Machine Learning

PCA, spectra, and low-rank approximations

Emanuele Rodolà
rodola@di.uniroma1.it

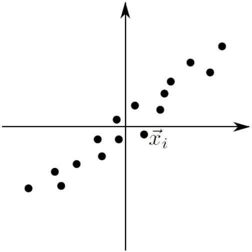
SAPIENZA
UNIVERSITÀ DI ROMA

2nd semester a.y. 2024/2025 · April 08, 2025

Principal component

Principal axis

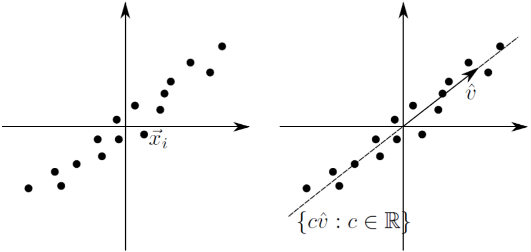
Consider the 2D data:



(a) Input data

Principal axis

Consider the 2D data:



(a) Input data (b) Principal axis

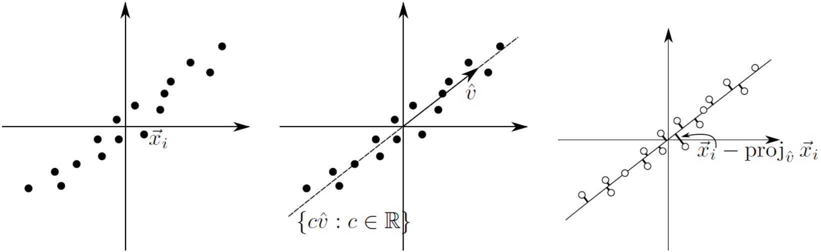
Q: Find the vector $\mathbf{v}$ such that each data point $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i = c_i \mathbf{v}$$

where each $\mathbf{x}_i$ has its own c_i

Principal axis

Consider the 2D data:



(a) Input data (b) Principal axis (c) Projection error

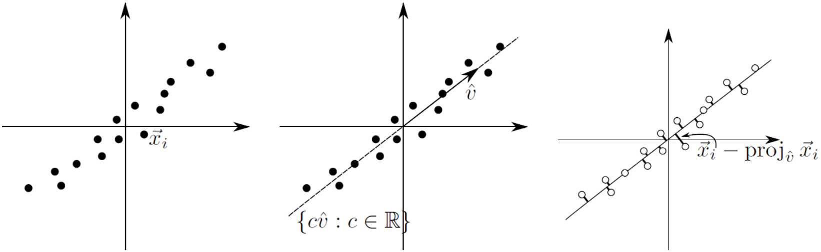
Q: Find the vector $\mathbf{v}$ such that each data point $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i = c_i \mathbf{v}$$

where each $\mathbf{x}_i$ has its own c_i

Principal axis

Consider the 2D data:



(a) Input data (b) Principal axis (c) Projection error

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \mathrm{proj}_{\mathbf{v}}\mathbf{x}_i\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \mathrm{proj}_{\mathbf{v}}\mathbf{x}_i\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \frac{\mathbf{x}_i^\top \mathbf{v}}{\mathbf{v}^\top \mathbf{v}} \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v})^\top (\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v})$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\begin{aligned} \min_{\mathbf{v}} \ & \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^\top \mathbf{v})(\mathbf{x}_i^\top \mathbf{v}) + (\mathbf{x}_i^\top \mathbf{v})^2 \|\mathbf{v}\|_2^2) \\ \text{s.t.} \ & \|\mathbf{v}\|_2 = 1 \end{aligned}$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^\top \mathbf{v})^2 + (\mathbf{x}_i^\top \mathbf{v})^2 \|\mathbf{v}\|_2^2)$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - (\mathbf{x}_i^\top \mathbf{v})^2)$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \ -\sum_i (\mathbf{x}_i^\top \mathbf{v})^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \; -\|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

Principal axis

$$\max_{\mathbf{v}} \|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

This can also be written as

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\max_{\mathbf{v}} \|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

This can also be written as

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

The global maximizer $\mathbf{v}^*$ of this problem is the principal component of the data contained in $\mathbf{X}$.

Eigenvectors and eigenvalues

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A} c\mathbf{x} = c\mathbf{A}\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

For this reason, we can restrict our search to eigenvectors with $\|\mathbf{x}\|_2 = 1$.

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

For this reason, we can restrict our search to eigenvectors with $\|\mathbf{x}\|_2 = 1$.

Clearly, $\mathbf{x}$ and $-\mathbf{x}$ are both eigenvectors with the same eigenvalue.

Eigenvalue equation

Suggestion:
Don't just memorize the expression, understand its implications!

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

Example:

$$\frac{d}{dx} e^{ax} = a e^{ax}$$

The exponential is an eigenfunction of the derivative operator!

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$$

In other words, $\mathbf{x}$ is an eigenvector of $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ with eigenvalue λ.

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$$

In other words, $\mathbf{x}$ is an eigenvector of $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ with eigenvalue λ.

We say that $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ is a similarity transformation. $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues.

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$\mathbf{Q}\mathbf{x} = \lambda \mathbf{x}$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$\|\mathbf{Qx}\|_2^2 = \|\lambda \mathbf{x}\|_2^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$(\mathbf{Qx})^\top \mathbf{Qx} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$\mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q} \mathbf{x} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$\mathbf{x}^\top \mathbf{x} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$\|\mathbf{x}\|_2^2 = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$1 = |\lambda|^2$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  Observe:

$$1 = |\lambda|$$

More basic facts

A few more useful facts:

- Orthogonal matrices

  We obtained:

$$\lambda = \pm 1$$

- Diagonal and upper-triangular matrices

  The eigenvalues are the entries along the main diagonal.

More basic facts

A few more useful facts:

- Orthogonal matrices

  We obtained:

$$\lambda = \pm 1$$

- Diagonal and upper-triangular matrices

  The eigenvalues are the entries along the main diagonal.

- Commuting matrices

  Consider two matrices $\mathbf{A}$ and $\mathbf{B}$. One can prove:

$$\mathbf{AB} = \mathbf{BA} \quad \Leftrightarrow \quad \mathbf{A} \text{ and } \mathbf{B} \text{ have the same eigenvectors}$$

Big questions

- How to compute eigenvalues and eigenvectors?

Big questions

- How to compute eigenvalues and eigenvectors?
- What to do with them?

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \mathbf{A}\sum_i \alpha_i \mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \sum_i \alpha_i \mathbf{A}\mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ means scaling $\mathbf{x}$.

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

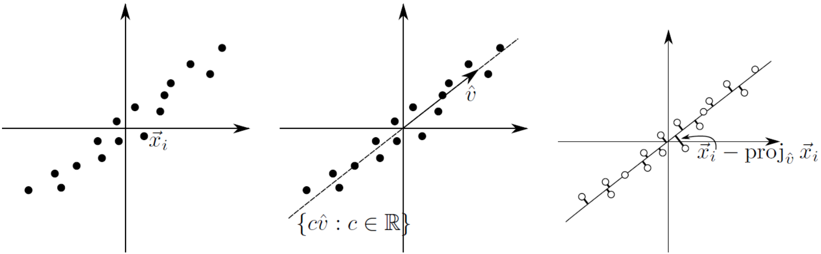
- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \sum_i \alpha_i \lambda_i \mathbf{x}_i$$

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$



(a) Input data (b) Principal axis (c) Projection error

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector.

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector. More in general:

$$\lambda_{\min} \leq \frac{\mathbf{v}^\top \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2} \leq \lambda_{\max}$$

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A}\mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector. More in general:

$$\lambda_{\min} \leq \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2} \leq \lambda_{\max}$$

The ratio $\frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$ is called Rayleigh quotient.

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\lambda \mathbf{x}^\top \mathbf{x} = (\lambda \mathbf{x})^\top \mathbf{x}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned} \lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\ &= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$
\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x})
\end{aligned}
$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$
\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}}
\end{aligned}
$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$(\mathbf{A}\mathbf{x}_i)^\top \mathbf{x}_j = \mathbf{x}_i^\top \mathbf{A}\mathbf{x}_j$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$(\lambda_i \mathbf{x}_i)^\top \mathbf{x}_j = \mathbf{x}_i^\top \lambda_j \mathbf{x}_j$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$
\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}
$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$
\lambda_i \mathbf{x}_i^\top \mathbf{x}_j = \lambda_j \mathbf{x}_i^\top \mathbf{x}_j
$$

Then it must be $\mathbf{x}_i^\top \mathbf{x}_j = 0$, i.e. $\mathbf{x}_i$ and $\mathbf{x}_j$ are orthogonal.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{AX} = \mathbf{X\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Observe the similarity with our motivational problem.

$$\max_{\mathbf{x}} \mathbf{x}^\top \mathbf{A} \mathbf{x}$$
$$\text{s.t. } \|\mathbf{x}\|_2 = 1$$

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Observe the similarity with our motivational problem. We can modify it to solve for all eigenvectors and eigenvalues:

$$\max_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})$$
$$\text{s.t. } \mathbf{X}^\top \mathbf{X} = \mathbf{I}$$

Finding eigenvalues

Power iteration

Very simple algorithm to find the largest eigenvalue/eigenvector:

```
function NORMALIZED-ITERATION(A)
    v⃗ ← ARBITRARY(n)
    for k ← 1, 2, 3, ...
        w⃗ ← Av⃗
        v⃗ ← w⃗/‖w⃗‖
    return v⃗
```

Power iteration

Very simple algorithm to find the largest eigenvalue/eigenvector:

function $\textsc{Normalized-Iteration}(A)$
  $\vec{v} \leftarrow \textsc{Arbitrary}(n)$
  for $k \leftarrow 1, 2, 3, \ldots$
    $\vec{w} \leftarrow A\vec{v}$
    $\vec{v} \leftarrow \vec{w}/\|\vec{w}\|$
  return $\vec{v}$

The normalization is needed to reduce the numerical error.

Without normalization, it will still converge to the principal eigenvector (but with a very large scale).

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Then, we can just apply the power method to $\mathbf{A}^{-1}$:

function $\textsc{Inverse-Iteration}(A)$
  $\vec{v} \leftarrow \textsc{Arbitrary}(n)$
  for $k \leftarrow 1, 2, 3, \ldots$
    $\vec{w} \leftarrow A^{-1}\vec{v}$
    $\vec{v} \leftarrow \vec{w}/\|\vec{w}\|$
  return $\vec{v}$

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Then, we can just apply the power method to $\mathbf{A}^{-1}$:

function INVERSE-ITERATION(A)
    $\vec{v} \leftarrow$ ARBITRARY(n)
    for $k \leftarrow 1, 2, 3, \ldots$
        $\vec{w} \leftarrow A^{-1}\vec{v}$
        $\vec{v} \leftarrow \vec{w}/\|\vec{w}\|$
    return $\vec{v}$

In practice, you don't invert $\mathbf{A}$ but apply LU decomposition.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma \mathbf{I})\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue
- Compute the shifted matrix $\mathbf{B} = \mathbf{A} - \sigma\mathbf{I}$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue
- Compute the shifted matrix $\mathbf{B} = \mathbf{A} - \sigma\mathbf{I}$
- Apply the inverse iteration on $\mathbf{B}$

Singular Value Decomposition (SVD)

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

By these properties, the map $\mathbf{x} \mapsto \mathbf{Qx}$ is an isometry of $\mathbb{R}^n$.

Isometries

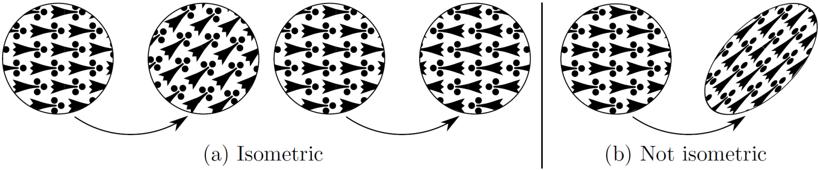
Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

By these properties, the map $\mathbf{x} \mapsto \mathbf{Qx}$ is an isometry of $\mathbb{R}^n$.



(a) Isometric (b) Not isometric

General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$

General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{U}}_{m \times m} \underbrace{\mathbf{\Sigma}}_{m \times n} \underbrace{\mathbf{V}^\top}_{n \times n}$$

General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{U}}_{m \times m} \underbrace{\mathbf{\Sigma}}_{m \times n} \underbrace{\mathbf{V}^\top}_{n \times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices

General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{U}}_{m \times m} \underbrace{\boldsymbol{\Sigma}}_{m \times n} \underbrace{\mathbf{V}^\top}_{n \times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices
- $\boldsymbol{\Sigma}$ is a rectangular diagonal matrix, e.g. $\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$

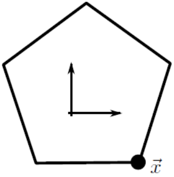
General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{U}}_{m \times m} \underbrace{\mathbf{\Sigma}}_{m \times n} \underbrace{\mathbf{V}^\top}_{n \times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices
- $\mathbf{\Sigma}$ is a rectangular diagonal matrix, e.g. $\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$



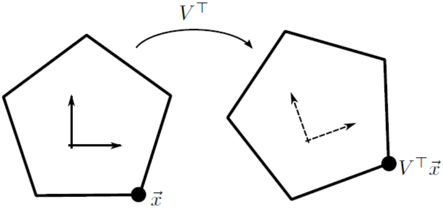
General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{U}}_{m \times m} \underbrace{\mathbf{\Sigma}}_{m \times n} \underbrace{\mathbf{V}^\top}_{n \times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices
- $\mathbf{\Sigma}$ is a rectangular diagonal matrix, e.g. $\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$



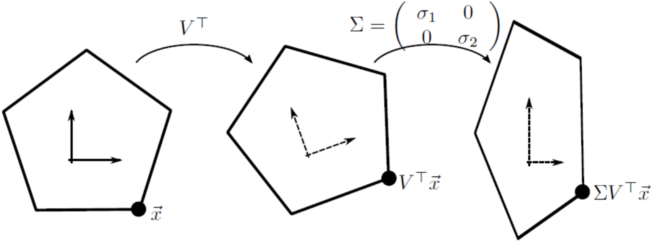
General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{U}}_{m\times m}\underbrace{\mathbf{\Sigma}}_{m\times n}\underbrace{\mathbf{V}^\top}_{n\times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices
- $\mathbf{\Sigma}$ is a rectangular diagonal matrix, e.g. $\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$



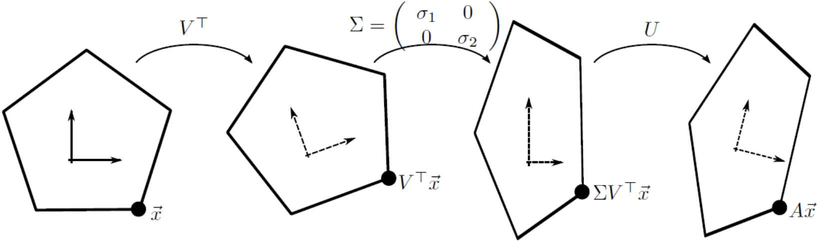
General transformations

Do we have a similar interpretation for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$?

Any matrix can be factorized as

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{U}}_{m\times m}\underbrace{\mathbf{\Sigma}}_{m\times n}\underbrace{\mathbf{V}^\top}_{n\times n}$$

- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices
- $\mathbf{\Sigma}$ is a rectangular diagonal matrix, e.g. $\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$



SVD

The factorization

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$

is called the singular value decomposition of matrix $\mathbf{A}$.

SVD

The factorization

$$\mathbf{A} = \underbrace{\mathbf{U}}_{\substack{\text{left}\\\text{singular}\\\text{vectors}}} \underbrace{\mathbf{\Sigma}}_{\substack{\text{singular}\\\text{values}}} \underbrace{\mathbf{V}^\top}_{\substack{\text{right}\\\text{singular}\\\text{vectors}}}$$

is called the singular value decomposition of matrix $\mathbf{A}$.

SVD

The factorization

$$\mathbf{A} = \underbrace{\mathbf{U}}_{\substack{\text{left}\\\text{singular}\\\text{vectors}}} \underbrace{\mathbf{\Sigma}}_{\substack{\text{singular}\\\text{values}}} \underbrace{\mathbf{V}^\top}_{\substack{\text{right}\\\text{singular}\\\text{vectors}}}$$

is called the singular value decomposition of matrix $\mathbf{A}$.

This can also be written as:

$$\mathbf{AV} = \mathbf{U\Sigma}$$

which looks quite similar to the eigenvalue equation.

Series of outer products

We can equivalently rewrite the decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ as

$$\mathbf{A} = \sum_{i=1}^{\ell} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th columns of $\mathbf{U}$ and $\mathbf{V}$.

Series of outer products

We can equivalently rewrite the decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ as

$$\mathbf{A} = \sum_{i=1}^{\ell} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th columns of $\mathbf{U}$ and $\mathbf{V}$.

Each outer product $\mathbf{u}_i \mathbf{v}_i^\top$ is a $m \times n$ matrix.

Series of outer products

We can equivalently rewrite the decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ as

$$\mathbf{A} = \sum_{i=1}^{\ell} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th columns of $\mathbf{U}$ and $\mathbf{V}$.

Each outer product $\mathbf{u}_i \mathbf{v}_i^\top$ is a $m \times n$ matrix.

Simply set $\ell = \min\{m, n\}$; the remaining columns are zeroed out by $\mathbf{\Sigma}$.

Series of outer products

We can equivalently rewrite the decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ as

$$\mathbf{A} = \sum_{i=1}^{\ell} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th columns of $\mathbf{U}$ and $\mathbf{V}$.

Each outer product $\mathbf{u}_i \mathbf{v}_i^\top$ is a $m \times n$ matrix.

Simply set $\ell = \min\{m, n\}$; the remaining columns are zeroed out by $\boldsymbol{\Sigma}$.

If we round small σ_i to zero, we approximate $\mathbf{A}$ with fewer terms:

$$\mathbf{A} \approx \mathbf{U}\tilde{\boldsymbol{\Sigma}}\mathbf{V}^\top$$

where $\tilde{\boldsymbol{\Sigma}}$ has the small σ_i truncated to zero.

Low-rank approximations

Construct the matrix:

$$\tilde{\mathbf{A}} \equiv \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$$

by truncating all but the first k largest singular values to zero.

Low-rank approximations

Construct the matrix:

$$\tilde{\mathbf{A}} \equiv \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$$

by truncating all but the first k largest singular values to zero.

Theorem (Eckart-Young) The matrix above minimizes the error $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F$ subject to the constraint that the column space of $\tilde{\mathbf{A}}$ has at most dimension k.

Low-rank approximations

Construct the matrix:

$$\tilde{\mathbf{A}} \equiv \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$$

by truncating all but the first k largest singular values to zero.

> **Theorem (Eckart-Young)** The matrix above minimizes the error $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F$ subject to the constraint that the column space of $\tilde{\mathbf{A}}$ has at most dimension k.

The rank of a matrix is the dimension of its column space.

Then, truncating the singular values gives a low-rank approximation (i.e. rank at most k) of the initial matrix $\mathbf{A}$.

Low-rank approximations

Construct the matrix:

$$\tilde{\mathbf{A}} \equiv \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$$

by truncating all but the first k largest singular values to zero.

Theorem (Eckart-Young) The matrix above minimizes the error $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F$ subject to the constraint that the column space of $\tilde{\mathbf{A}}$ has at most dimension k.

The rank of a matrix is the dimension of its column space.

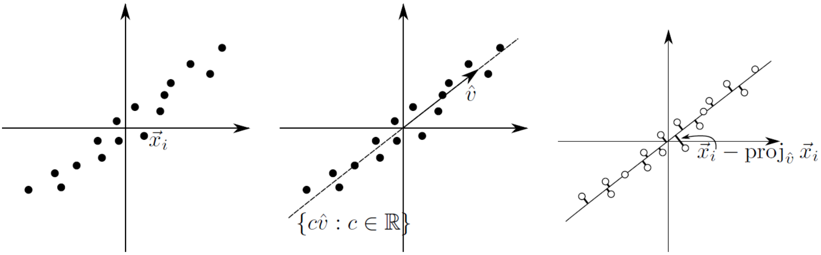
Then, truncating the singular values gives a low-rank approximation (i.e. rank at most k) of the initial matrix $\mathbf{A}$.

Low-rank approximations have numerous applications!

Principal Component Analysis (PCA)

Principal component

Consider the two-dimensional data in this plot:



(a) Input data (b) Principal axis (c) Projection error

Q: Find the vector $\mathbf{v}$ such that each data point $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i = c_i \mathbf{v}$$

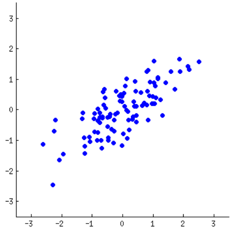
where each $\mathbf{x}_i$ has its own c_i

Another perspective

Let us be given n data points stored in matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$:

$$\mathbf{X}^\top = \begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}$$

.

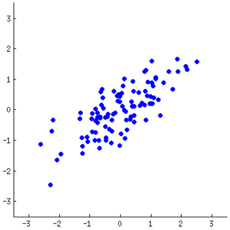


Another perspective

Let us be given n data points stored in matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$:

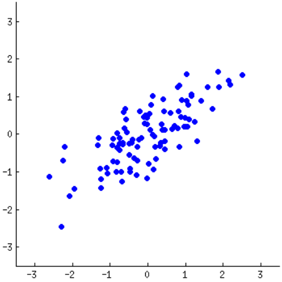
$$\mathbf{X}^\top = \begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix} \approx \begin{pmatrix} - & \tilde{\mathbf{x}}_1^\top & - \\ & \vdots & \\ - & \tilde{\mathbf{x}}_n^\top & - \end{pmatrix} = \tilde{\mathbf{X}}^\top$$

We want to replace them with a lower-dimensional approximation $\tilde{\mathbf{X}} \in \mathbb{R}^{k\times n}$, with $k \ll d$.



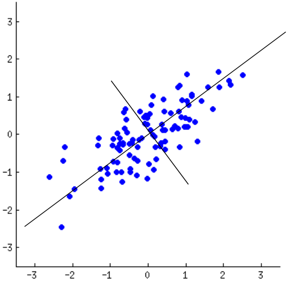
Principal component analysis (PCA)

Regard our data as n points in $\mathbb{R}^d$:



Principal component analysis (PCA)

Regard our data as n points in $\mathbb{R}^d$:

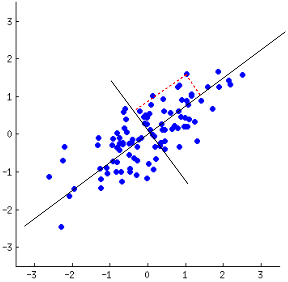


Overall idea:

- Find $k \leq d$ orthogonal directions with the most variance.
  These span a k-dimensional subspace of the data.

Principal component analysis (PCA)

Regard our data as n points in $\mathbb{R}^d$:



Overall idea:

- Find $k \leq d$ orthogonal directions with the most variance.
  These span a k-dimensional subspace of the data.
- Project all the data points onto these directions.
  This is lossy, but can be done with the smallest possible error.

Principal component analysis (PCA)

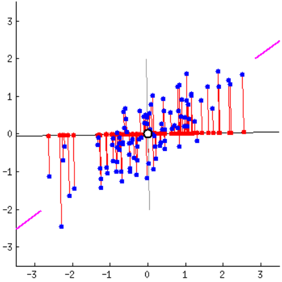
We seek the direction $\mathbf{w}$ that:

- Minimizes the projection/reconstruction error.
- Maximizes the variance of the projected data.

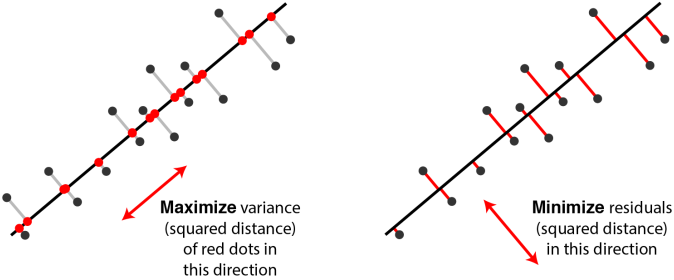
Principal component analysis (PCA)

We seek the direction $\mathbf{w}$ that:

- Minimizes the projection/reconstruction error.
- Maximizes the variance of the projected data.



Principal component analysis (PCA)



Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n\times d}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n \times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d \times k}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{x}_n^\top & \text{---} \end{pmatrix}}_{n \times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d \times k} = \underbrace{\begin{pmatrix} \text{---} & \mathbf{z}_1^\top & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{z}_n^\top & \text{---} \end{pmatrix}}_{n \times k}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n \times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d \times k} = \underbrace{\begin{pmatrix} — & \mathbf{z}_1^\top & — \\ & \vdots & \\ — & \mathbf{z}_n^\top & — \end{pmatrix}}_{n \times k}$$

Assuming $\mathbf{W}^\top \mathbf{W} = \mathbf{I}$, for $k = d$ we get:

$$\begin{aligned} \mathbf{X}^\top \mathbf{W} &= \mathbf{Z}^\top \\ \mathbf{X} &= \mathbf{W}\mathbf{Z} \end{aligned}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n\times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d\times k} = \underbrace{\begin{pmatrix} — & \mathbf{z}_1^\top & — \\ & \vdots & \\ — & \mathbf{z}_n^\top & — \end{pmatrix}}_{n\times k}$$

Assuming $\mathbf{W}^\top\mathbf{W} = \mathbf{I}$, for $k < d$ we get:

$$\mathbf{X}^\top\mathbf{W} = \mathbf{Z}^\top$$
$$\mathbf{X} \approx \mathbf{W}\mathbf{Z}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n \times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d \times k} = \underbrace{\begin{pmatrix} — & \mathbf{z}_1^\top & — \\ & \vdots & \\ — & \mathbf{z}_n^\top & — \end{pmatrix}}_{n \times k}$$

Assuming $\mathbf{W}^\top \mathbf{W} = \mathbf{I}$, for $k < d$ we get:

$$\begin{aligned} \mathbf{X}^\top \mathbf{W} &= \mathbf{Z}^\top \quad \text{projection} \\ \mathbf{X} &\approx \mathbf{W}\mathbf{Z} \quad \text{reconstruction} \end{aligned}$$

Principal component analysis (PCA)

In matrix notation:

$$\underbrace{\begin{pmatrix} — & \mathbf{x}_1^\top & — \\ & \vdots & \\ — & \mathbf{x}_n^\top & — \end{pmatrix}}_{n \times d} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_k \\ | & & | \end{pmatrix}}_{d \times k} = \underbrace{\begin{pmatrix} — & \mathbf{z}_1^\top & — \\ & \vdots & \\ — & \mathbf{z}_n^\top & — \end{pmatrix}}_{n \times k}$$

Assuming $\mathbf{W}^\top \mathbf{W} = \mathbf{I}$, for $k < d$ we get:

$$\begin{aligned} \mathbf{X}^\top \mathbf{W} &= \mathbf{Z}^\top && \text{projection} \\ \mathbf{X} &\approx \mathbf{W}\mathbf{Z} && \text{reconstruction} \end{aligned}$$

We call the columns of $\mathbf{W}$ principal components.

They are unknown and must be computed.

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top\mathbf{w}$.

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top \mathbf{w}$.

The variance to maximize is $\|\mathbf{X}^\top \mathbf{w}\|_2^2$:

$$(\mathbf{X}^\top \mathbf{w})^\top (\mathbf{X}^\top \mathbf{w})$$

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top \mathbf{w}$.

The variance to maximize is $\|\mathbf{X}^\top \mathbf{w}\|_2^2$:

$$(\mathbf{X}^\top \mathbf{w})^\top (\mathbf{X}^\top \mathbf{w}) = \mathbf{w}^\top (\mathbf{X}\mathbf{X}^\top)\mathbf{w}$$

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top \mathbf{w}$.

The variance to maximize is $\|\mathbf{X}^\top \mathbf{w}\|_2^2$:

$$(\mathbf{X}^\top \mathbf{w})^\top (\mathbf{X}^\top \mathbf{w}) = \mathbf{w}^\top \underbrace{(\mathbf{X}\mathbf{X}^\top)}_{\mathbf{C}} \mathbf{w}$$

where $\mathbf{C} \in \mathbb{R}^{d \times d}$ is the symmetric covariance matrix.

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top \mathbf{w}$.

The variance to maximize is $\|\mathbf{X}^\top \mathbf{w}\|_2^2$:

$$(\mathbf{X}^\top \mathbf{w})^\top (\mathbf{X}^\top \mathbf{w}) = \mathbf{w}^\top \underbrace{(\mathbf{X}\mathbf{X}^\top)}_{\mathbf{C}} \mathbf{w}$$

where $\mathbf{C} \in \mathbb{R}^{d \times d}$ is the symmetric covariance matrix.

We want to solve the problem:

$$\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{C} \mathbf{w} \quad \text{s.t. } \|\mathbf{w}\|_2 = 1$$

Principal component analysis (PCA)

Assume the data points $\mathbf{X}$ are centered at zero.

For a given $\mathbf{w}$, the projection of all n points onto $\mathbf{w}$ is $\mathbf{X}^\top \mathbf{w}$.

The variance to maximize is $\|\mathbf{X}^\top \mathbf{w}\|_2^2$:

$$(\mathbf{X}^\top \mathbf{w})^\top (\mathbf{X}^\top \mathbf{w}) = \mathbf{w}^\top \underbrace{(\mathbf{X}\mathbf{X}^\top)}_{\mathbf{C}} \mathbf{w}$$

where $\mathbf{C} \in \mathbb{R}^{d \times d}$ is the symmetric covariance matrix.

We want to solve the problem:

$$\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{C} \mathbf{w} \quad \text{s.t. } \|\mathbf{w}\|_2 = 1$$

The solution is $\mathbf{w} =$ principal eigenvector of $\mathbf{C}$ (Courant minmax principle), and the value $\mathbf{w}^\top \mathbf{C} \mathbf{w}$ is the corresponding eigenvalue.

Principal component analysis (PCA)

After solving the problem:

$$\mathbf{w}_1 = \arg\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{C}\mathbf{w} \quad \text{s.t. } \|\mathbf{w}\|_2 = 1$$

Principal component analysis (PCA)

After solving the problem:

$$\begin{aligned}\mathbf{w}_1 = \arg\max_{\mathbf{w}} \ & \mathbf{w}^\top \mathbf{C} \mathbf{w} \\ \text{s.t. } & \|\mathbf{w}\|_2 = 1\end{aligned}$$

The successive orthogonal direction can be found by solving:

$$\begin{aligned}\mathbf{w}_2 = \arg\max_{\mathbf{w}} \ & \mathbf{w}^\top \mathbf{C} \mathbf{w} \\ \text{s.t. } & \|\mathbf{w}\|_2 = 1 \\ & \mathbf{w}_1^\top \mathbf{w} = 0\end{aligned}$$

Principal component analysis (PCA)

After solving the problem:

$$\mathbf{w}_1 = \arg\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{C}\mathbf{w} \quad \text{s.t. } \|\mathbf{w}\|_2 = 1$$

The successive orthogonal direction can be found by solving:

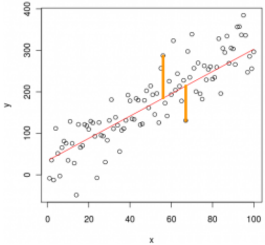
$$\begin{aligned} \mathbf{w}_2 = \arg\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{C}\mathbf{w} \\ \text{s.t. } \|\mathbf{w}\|_2 = 1 \\ \mathbf{w}_1^\top \mathbf{w} = 0 \end{aligned}$$

which is the second eigenvector of $\mathbf{C}$, and so on for all $\mathbf{w}_{i=2\ldots k}$.

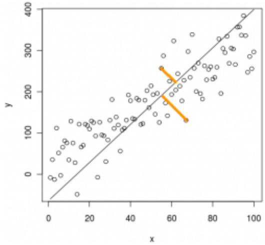
The principal components are thus the first $k \ll d$ eigenvectors of $\mathbf{C}$, sorted by decreasing eigenvalue.

PCA is not linear regression

With linear regression we measure the error along the y coordinate:



With PCA we measure the error orthogonal to the principal direction:



PCA as a generative model

Given the $\mathbf{W}$ satisfying, for the observations $\mathbf{X}$:

$$\begin{aligned} \mathbf{X}^\top \mathbf{W} &= \mathbf{Z}^\top \quad && \text{projection} \\ \mathbf{X} &\approx \mathbf{W}\mathbf{Z} && \text{reconstruction} \end{aligned}$$

We can generate new data just by sampling $\mathbf{z}_{\text{new}} \in \mathbb{R}^k$ and computing:

$$\mathbf{x}_{\text{new}} = \mathbf{W}\mathbf{z}_{\text{new}}$$

PCA as a generative model

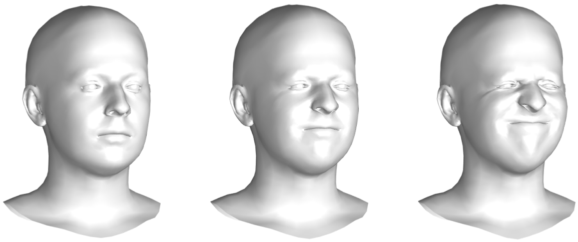
Given the $\mathbf{W}$ satisfying, for the observations $\mathbf{X}$:

$$\begin{aligned} \mathbf{X}^\top \mathbf{W} &= \mathbf{Z}^\top \quad \text{projection} \\ \mathbf{X} &\approx \mathbf{W}\mathbf{Z} \quad \text{reconstruction} \end{aligned}$$

We can generate new data just by sampling $\mathbf{z}_{\text{new}} \in \mathbb{R}^k$ and computing:

$$\mathbf{x}_{\text{new}} = \mathbf{W}\mathbf{z}_{\text{new}}$$

Example:



Data point $\mathbf{x}_1$ | **Generated** $\mathbf{x}_{\text{new}} = \frac{1}{2}\mathbf{W}(\mathbf{z}_1 + \mathbf{z}_2)$ | Data point $\mathbf{x}_2$

Suggested reading

Read Sections 6.1.1, 6.2, 6.2.1, 6.3.1, 6.3.2, 6.3.3, 6.4.2, 7.1, 7.2.2, 7.2.5 of the book:

J. Solomon, “Numerical Algorithms”