COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 13

Next Few Classes: Low-rank approximation, the SVD, and principal component analysis (PCA).

- Reduce d-dimensional data points to a smaller dimension m.
- Like JL, compression is linear, i.e., by applying a matrix.
- Chose matrix taking into account structure of dataset.
- Can give better compression than random projection.

Next Few Classes: Low-rank approximation, the SVD, and principal component analysis (PCA).

- Reduce d-dimensional data points to a smaller dimension m.
- Like JL, compression is linear, i.e., by applying a matrix.
- Chose matrix taking into account structure of dataset.
- Can give better compression than random projection.

Will be using a fair amount of linear algebra. Today we'll use:

- Vectors $\vec{v}_1, \ldots, \vec{v}_k$ are orthonormal if $\|\vec{v}_i\|_2 = 1$ and $\langle \vec{v}_i, \vec{v}_j \rangle = 0 \quad \forall i \neq j$

Next Few Classes: Low-rank approximation, the SVD, and principal component analysis (PCA).

- Reduce d-dimensional data points to a smaller dimension m.
- Like JL, compression is linear, i.e., by applying a matrix.
- Chose matrix taking into account structure of dataset.
- Can give better compression than random projection.

Will be using a fair amount of linear algebra. Today we'll use:

- Vectors $\vec{v}_1, \ldots, \vec{v}_k$ are orthonormal if $\|\vec{v}_i\|_2 = 1$ and $\langle \vec{v}_i, \vec{v}_j \rangle = 0 \quad \forall i \neq j$
- $\|\vec{v}\|_2^2 = \vec{v}^T \vec{v}$ and $(AB)^T = B^T A^T$ where $(\cdot)^T$ means transpose, i.e., swapping rows and columns.

Next Few Classes: Low-rank approximation, the SVD, and principal component analysis (PCA).

- Reduce d-dimensional data points to a smaller dimension m.
- Like JL, compression is linear, i.e., by applying a matrix.
- Chose matrix taking into account structure of dataset.
- Can give better compression than random projection.

Will be using a fair amount of linear algebra. Today we'll use:

- Vectors $\vec{v}_1, \ldots, \vec{v}_k$ are orthonormal if $\|\vec{v}_i\|_2 = 1$ and $\langle \vec{v}_i, \vec{v}_j \rangle = 0 \quad \forall i \neq j$
- $\|\vec{v}\|_2^2 = \vec{v}^T \vec{v}$ and $(AB)^T = B^T A^T$ where $(\cdot)^T$ means transpose, i.e., swapping rows and columns.
- A set of vectors $\mathcal{B}$ is a basis for a set of vectors $\mathcal{V}$, if every vector in $\mathcal{V}$ is a linear combination of vectors in $\mathcal{B}$.

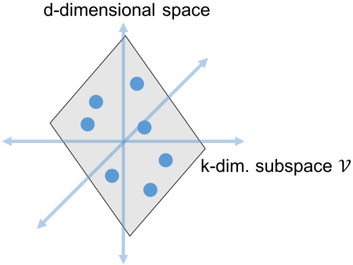
Next Few Classes: Low-rank approximation, the SVD, and principal component analysis (PCA).

- Reduce d-dimensional data points to a smaller dimension m.
- Like JL, compression is linear, i.e., by applying a matrix.
- Chose matrix taking into account structure of dataset.
- Can give better compression than random projection.

Will be using a fair amount of linear algebra. Today we'll use:

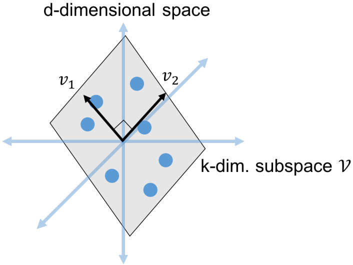
- Vectors $\vec{v}_1, \ldots, \vec{v}_k$ are orthonormal if $\|\vec{v}_i\|_2 = 1$ and $\langle \vec{v}_i, \vec{v}_j \rangle = 0 \quad \forall i \neq j$
- $\|\vec{v}\|_2^2 = \vec{v}^T \vec{v}$ and $(AB)^T = B^T A^T$ where $(\cdot)^T$ means transpose, i.e., swapping rows and columns.
- A set of vectors $\mathcal{B}$ is a basis for a set of vectors $\mathcal{V}$, if every vector in $\mathcal{V}$ is a linear combination of vectors in $\mathcal{B}$.
- The dimension of $\mathcal{V}$ is the size of its smallest basis.

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



EMBEDDING WITH ASSUMPTIONS

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

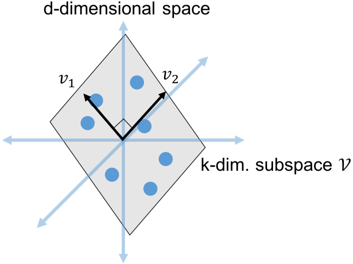


Claim: Let $\vec{v}_1, \ldots, \vec{v}_k$ be an orthonormal basis for $\mathcal{V}$ and $\mathbf{V} \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns. For all $\vec{x}_i, \vec{x}_j$:

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

EMBEDDING WITH ASSUMPTIONS

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



Claim: 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. For all $\vec{x}_i, \vec{x}_j$:

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

That is, $\mathbf{V}^T \in \mathbb{R}^{k \times d}$ is a linear embedding of $\vec{x}_1, \ldots, \vec{x}_n$ into k dimensions with no distortion.

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$

DOT PRODUCT TRANSFORMATION

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y}$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T (\mathbf{V}\vec{c})$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T\vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i\vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T\vec{y} = (\mathbf{V}\vec{c})^T(\mathbf{V}\vec{c}) = \vec{c}^T\mathbf{V}^T\mathbf{V}\vec{c}$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T (\mathbf{V}\vec{c}) = \vec{c}^T \mathbf{V}^T \mathbf{V} \vec{c}$
- $\|\mathbf{V}^T \vec{y}\|_2^2 = (\mathbf{V}^T \vec{y})^T (\mathbf{V}^T \vec{y}) = \vec{y}^T \mathbf{V}\mathbf{V}^T \vec{y}$

DOT PRODUCT TRANSFORMATION

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T (\mathbf{V}\vec{c}) = \vec{c}^T \mathbf{V}^T \mathbf{V}\vec{c}$
- $\|\mathbf{V}^T \vec{y}\|_2^2 = (\mathbf{V}^T \vec{y})^T (\mathbf{V}^T \vec{y}) = \vec{y}^T \mathbf{V}\mathbf{V}^T \vec{y} = \vec{c}^T \mathbf{V}^T \mathbf{V}\mathbf{V}^T \mathbf{V}\vec{c}$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T (\mathbf{V}\vec{c}) = \vec{c}^T \mathbf{V}^T \mathbf{V} \vec{c}$
- $\|\mathbf{V}^T \vec{y}\|_2^2 = (\mathbf{V}^T \vec{y})^T (\mathbf{V}^T \vec{y}) = \vec{y}^T \mathbf{V}\mathbf{V}^T \vec{y} = \vec{c}^T \mathbf{V}^T \mathbf{V}\mathbf{V}^T \mathbf{V} \vec{c}$
- But $\mathbf{V}^T \mathbf{V} = I$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T (\mathbf{V}\vec{c}) = \vec{c}^T \mathbf{V}^T \mathbf{V} \vec{c}$
- $\|\mathbf{V}^T \vec{y}\|_2^2 = (\mathbf{V}^T \vec{y})^T (\mathbf{V}^T \vec{y}) = \vec{y}^T \mathbf{V}\mathbf{V}^T \vec{y} = \vec{c}^T \mathbf{V}^T \mathbf{V}\mathbf{V}^T \mathbf{V} \vec{c}$
- But $\mathbf{V}^T \mathbf{V} = I$ since

$$[\mathbf{V}^T \mathbf{V}]_{i,j} = \vec{v}_i^T \vec{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Claim: 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. For all $\vec{y} \in \mathcal{V}$:

$$\|\mathbf{V}^T \vec{y}\|_2 = \|\vec{y}\|_2.$$

Proof:

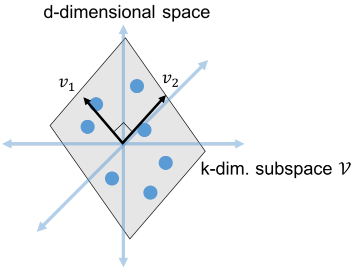
- If $\vec{y} = \sum_i c_i \vec{v}_i$ then $\vec{y} = \mathbf{V}\vec{c}$ where $\vec{c}^T = (c_1, \ldots, c_k)$
- $\|\vec{y}\|_2^2 = \vec{y}^T \vec{y} = (\mathbf{V}\vec{c})^T(\mathbf{V}\vec{c}) = \vec{c}^T \mathbf{V}^T \mathbf{V} \vec{c}$
- $\|\mathbf{V}^T \vec{y}\|_2^2 = (\mathbf{V}^T \vec{y})^T (\mathbf{V}^T \vec{y}) = \vec{y}^T \mathbf{V}\mathbf{V}^T \vec{y} = \vec{c}^T \mathbf{V}^T \mathbf{V}\mathbf{V}^T \mathbf{V} \vec{c}$
- But $\mathbf{V}^T \mathbf{V} = I$ since

$$[\mathbf{V}^T \mathbf{V}]_{i,j} = \vec{v}_i^T \vec{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

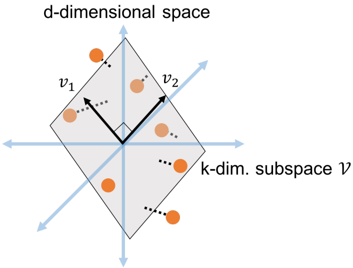
- So $\|\vec{y}\|_2^2 = \vec{c}^T \vec{c} = \|\mathbf{V}^T \vec{y}\|_2^2$.

EMBEDDING WITH ASSUMPTIONS

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

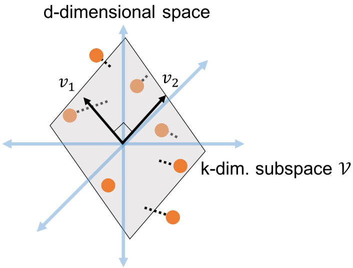


Now assume that data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to any k-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.



EMBEDDING WITH ASSUMPTIONS

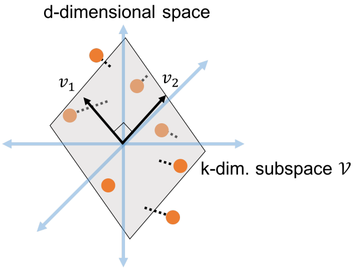
Now assume that data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to any k-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.



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 $x_i \in \mathbb{R}^d$.

EMBEDDING WITH ASSUMPTIONS

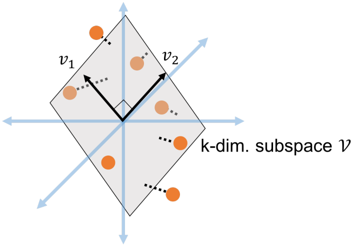
Now assume that data points $\vec{x}_1, \ldots, \vec{x}_n$ lie close to any k-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.



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 $x_i \in \mathbb{R}^d$. This is the key idea behind low-rank approximation and principal component analysis (PCA).

EMBEDDING WITH ASSUMPTIONS

Now 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 $x_i \in \mathbb{R}^d$. This is the key idea behind low-rank approximation and principal component analysis (PCA).

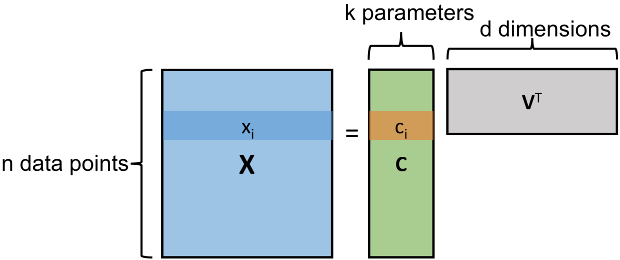
- How do we find $\mathcal{V}$ and $\mathbf{V}$?
- How good is the embedding?

- Every data point $\vec{x}_i$ (row of $\mathbf{X}$) can be written as

$$\vec{x}_i = \mathbf{V}\vec{c}_i = c_{i,1} \cdot \vec{v}_1 + \ldots + c_{i,k} \cdot \vec{v}_k$$

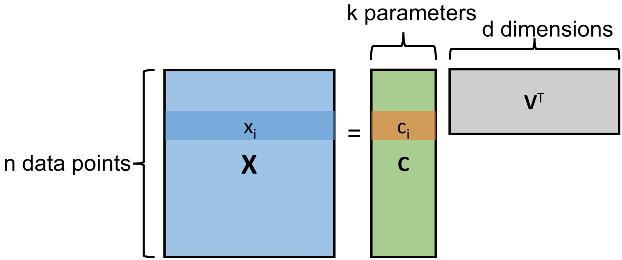
- Every data point $\vec{x}_i$ (row of $\mathbf{X}$) can be written as

$$\vec{x}_i = \mathbf{V}\vec{c}_i = c_{i,1} \cdot \vec{v}_1 + \ldots + c_{i,k} \cdot \vec{v}_k$$



- Every data point $\vec{x}_i$ (row of $\mathbf{X}$) can be written as

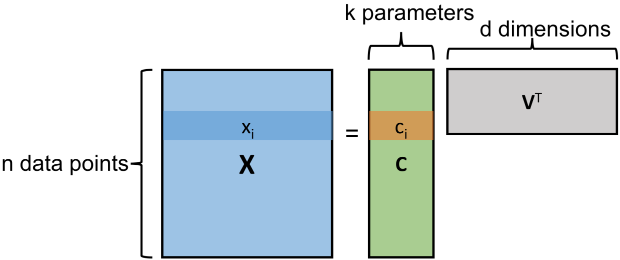
$$\vec{x}_i = \mathbf{V}\vec{c}_i = c_{i,1} \cdot \vec{v}_1 + \ldots + c_{i,k} \cdot \vec{v}_k$$



- $\mathbf{X}$ can be represented by $(n + d) \cdot k$ parameters vs. $n \cdot d$.

- Every data point $\vec{x}_i$ (row of $\mathbf{X}$) can be written as

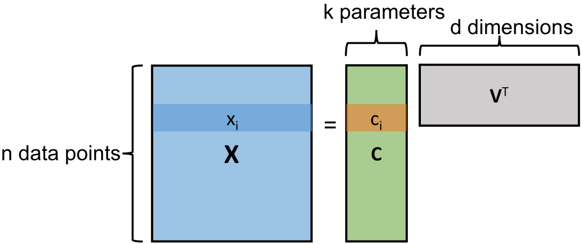
$$\vec{x}_i = \mathbf{V}\vec{c}_i = c_{i,1} \cdot \vec{v}_1 + \ldots + c_{i,k} \cdot \vec{v}_k$$



- $\mathbf{X}$ can be represented by $(n+d) \cdot k$ parameters vs. $n \cdot d$.
- The rows of $\mathbf{X}$ are spanned by k vectors: the columns of $\mathbf{V}$ $\implies$ the columns of $\mathbf{X}$ are spanned by k vectors: the columns of $\mathbf{C}$.

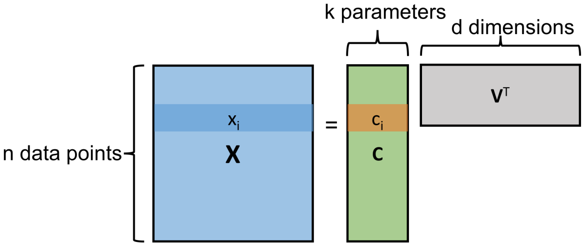
LOW-RANK FACTORIZATION

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



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

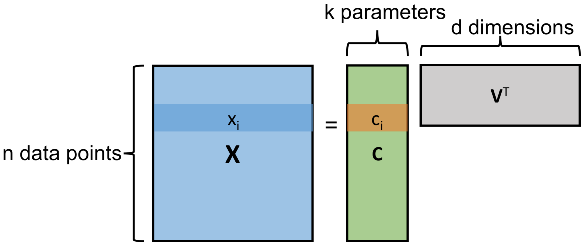
Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie in a k-dimensional subspace with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as $\mathbf{X} = \mathbf{C}\mathbf{V}^T$.



Exercise: What is this coefficient matrix C? **Hint:** Use that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.

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

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

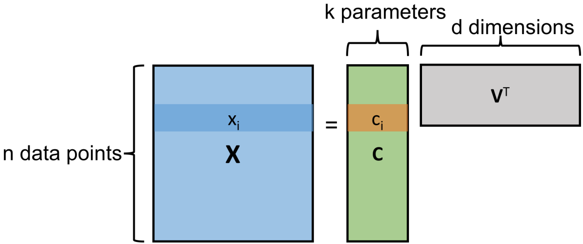


Exercise: What is this coefficient matrix C? **Hint:** Use that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.

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

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

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

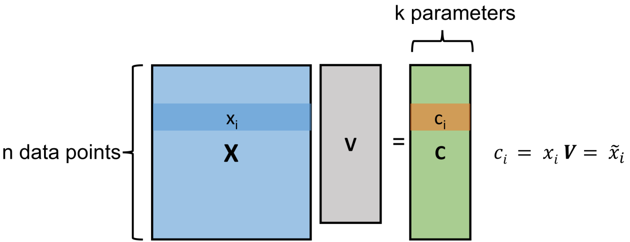


Exercise: What is this coefficient matrix $\mathbf{C}$? **Hint:** Use that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.

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

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as $\mathbf{X} = \mathbf{C}\mathbf{V}^T$.



Exercise: What is this coefficient matrix C? **Hint:** Use that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.

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

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: 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 $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as

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

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

PROJECTION VIEW

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

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

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

PROJECTION VIEW

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

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

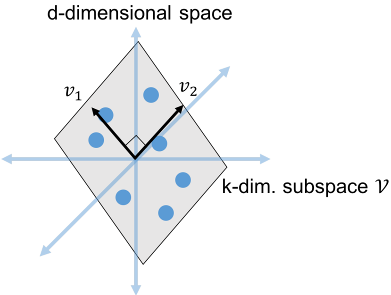
- $\mathbf{V}\mathbf{V}^T$ is a projection matrix, which projects the rows of $\mathbf{X}$ (the data points $\vec{x}_1, \ldots, \vec{x}_n$) onto the subspace $\mathcal{V}$.

$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthonormal 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 $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be written as

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

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



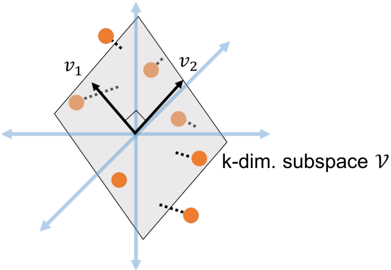
$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthonormal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

PROJECTION VIEW

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

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

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

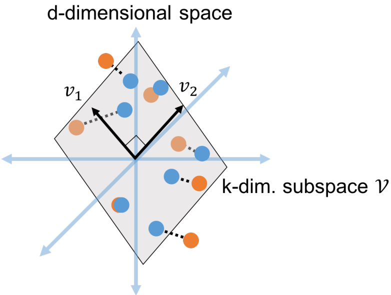


$\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$: data points, $\mathbf{X} \in \mathbb{R}^{n \times d}$: data matrix, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^d$: orthonormal basis for subspace $\mathcal{V}$. $\mathbf{V} \in \mathbb{R}^{d \times k}$: matrix with columns $\vec{v}_1, \ldots, \vec{v}_k$.

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

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

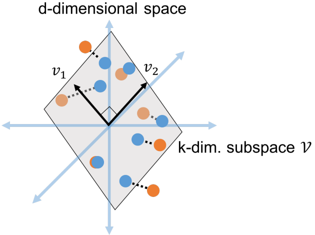
- $\mathbf{V}\mathbf{V}^T$ is a projection matrix, which projects the rows of $\mathbf{X}$ (the data points $\vec{x}_1, \ldots, \vec{x}_n$) onto the subspace $\mathcal{V}$.



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

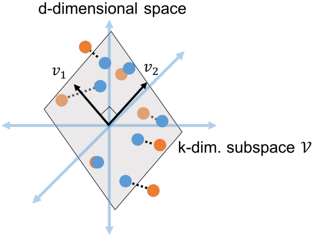
Claim: If $\vec{x}_1, \ldots, \vec{x}_n$ lie close to a k-dimensional subspace $\mathcal{V}$ with orthonormal basis $\mathbf{V} \in \mathbb{R}^{d \times k}$, the data matrix can be approximated as:

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



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

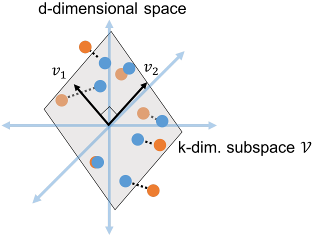
$$\mathbf{X} \approx \mathbf{X}\mathbf{V}\mathbf{V}^T$$



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

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

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



Note: $\mathbf{X}\mathbf{V}\mathbf{V}^T$ has rank k. It is a low-rank approximation of $\mathbf{X}$. Later we'll show that:

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

where $\|\mathbf{A}\|_F$ is defined as $\sqrt{\sum_{i,j} \mathbf{A}_{i,j}^2}$.

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 $\mathcal{V}$).

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 $\mathcal{V}$).

- Letting $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i, (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j$ be the i^{th} and j^{th} projected data points, i.e., the ith and jth rows of $\mathbf{X}\mathbf{V}\mathbf{V}^T$:

  $$\|(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i - (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j\|_2 = \|[(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j]\mathbf{V}^T\|_2 = \|(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j\|_2.$$

  The first equality uses $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i = (\mathbf{X}\mathbf{V})_i\mathbf{V}^T$, $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_j = (\mathbf{X}\mathbf{V})_j\mathbf{V}^T$ and the second equality uses the orthonormality of the columns of $\mathbf{V}$, i.e., for any row vector $\mathbf{a}$

  $$\|\mathbf{a}\mathbf{V}^T\|_2^2 = (\mathbf{a}\mathbf{V}^T)(\mathbf{a}\mathbf{V}^T)^T = \mathbf{a}\mathbf{V}^T\mathbf{V}\mathbf{a}^T = \mathbf{a}\mathbf{a}^T = \|\mathbf{a}\|_2^2$$

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 $\mathcal{V}$).

- Letting $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i, (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j$ be the i^{th} and j^{th} projected data points, i.e., the ith and jth rows of $\mathbf{X}\mathbf{V}\mathbf{V}^T$:

$$\|(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i - (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j\|_2 = \|[(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j]\mathbf{V}^T\|_2 = \|(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j\|_2.$$

  The first equality uses $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i = (\mathbf{X}\mathbf{V})_i\mathbf{V}^T$, $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_j = (\mathbf{X}\mathbf{V})_j\mathbf{V}^T$ and the second equality uses the orthonormality of the columns of $\mathbf{V}$, i.e., for any row vector $\mathbf{a}$

$$\|\mathbf{a}\mathbf{V}^T\|_2^2 = (\mathbf{a}\mathbf{V}^T)(\mathbf{a}\mathbf{V}^T)^T = \mathbf{a}\mathbf{V}^T\mathbf{V}\mathbf{a}^T = \mathbf{a}\mathbf{a}^T = \|\mathbf{a}\|_2^2$$

- Can use $\mathbf{X}\mathbf{V} \in \mathbb{R}^{n \times k}$ as a compressed approximate data set.

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 $\mathcal{V}$).

- Letting $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i, (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j$ be the i^{th} and j^{th} projected data points, i.e., the ith and jth rows of $\mathbf{X}\mathbf{V}\mathbf{V}^T$:

  $$\|(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i - (\mathbf{X}\mathbf{V}\mathbf{V}^T)_j\|_2 = \|[(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j]\mathbf{V}^T\|_2 = \|(\mathbf{X}\mathbf{V})_i - (\mathbf{X}\mathbf{V})_j\|_2.$$

  The first equality uses $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_i = (\mathbf{X}\mathbf{V})_i\mathbf{V}^T$, $(\mathbf{X}\mathbf{V}\mathbf{V}^T)_j = (\mathbf{X}\mathbf{V})_j\mathbf{V}^T$ and the second equality uses the orthonormality of the columns of $\mathbf{V}$, i.e., for any row vector $\mathbf{a}$

  $$\|\mathbf{a}\mathbf{V}^T\|_2^2 = (\mathbf{a}\mathbf{V}^T)(\mathbf{a}\mathbf{V}^T)^T = \mathbf{a}\mathbf{V}^T\mathbf{V}\mathbf{a}^T = \mathbf{a}\mathbf{a}^T = \|\mathbf{a}\|_2^2$$

- Can use $\mathbf{X}\mathbf{V} \in \mathbb{R}^{n \times k}$ as a compressed approximate data set.

Next Time: How to find the subspace $\mathcal{V}$ and correspondingly $\mathbf{V}$.

a step back: why low-rank approximation?

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

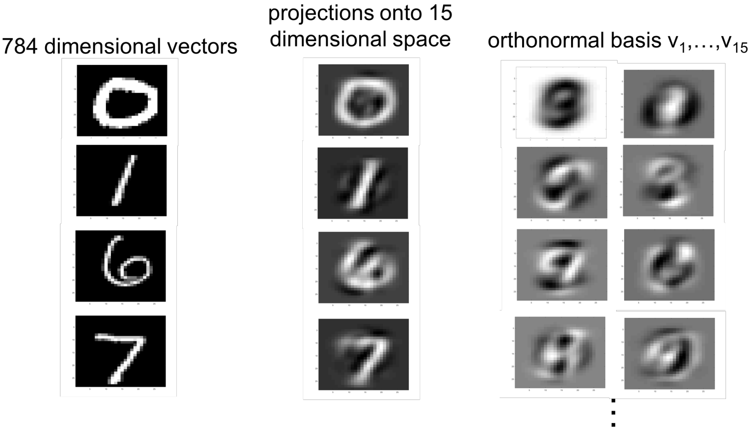
A STEP BACK: WHY LOW-RANK APPROXIMATION?

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

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

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

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



DUAL VIEW OF LOW-RANK APPROXIMATION

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

DUAL VIEW OF LOW-RANK APPROXIMATION

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

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

DUAL VIEW OF LOW-RANK APPROXIMATION

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

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

Linearly Dependent Variables:

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

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

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

Linearly Dependent Variables:

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

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

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

Linearly Dependent Variables:

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

DUAL VIEW OF LOW-RANK APPROXIMATION

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

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

Linearly Dependent Variables:

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

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