Math Refresher for DS

Practical Session 6

girafe
ai

Linear Transformations

- Every $n \times n$ matrix A represents a linear transformation of $\mathbb{R}^n$.
- Columns of A = what happens to the basis vectors.

Linear Transformations

- Another example:

$$A = \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix}$$

Linear Transformations

- Another example:

$$A = \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix}$$

Every vector in $\mathbb{R}^2$ is mapped onto a line, a one-dimensional subspace of $\mathbb{R}^2$ (but we still stay in $\mathbb{R}^2$)

$$Ax = \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 4x_2 \\ 0 \end{bmatrix}$$

girafe ai

Linear Transformations

- Another example:

$$A = \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix}$$

Every vector in $\mathbb{R}^2$ is mapped onto a line, a one-dimensional subspace of $\mathbb{R}^2$ (but we still stay in $\mathbb{R}^2$)

$$Ax = \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 4x_2 \\ 0 \end{bmatrix}$$

Many vectors are mapped onto the same one (no inverse!):

Example: $\begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 14 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 0 \end{bmatrix}$

Linear Transformations

So far: only square matrices.

But about non-square ones?

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} - 3 \times 2 \text{ matrix.}$$

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} - 3 \times 2 \text{ matrix.}$$

$$x \in \mathbb{R}^2, \qquad Ax = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} =$$

girafe ai

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} - 3 \times 2 \text{ matrix.}$$

$$x \in \mathbb{R}^2, \qquad Ax = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 14 \end{bmatrix} \in \mathbb{R}^3$$

girafe ai

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} - 3 \times 2 \text{ matrix.}$$

$$x \in \mathbb{R}^2, \qquad Ax = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 14 \end{bmatrix} \in \mathbb{R}^3$$

A is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$!

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} - 3 \times 2 \text{ matrix.}$$

$$x \in \mathbb{R}^2, \qquad Ax = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 14 \end{bmatrix} \in \mathbb{R}^3$$

A is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$!

$rank(A) = 2$: vectors that were <u>independent</u> in $\mathbb{R}^2$ will be mapped on <u>independent</u> vectors in $\mathbb{R}^3$.

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} - 2 \times 3 \text{ matrix.}$$

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} - 2 \times 3 \text{ matrix.}$$

$$x \in \mathbb{R}^3, \qquad Ax = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} =$$

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} - 2 \times 3 \text{ matrix.}$$

$$x \in \mathbb{R}^3, \qquad Ax = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 15 \end{bmatrix} \in \mathbb{R}^2$$

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} - 2 \times 3 \text{ matrix.}$$

$$x \in \mathbb{R}^3, \qquad Ax = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 15 \end{bmatrix} \in \mathbb{R}^2$$

A is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$!

Linear Transformations

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} - 2 \times 3 \text{ matrix.}$$

$$x \in \mathbb{R}^3, \qquad Ax = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 15 \end{bmatrix} \in \mathbb{R}^2$$

A is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$!

$rank(A) = 2$: vectors that were <u>independent</u> in $\mathbb{R}^3$ may be mapped on <u>dependent</u> vectors in $\mathbb{R}^2$.

Linear Transformations

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} - m \times n \text{ matrix.}$$

$$x \in \mathbb{R}^n, \qquad Ax = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$$

A is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$!

Linear Transformations From $\mathbb{R}^n$ to $\mathbb{R}^m$

Why this is useful?

Dimensionality reduction!

Dimensionality Reduction

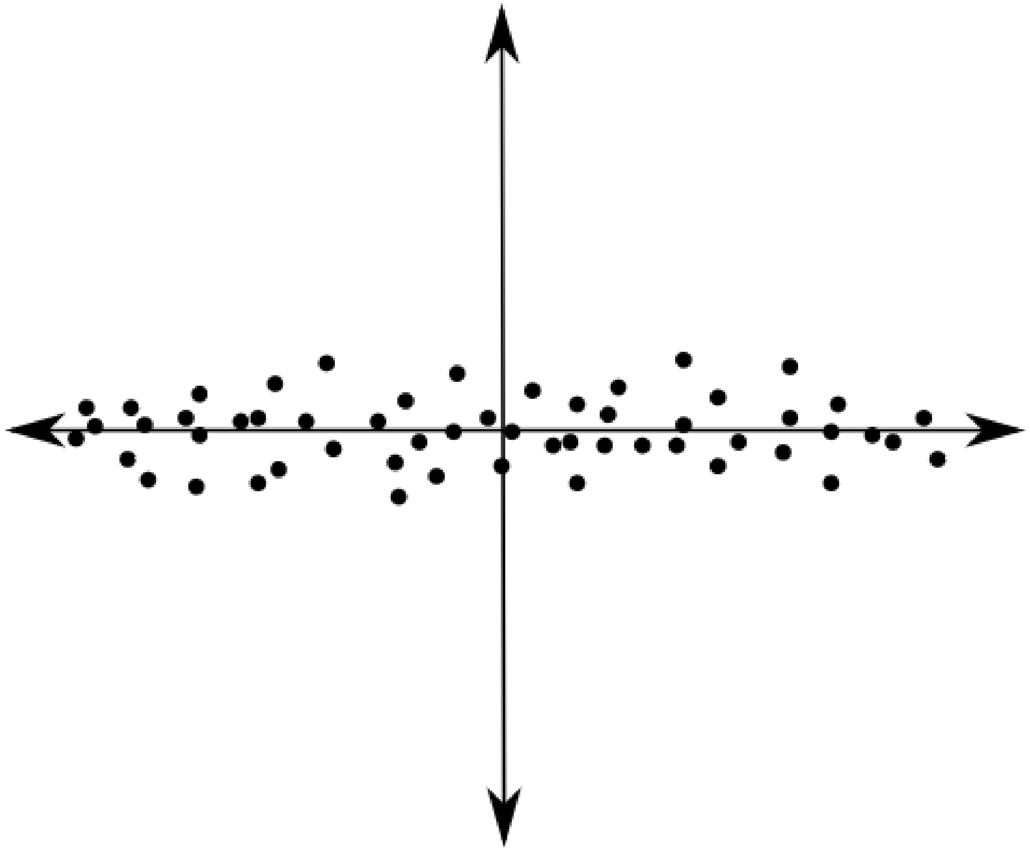
- Imagine that you have some data: m features, n examples.
- Each example $x = (x_1, \dots, x_m)$ – a point in $\mathbb{R}^m$.
- How to visualize this data?

Dimensionality Reduction

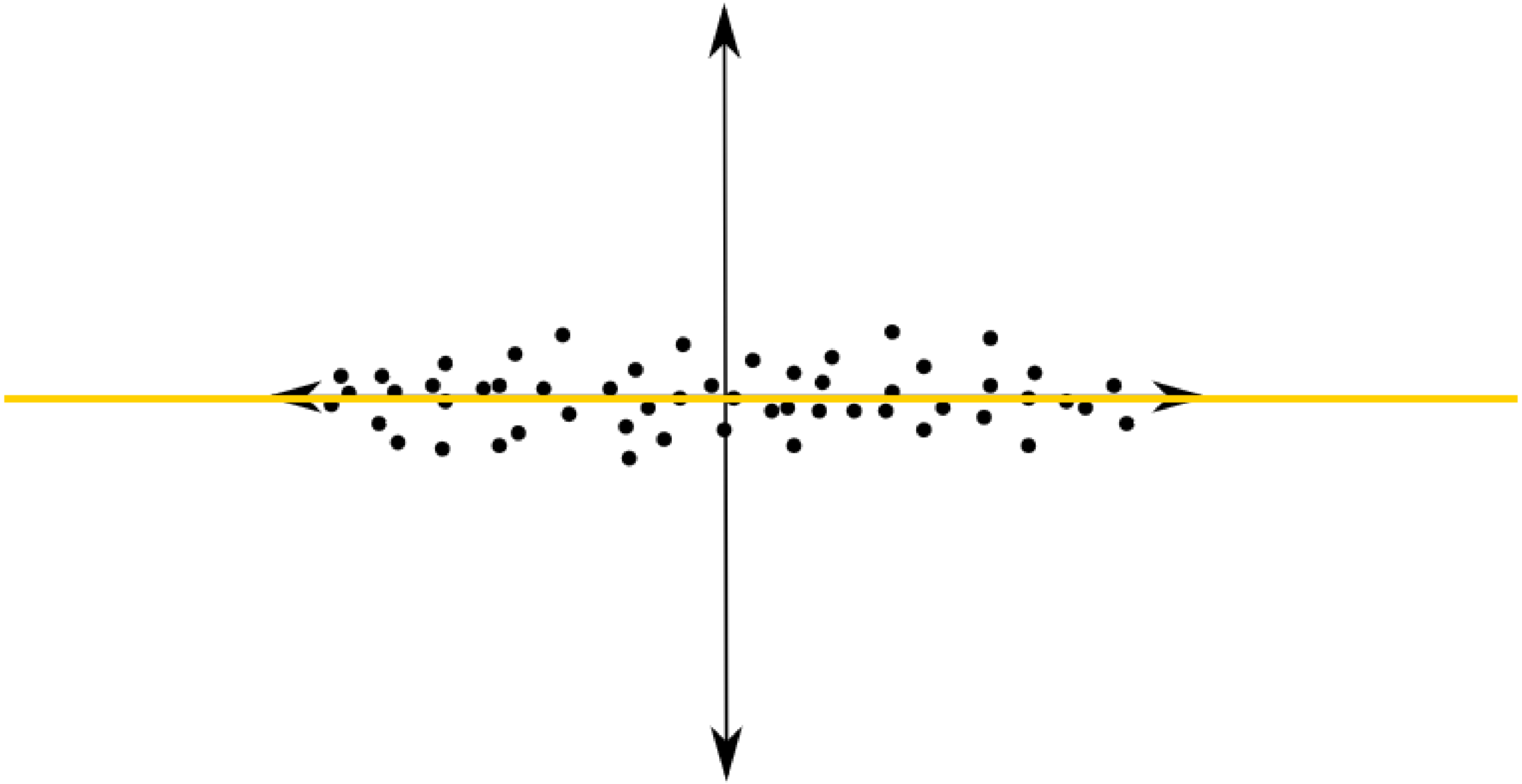
- Imagine that you have some data: m features, n examples.
- Each example $x = (x_1, \ldots, x_m)$ – a point in $\mathbb{R}^m$.
- How to visualize this data?

Map into onto a lower-dimensional space!
But preserve as much variance in data as possible.

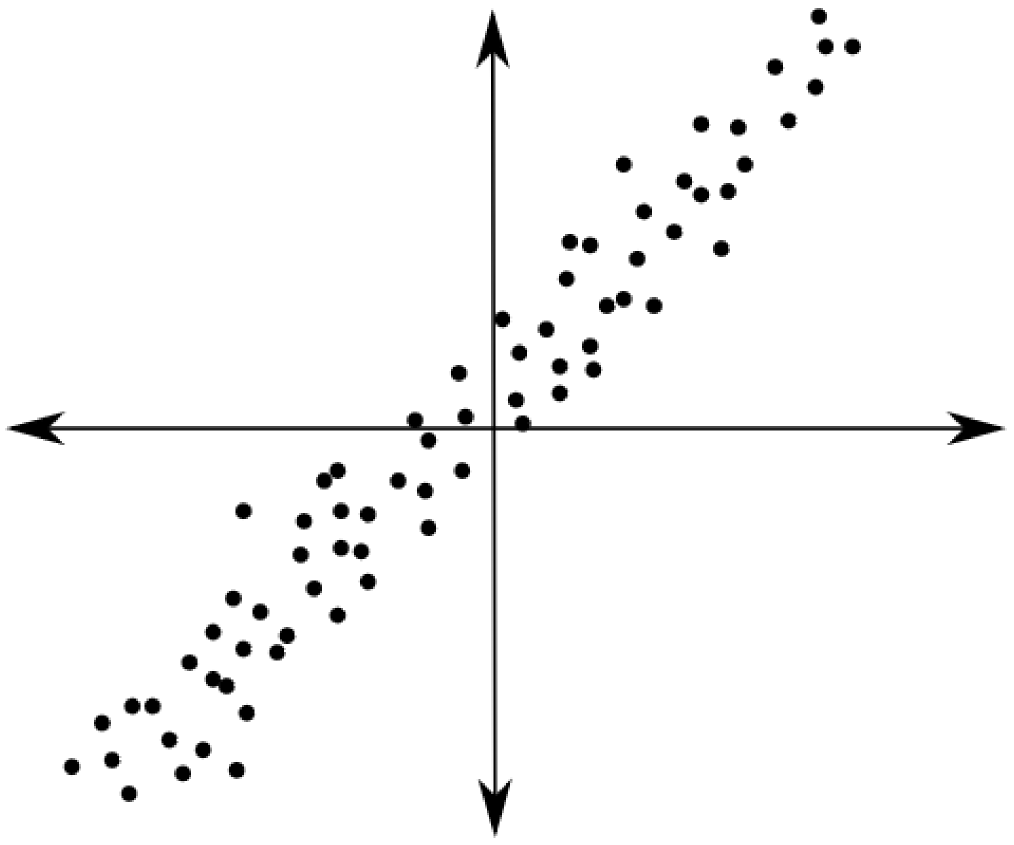
Dimensionality Reduction



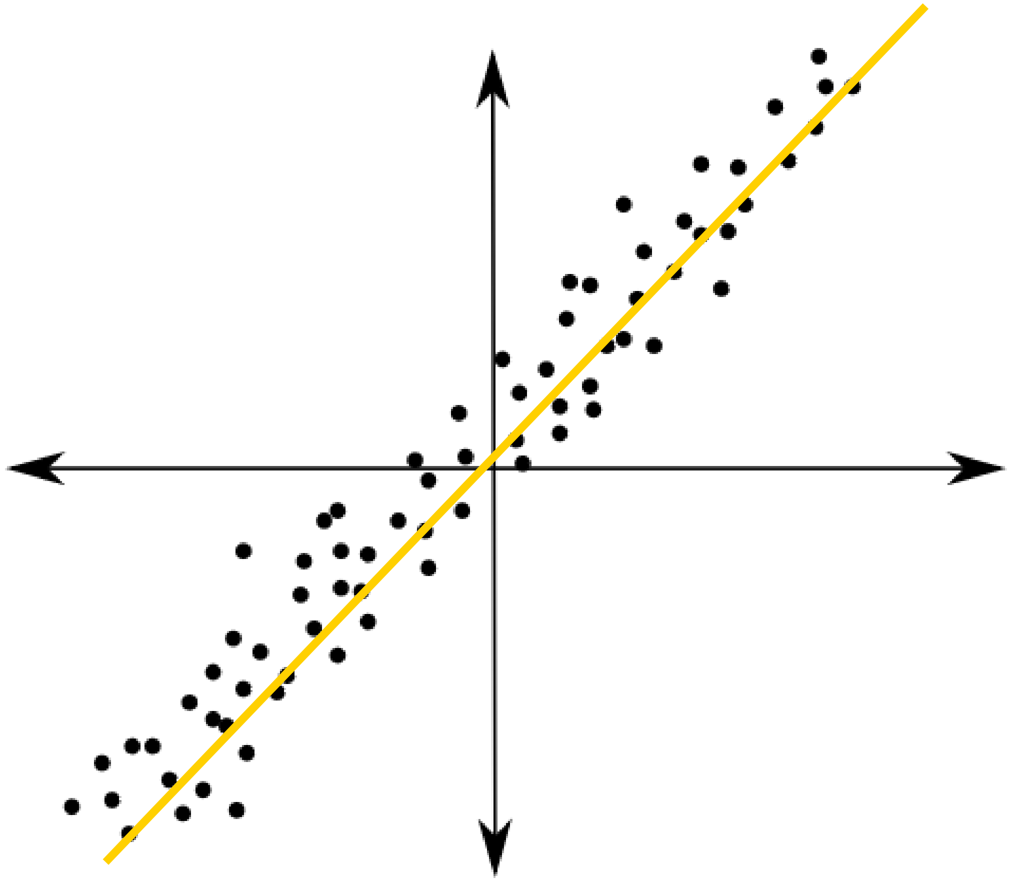
Dimensionality Reduction



Dimensionality Reduction

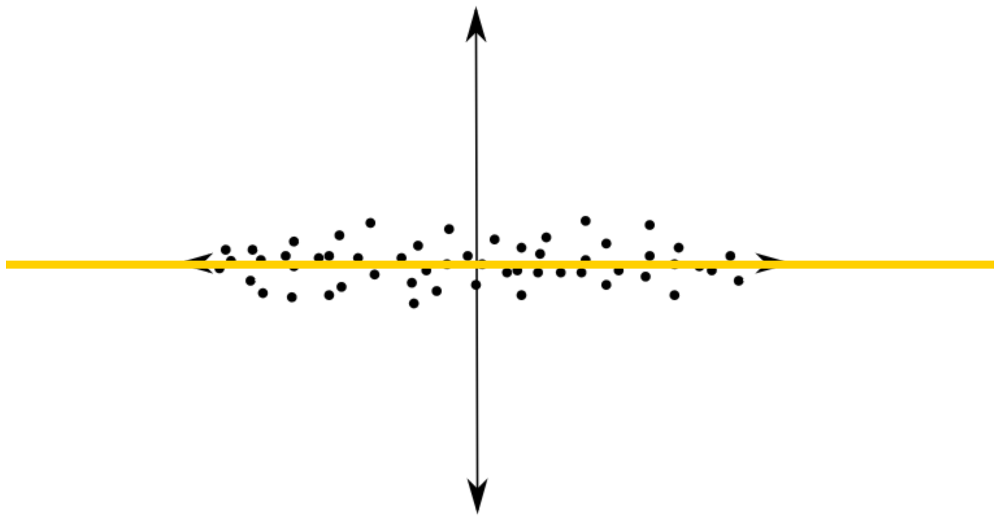


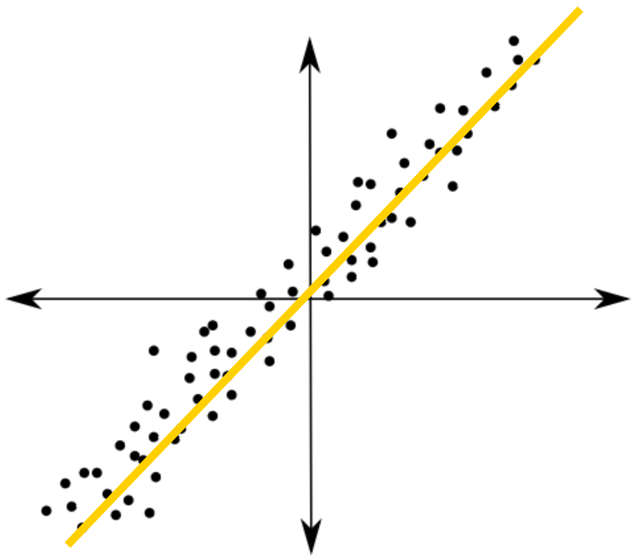
Dimensionality Reduction



Dimensionality Reduction

- How to find this direction?





Some theory first...

girafe
ai

Eigenvalues & Eigenvectors

- Consider an $n \times n$ matrix A.
- A – linear transformation of $\mathbb{R}^n$. Every vector gets scaled and rotated:

$$x' = Ax$$

Eigenvalues & Eigenvectors

- Consider an $n \times n$ matrix A.
- A – linear transformation of $\mathbb{R}^n$. Every vector gets scaled and rotated:

$$x' = Ax$$

- Some vectors only get scaled:

$$Av = \lambda v, \qquad \lambda \in \mathbb{R}, \qquad v \neq 0$$

$v \neq 0$ – eigenvector, λ – corresponding eigenvalue.

Eigenvalues & Eigenvectors

- Consider an $n \times n$ matrix A.
- $v \neq 0$ – eigenvector, λ – corresponding eigenvalue:

$$Av = \lambda v$$

How to find λ and v?

$$Av = \lambda v \Leftrightarrow Av - \lambda v = 0 \Leftrightarrow (A - \lambda E)v = 0 \Leftrightarrow$$

Eigenvalues & Eigenvectors

- Consider an $n \times n$ matrix A.
- $v \neq 0$ – eigenvector, λ – corresponding eigenvalue:

$$Av = \lambda v$$

How to find λ and v?

$$Av = \lambda v \Leftrightarrow Av - \lambda v = 0 \Leftrightarrow (A - \lambda E)v = 0 \Leftrightarrow$$
$$\det(A - \lambda E) = 0$$

$\det(A - \lambda E)$ – characteristic polynomial of A.

girafe ai

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

- Characteristic polynomial:

$$\det(A - \lambda E) =$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$
- Let's find eigenvalues and eigenvectors of A.
- Characteristic polynomial:

$$\det(A - \lambda E) = \begin{vmatrix} -5-\lambda & 2 \\ -7 & 4-\lambda \end{vmatrix} =$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$
- Let's find eigenvalues and eigenvectors of A.
- Characteristic polynomial:

$$\det(A - \lambda E) = \begin{vmatrix} -5-\lambda & 2 \\ -7 & 4-\lambda \end{vmatrix} = (-5-\lambda)(4-\lambda) + 14 = 0 \Leftrightarrow$$

$$\lambda^2 + \lambda - 6 = 0 \Leftrightarrow$$

girafe ai

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$
- Let's find eigenvalues and eigenvectors of $\boldsymbol{A}$.
- Characteristic polynomial:

$$\det(A - \lambda E) = \begin{vmatrix} -5-\lambda & 2 \\ -7 & 4-\lambda \end{vmatrix} = (-5-\lambda)(4-\lambda) + 14 = 0 \Leftrightarrow$$

$$\lambda^2 + \lambda - 6 = 0 \Leftrightarrow$$

$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3}$ – eigenvalues of $\boldsymbol{A}$.

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3}\ \text{ – eigenvalues of } \boldsymbol{A}.$$

$$Av_1 = \lambda_1 v_1 \Leftrightarrow$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3} \ \text{ – eigenvalues of } \boldsymbol{A}.$$

$$Av_1 = \lambda_1 v_1 \Leftrightarrow \begin{bmatrix} -5-2 & 2 \\ -7 & 4-2 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = 0 \ \Leftrightarrow$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3} \ \text{ – eigenvalues of } \boldsymbol{A}.$$

$$Av_1 = \lambda_1 v_1 \Leftrightarrow \begin{bmatrix} -5-2 & 2 \\ -7 & 4-2 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = 0 \Leftrightarrow v_{11} = \frac{2}{7} v_{21}$$

For example, $v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$.

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3}\ \text{ – eigenvalues of } \boldsymbol{A}.$$

$$Av_2 = \lambda_2 v_2 \Leftrightarrow$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3}$ – eigenvalues of $\boldsymbol{A}$.

$$Av_2 = \lambda_2 v_2 \Leftrightarrow \begin{bmatrix} -5+3 & 2 \\ -7 & 4+3 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 0$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Let's find eigenvalues and eigenvectors of A.

$$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3} \text{ – eigenvalues of } \boldsymbol{A}.$$

$$Av_2 = \lambda_2 v_2 \Leftrightarrow \begin{bmatrix} -5+3 & 2 \\ -7 & 4+3 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 0 \Leftrightarrow v_{12} = v_{22}$$

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$
- Let's find eigenvalues and eigenvectors of A.

$\boldsymbol{\lambda_1 = 2,\ \lambda_2 = -3}$ – eigenvalues of $\boldsymbol{A}$.

$$Av_2 = \lambda_2 v_2 \Leftrightarrow \begin{bmatrix} -5+3 & 2 \\ -7 & 4+3 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 0 \Leftrightarrow v_{12} = v_{22}$$

For example, $v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$
- Eigenvalues and eigenvectors of A:

$$\lambda_1 = 2, \qquad v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

$$\lambda_2 = -3, \qquad v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

girafe ai

Eigenvalues & Eigenvectors

- Example: $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$

- Eigenvalues and eigenvectors of A:

$$\lambda_1 = 2, \qquad v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

$$\lambda_2 = -3, \qquad v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Note that v_1 *and* v_2 *are linearly independent.*

girafe ai

Eigenbasis

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- v_1 and v_2 are linearly independent. They form eigenbasis.
  What happens to A if we change to basis $\{v_1, v_2\}$?

Eigenbasis

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- v_1 and v_2 are linearly independent. They form **eigenbasis**.
  What happens to A if we change to basis $\{v_1, v_2\}$?

$[A]_V$ = {what happens to basis vectors v_1, v_2 after applying A} =

Eigenbasis

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$

- v_1 and v_2 are linearly independent. They form **eigenbasis**.
  What happens to A if we change to basis $\{v_1, v_2\}$?

$$[A]_V = \{\text{what happens to basis vectors } v_1, v_2 \text{ after applying } A\} =$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix} = \Lambda - \text{it becomes diagonal!}$$

girafe ai

Eigendecomposition

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- How do we change to basis $\{v_1, v_2\}$?

Eigendecomposition

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- How do we change to basis $\{v_1, v_2\}$?

$V = [v_1 \mid v_2]$ – transition from standard to eigenbasis.

$$A = V[A]_V V^{-1}$$

$A = V\Lambda V^{-1}$ – eigendecomposition.

Eigendecomposition

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- How do we change to basis $\{v_1, v_2\}$?

$A = V\Lambda V^{-1}$ – eigendecomposition.

Let's check this:

$$\begin{bmatrix} 2 & 1 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 7 & 1 \end{bmatrix}^{-1} =$$

Eigendecomposition

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

- Eigenvalues and eigenvectors: $\lambda_1 = 2,\ v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \quad \lambda_2 = -3,\ v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$
- How do we change to basis $\{v_1, v_2\}$?

$A = V\Lambda V^{-1}$ – eigendecomposition.

Let's check this:

$$\begin{bmatrix} 2 & 1 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 7 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 1 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} -0.2 & 0.2 \\ 1.4 & -0.4 \end{bmatrix} = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$$

Eigendecomposition

Is it always possible to find an eigenbasis?

No ☹ (see lectures for examples).

But there are good news ☺

The Spectral Theorem

If A is an $n \times n$ symmetric matrix,
then A always n linearly independent eigenvectors $v_1, \ldots, v_n$.

What's more, $v_1, \ldots, v_n$ are mutually orthogonal!
Since we are choosing the scaling, we can make them orthonormal.

The Spectral Theorem

If A is an $n \times n$ symmetric matrix,
then A always n linearly independent eigenvectors $v_1, \ldots, v_n$.

What's more, $v_1, \ldots, v_n$ are mutually orthogonal!
Since we are choosing the scaling, we can make them orthonormal.

So, we can always decompose square symmetric matrices as

$A = V\Lambda V^{-1}$, where

V – orthogonal matrix (columns = eigenvecotrs),

Λ – diagonal matrix (diagonal elements = eigenvalues)

girafe ai

Orthogonal Matrices

- A matrix where all columns are mutually orthonormal:

$$A^T A = AA^T = E$$

Orthogonal Matrices

- A matrix where all columns are mutually orthonormal:

$$A^T A = AA^T = E$$

- That means that orthogonal matrices are easy to invert:

$$A^{-1} = A^T.$$

The Spectral Theorem

If A is an $n \times n$ symmetric matrix,
then A always n linearly independent eigenvectors $v_1, \dots, v_n$.

What's more, $v_1, \dots, v_n$ are mutually orthogonal!
Since we are choosing the scaling, we can make them orthonormal.

So, we can always decompose square symmetric matrices as

$\boldsymbol{A = V\Lambda V^{-1} = V\Lambda V^T}$=, where

V – orthogonal matrix (columns = eigenvecotrs),

Λ – diagonal matrix (diagonal elements = eigenvalues)

Back to Dimensionality Reduction

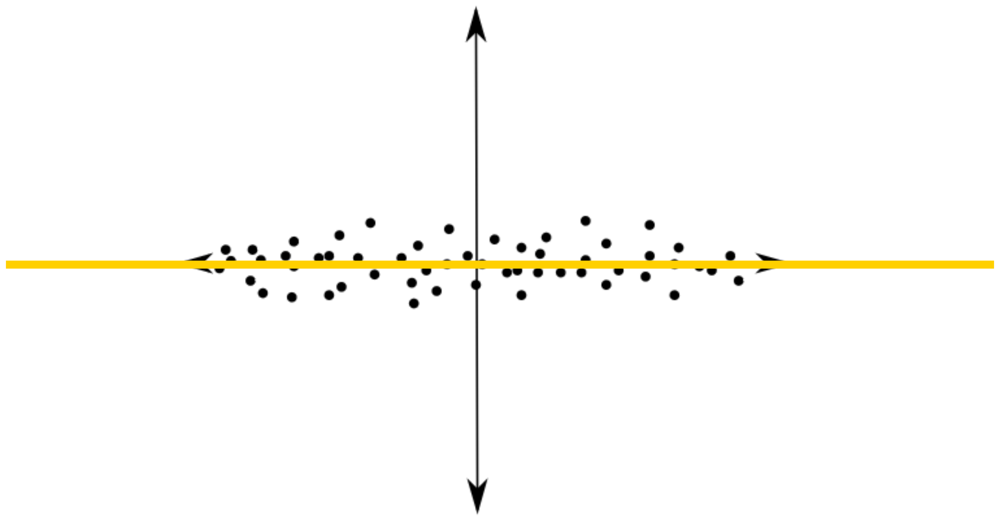
girafe
ai

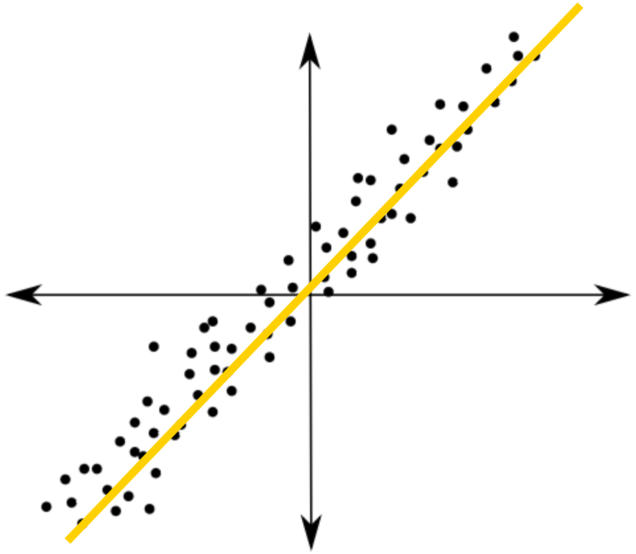
Dimensionality Reduction

- Imagine that you have some data: m features, n examples.
- Each example $x = (x_1, \dots, x_m)$ – a point in $\mathbb{R}^m$.
- How to visualize this data?
- Map into onto a lower-dimensional space!
  But preserve as much variance in data as possible.

Dimensionality Reduction

- How to find this direction?





PCA

- Let's construct data covariance matrix.

$$S = \frac{1}{n-1} XX^T - m \times m \text{ symmetric matrix.}$$

s_{ij} – covariance between features i and j,

s_{ii} – variance of feature i

PCA

- Let's construct data covariance matrix.

$$S = \frac{1}{n-1} XX^T - m \times m \text{ symmetric matrix.}$$

s_{ij} – covariance between features i and j,

s_{ii} – variance of feature i

- We can apply eigendecomposition to S:

$$S = V\Lambda V^T$$

PCA

$$S = V\Lambda V^{-1} = V\Lambda V^{T}$$

$$\begin{bmatrix} \boldsymbol{s_{11}} & s_{12} & \cdots & s_{1m} \\ s_{21} & \boldsymbol{s_{22}} & \cdots & s_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ s_{m1} & s_{m2} & \cdots & \boldsymbol{s_{mm}} \end{bmatrix} =$$

$$= \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda_m} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix}^{T}$$

PCA

$$S = V\Lambda V^{-1} = V\Lambda V^{T}$$

$$\begin{bmatrix} \boldsymbol{s_{11}} & s_{12} & \cdots & s_{1m} \\ s_{21} & \boldsymbol{s_{22}} & \cdots & s_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ s_{m1} & s_{m2} & \cdots & \boldsymbol{s_{mm}} \end{bmatrix} =$$

$$= \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda_m} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix}^{T}$$

Total variance of the data

$$\mathrm{T} = tr(S) = s_{11} + \cdots + s_{nn} = \lambda_1 + \cdots \lambda_m$$

Projecting the Data

- What happens if we project the data onto the eigenvectors of S? Remember: $\{v_1, .., v_m\}$ is an orthonormal basis!

Projecting the Data

- What happens if we project the data onto the eigenvectors of S? Remember: $\{v_1, .., v_m\}$ is an orthonormal basis!

$$x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n,$$

Projecting the Data

- What happens if we project the data onto the eigenvectors of S? Remember: $\{v_1, .., v_m\}$ is an orthonormal basis!

$$x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n, \qquad (x, v_i) = x_i (v_i, v_i) + 0 = x_i$$

Coordinates in an orthonormal basis = dot products!

Projecting the Data

- What happens if we project the data onto the eigenvectors of S? Remember: $\{v_1, .., v_m\}$ is an orthonormal basis!

$$x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n, \qquad (x, v_i) = x_i (v_i, v_i) + 0 = x_i$$

Coordinates in an orthonormal basis = dot products!

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \text{– one example,} \qquad x_{proj} = V^T x$$

Projecting the Data

- What happens if we project the data onto the eigenvectors of S? Remember: $\{v_1, .., v_m\}$ is an orthonormal basis!

$$x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n, \qquad (x, v_i) = x_i (v_i, v_i) + 0 = x_i$$

Coordinates in an orthonormal basis = dot products!

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \text{– one example,} \qquad x_{proj} = V^T x$$

The whole data: $X_{proj} = V^T X$

Projecting the Data

- How to project the data onto $p < m$ eigenvectors of S?

$$V_p = \{v_1 \mid \dots \mid v_p\}$$

$$X_p = V_p^T X$$

Projecting the Data

- How to project the data onto $p < m$ eigenvectors of S?

$$V_p = \{v_1 \mid \dots \mid v_p\}$$

$$X_p = V_p^T X$$

- What happens to the covariance matrix?

$$S_{proj} \sim X_p X_p^T = V_p^T X X^T V_p$$

Projecting the Data

- How to project the data onto $p < m$ eigenvectors of S?

$$V_p = \{v_1 \mid \dots \mid v_p\}$$

$$X_p = V_p^T X$$

- What happens to the covariance matrix?

$$S_{proj} \sim X_p X_p^T = V_p^T X X^T V_p \sim V_p^T (S V_p) =$$

Projecting the Data

- How to project the data onto $p < m$ eigenvectors of S?

$$V_p = \{v_1 \mid \dots \mid v_p\}$$

$$X_p = V_p^T X$$

- What happens to the covariance matrix?

$$S_{proj} \sim X_p X_p^T = V_p^T X X^T V_p \sim V_p^T (S V_p) = V_p^T \Lambda_p V_p$$

Projecting the Data

- How to project the data onto $p < m$ eigenvectors of S?

$$V_p = \{v_1 \mid \ldots \mid v_p\}$$

$$X_p = V_p^T X$$

- What happens to the covariance matrix?

$$S_{proj} \sim X_p X_p^T = V_p^T X X^T V_p \sim V_p^T (S V_p) = V_p^T \Lambda_p V_p$$

Total variance of the projected data:

$$T = tr\left(\Lambda_p\right) = \lambda_1 + \cdots \lambda_p.$$

PCA

$$S_p = V_{\mathrm{p}}\Lambda_{\mathrm{p}}V_p^T$$

$$\begin{bmatrix} \boldsymbol{s_{11}} & s_{12} & \cdots & s_{1m} \\ s_{21} & \boldsymbol{s_{22}} & \cdots & s_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ s_{p1} & s_{p2} & \cdots & \boldsymbol{s_{pp}} \end{bmatrix} =$$

$$= \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mp} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda_p} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mp} \end{bmatrix}^T$$

Total variance of the projected data

$$\mathrm{T} = tr(S) = s_{11} + \cdots + s_{pp} = \lambda_1 + \cdots \lambda_p$$

girafe ai

PCA

- Let's order eigenvalues and eigenvectors so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$.

Projecting on the first p eigenvectors =
explaining $(\lambda_1 + \cdots + \lambda_p)$ out of the total variance.

$$\begin{bmatrix} \boldsymbol{s_{11}} & s_{12} & \cdots & s_{1m} \\ s_{21} & \boldsymbol{s_{22}} & \cdots & s_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ s_{p1} & s_{p2} & \cdots & \boldsymbol{s_{pp}} \end{bmatrix} =$$

$$= \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mp} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda_p} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mp} \end{bmatrix}^T$$

Let’s practice!

https://colab.research.google.com/drive/1tx5IXfGheU4fZRN5OkkBHAU4mbna0s7R?usp=sharing