \vec{x}$ for some scalar $\lambda$ (the eigenvalue corresponding to $\vec{x}$).

- That is, $A$ just ‘stretches’ $\vec{x}$. 

**Eigendecomposition:**

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

- That is, \( \mathbf{A} \) just ‘stretches’ \( \mathbf{x} \).
- If \( \mathbf{A} \) is symmetric, it has \( d \) orthonormal eigenvectors \( \mathbf{v}_1, \ldots, \mathbf{v}_d \).

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.

Yields eigendecomposition:

\[
\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T
\]

where the first inequality follows since rows of \( \mathbf{A} \) are in span of the eigenvectors.
EIGENVECTORS AND EIGENDECOMPOSITION

**Eigenvector:** \( \vec{x} \in \mathbb{R}^d \) is an eigenvector of a matrix \( A \in \mathbb{R}^{d \times d} \) if \( A\vec{x} = \lambda \vec{x} \) for some scalar \( \lambda \) (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 \( V \in \mathbb{R}^{d \times d} \) have these vectors as columns and \( \Lambda \) be the diagonal matrix with corresponding eigenvalues on the diagonal.

\[
AV = \begin{bmatrix}
A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_d
\end{bmatrix}
\]

Yields eigendecomposition: 
\[
AV = \Lambda \]

where the first inequality follows since rows of \( A \) are in span of the eigenvectors.
**Eigenvectors and Eigendecomposition**

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

- That is, \( A \) just 'stretches' \( \vec{x} \).
- If \( A \) is **symmetric**, it has \( d \) orthonormal eigenvectors \( \vec{v}_1, \ldots, \vec{v}_d \).

Let \( V \in \mathbb{R}^{d \times d} \) have these vectors as columns and \( \Lambda \) be the diagonal matrix with corresponding eigenvalues on the diagonal.

\[
AV = \begin{bmatrix}
A \vec{v}_1 & A \vec{v}_2 & \cdots & A \vec{v}_d
\end{bmatrix} = \begin{bmatrix}
\lambda_1 \vec{v}_1 & \lambda_2 \vec{v}_2 & \cdots & \lambda_d \vec{v}_d
\end{bmatrix}
\]
**EIGENVECTORS AND EIGENDECOMPOSITION**

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

- That is, \( A \) just ‘stretches’ \( \vec{x} \).
- If \( A \) is symmetric, it has \( d \) orthonormal eigenvectors \( \vec{v}_1, \ldots, \vec{v}_d \).

Let \( V \in \mathbb{R}^{d \times d} \) have these vectors as columns and \( \Lambda \) be the diagonal matrix with corresponding eigenvalues on the diagonal.

\[
AV = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_d \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{v}_1 & \lambda_2 \vec{v}_2 & \cdots & \lambda_d \vec{v}_d \end{bmatrix} = V\Lambda
\]
**Eigenvectors and Eigendecomposition**

**Eigenvector:** \( \vec{x} \in \mathbb{R}^d \) is an eigenvector of a matrix \( A \in \mathbb{R}^{d \times d} \) if \( A\vec{x} = \lambda \vec{x} \) for some scalar \( \lambda \) (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 \( V \in \mathbb{R}^{d \times d} \) have these vectors as columns and \( \Lambda \) be the diagonal matrix with corresponding eigenvalues on the diagonal.

\[
AV = \begin{bmatrix}
A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_d
\end{bmatrix} = \begin{bmatrix}
\lambda_1 \vec{v}_1 & \lambda_2 \vec{v}_2 & \cdots & \lambda_d \vec{v}_d
\end{bmatrix} = V\Lambda
\]

Yields eigendecomposition: \( AVV^T = A = V\Lambda V^T \) where the first inequality follows since rows of \( A \) 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 $A$, the eigenvectors are given via the greedy optimization:

$$\vec{v}_1 = \arg\max_{\vec{v}} \vec{v}^T A \vec{v}.$$  
$\vec{v}$ with $\|\vec{v}\|_2 = 1$

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

$$\ldots$$

$$\vec{v}_d = \arg\max_{\vec{v}} \vec{v}^T A \vec{v}.$$  
$\vec{v}$ with $\|\vec{v}\|_2 = 1$, $\langle \vec{v}, \vec{v}_j \rangle = 0$ $\forall j < d$
Courant-Fischer Principal: For symmetric $A$, the eigenvectors are given via the greedy optimization:

$$\vec{v}_1 = \arg \max_{\vec{v}} \vec{v}^T A \vec{v} \text{ with } \|\vec{v}\|_2 = 1$$

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

$$\ldots$$

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

- $\vec{v}_j^T A \vec{v}_j = \lambda_j \cdot \vec{v}_j^T \vec{v}_j = \lambda_j$, the $j^{th}$ largest eigenvalue.
Courant-Fischer Principal: For symmetric $A$, the eigenvectors are given via the greedy optimization:

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

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

\[ \cdots \]

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

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

- The first $k$ eigenvectors of $X^T X$ (corresponding to the largest $k$ eigenvalues) are exactly the directions of greatest "variance" in $X$ that we use for low-rank approximation.