Data analysis, autumn 2021

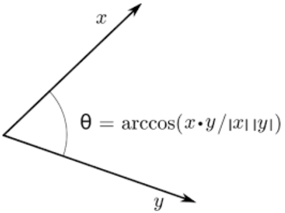
Principal Component Analysis

Benazha Hamed

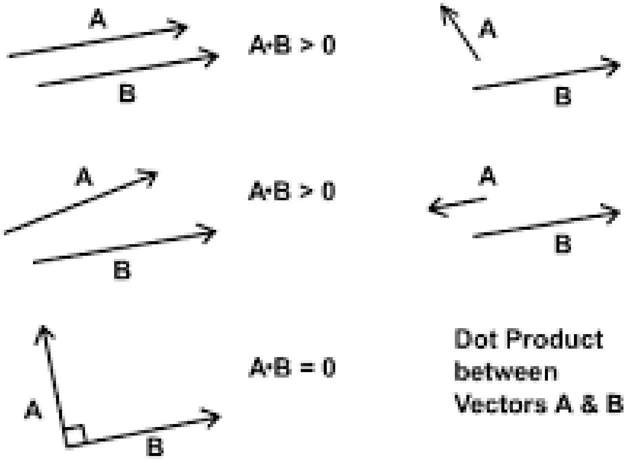
Vector dot product

Dot product between two vectors:

$$\mathrm{x} \cdot \mathrm{y} = \mathrm{x}^\top \mathrm{y} = \sum_{i=1}^{d} x_i y_i$$
$$= \|\mathrm{x}\| \, \|\mathrm{y}\| \cos(\theta) \in \mathbb{R}$$



Vector dot product



Eigenvalues and eigenvectors

If we can write

$$\mathsf{M}\mathsf{v} = \lambda \mathsf{v}$$

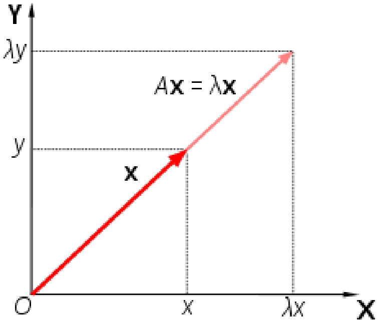
then v is called a eigenvector of matrix M, with eigenvalue λ.

Eigenvalues and eigenvectors

If we can write

$$\mathrm{Mv} = \lambda \mathrm{v}$$

then v is called a eigenvector of matrix M, with eigenvalue λ.



Eigenvalues and eigenvectors

If we can write

$$Mv = \lambda v$$

then v is called a eigenvector of matrix M, with eigenvalue λ.

Examples:

- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 1 \times (-1) \\ 1 \times 1 + 2 \times (-1) \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 1 \times 1 \\ 1 \times 1 + 2 \times 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \times 3 + 1 \times 1 \\ 1 \times 3 + 2 \times 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Eigenvalues and eigenvectors

If we can write

$$Mv = \lambda v$$

then v is called a eigenvector of matrix M, with eigenvalue λ.

Examples:

- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 1 \times (-1) \\ 1 \times 1 + 2 \times (-1) \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 1 \times 1 \\ 1 \times 1 + 2 \times 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \times 3 + 1 \times 1 \\ 1 \times 3 + 2 \times 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Notice that v multiplied by any scalar is still eigenvector in this definition! $\rightarrow$ select as eigenvector the one with length 1.

Eigenvalues and eigenvectors

If we can write

$$Mv = \lambda v$$

then v is called a eigenvector of matrix M, with eigenvalue λ.

- Matrix operation to this vector achieves only scaling with scalar λ.
- For positive symmetric matrices G it holds that $G = V\Lambda V^T$, with eigenvalues at diagonal of Λ, eigenvectors at columns of V
- Property of eigenvectors: orthogonality.

Variance and covariance

1. How different are the values of my variable? **Variance**

$$var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \approx \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{X})^2$$

2. Is there some relation between two variables? **Covariance**

$$\begin{aligned} cov(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &\approx \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{X})(y_i - \tilde{Y}) \end{aligned}$$

- > 0: if one increases/decreases, so does the other
- 0: no relation
- < 0: if one increases/decreases, the other decreases/increases

Principal Component Analysis (PCA)

L'analyse en composantes principales (ACP)

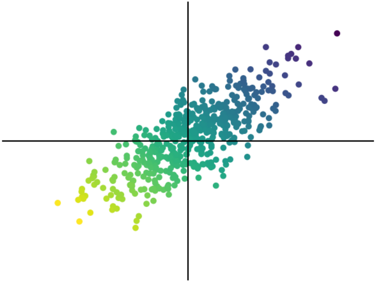
Answers the problem of **dimensionality reduction**, and allows one to **visualize high-dimensional data**.

- invented in 1901 by Karl Pearson
- sometimes called the discrete Karhunen–Loève transform (KLT) in signal processing & some other names in some other fields

PCA - general idea

Find the direction of **maximum variance** in data.

Find the direction of **maximum variance** in data.

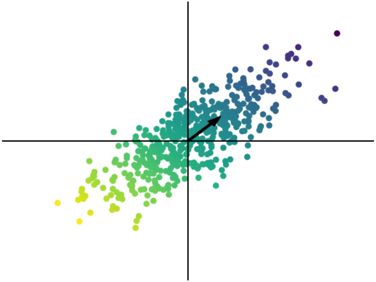


Data created with covariance matrix $\begin{bmatrix} 1 & 0.75 \\ 0.75 & 1 \end{bmatrix}$

(From this point on I assume my data is **centered**!)

PCA - general idea

Find the direction of **maximum variance** in data.

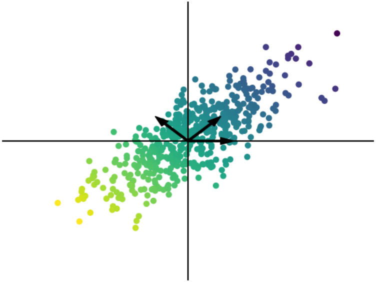


PCA - general idea

Find the direction of **maximum variance** in data.

PCA - maths version

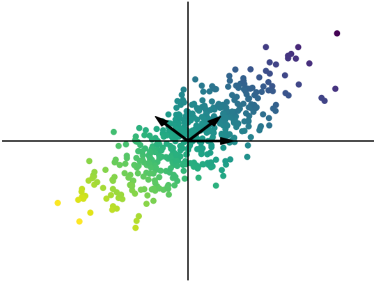
How do we find the direction that maximizes the variance?



PCA - maths version

How do we find the direction that maximizes the variance?

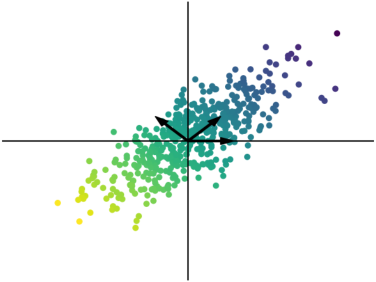
Dot product



PCA - maths version

How do we find the direction that maximizes the variance?

Dot product



Direction that maximizes the variance

$\Leftrightarrow$

Direction that maximizes the variance of dot products to data!

PCA - maths version

Consider the data x :

$$var(a^\top x) = \mathbb{E}_x[(a^\top x)^2]$$

Mean is zero!

PCA - maths version

Consider the data x :

$$
\begin{aligned}
var(a^\top x) &= \mathbb{E}_x[(a^\top x)^2] \\
&= \mathbb{E}_x[(a^\top x)(a^\top x)^T]
\end{aligned}
$$

PCA - maths version

Consider the data x :

$$\begin{aligned} var(a^\top x) &= \mathbb{E}_x[(a^\top x)^2] \\ &= \mathbb{E}_x[(a^\top x)(a^\top x)^T] \\ &= \mathbb{E}_x[a^\top x x^\top a] \end{aligned}$$

PCA - maths version

Consider the data x :

$$
\begin{aligned}
var(a^\top x) &= \mathbb{E}_x[(a^\top x)^2] \\
&= \mathbb{E}_x[(a^\top x)(a^\top x)^T] \\
&= \mathbb{E}_x[a^\top x x^\top a] \\
&= a^\top \mathbb{E}[x x^\top] a
\end{aligned}
$$

PCA - maths version

Consider the data x :

$$\begin{aligned} var(a^\top x) &= \mathbb{E}_x[(a^\top x)^2] \\ &= \mathbb{E}_x[(a^\top x)(a^\top x)^T] \\ &= \mathbb{E}_x[a^\top x x^\top a] \\ &= a^\top \mathbb{E}[x x^\top] a \\ &= a^\top C\, a \end{aligned}$$

PCA - maths version

Optimization problem with sample covariance matrix $\mathsf{C} \in \mathbb{R}^{d \times d}$:

$$\max_{\mathsf{a}} \; \mathsf{a}^\top \mathsf{C} \mathsf{a}$$

PCA - maths version

Optimization problem with sample covariance matrix $\mathsf{C} \in \mathbb{R}^{d \times d}$:

$$\max_{\mathsf{a}} \mathsf{a}^\top \mathsf{C} \mathsf{a} \quad s.t. \ \|\mathsf{a}\| = 1$$

PCA - maths version

Optimization problem with sample covariance matrix $\mathsf{C} \in \mathbb{R}^{d \times d}$:

$$\max_{\mathsf{a}} \ \mathsf{a}^\top \mathsf{C} \mathsf{a} \quad s.t. \ \|\mathsf{a}\| = 1$$

Eigenvalue decomposition:

$$\max_{\mathsf{a}} \ \mathsf{a}^\top \mathsf{V} \Lambda \mathsf{V}^\top \mathsf{a} \quad s.t. \ \|\mathsf{a}\| = 1$$

PCA - maths version

Optimization problem with sample covariance matrix $\mathrm{C} \in \mathbb{R}^{d \times d}$:

$$\max_{\mathrm{a}} \ \mathrm{a}^\top \mathrm{C} \mathrm{a} \quad s.t. \ \|\mathrm{a}\| = 1$$

Eigenvalue decomposition:

$$\max_{\mathrm{a}} \ \mathrm{a}^\top \mathrm{V} \Lambda \mathrm{V}^\top \mathrm{a} \quad s.t. \ \|\mathrm{a}\| = 1$$

Change of variables $\mathrm{V}^\top \mathrm{a} = \mathrm{b}$:

$$\max_{\mathrm{b}} \ \mathrm{b}^\top \Lambda \mathrm{b} \quad s.t. \ \|\mathrm{b}\| = 1$$

PCA - maths version

$$\max_{\mathrm{b}} \mathrm{b}^\top \Lambda \mathrm{b} \quad s.t.\ \|\mathrm{b}\| = 1$$

$$\Leftrightarrow$$

$$\max_{\mathrm{b}} \sum_{i=1}^{d} b_i^2 \lambda_i \quad s.t.\ \sum_{i=1}^{d} b_i^2 = 1$$

- each $b_i^2 \geq 0$
- sum of positive b_is should be one
- each $\lambda_i \geq 0$

$\Rightarrow$ choose b_i corresponding to largest eigenvalue to be 1, others 0!

PCA - maths version

Now b is vector with one 1 and rest zeros.
Remember $V^\top a = b$

- V consists of orthogonal columns
- a has to be the eigenvector corresponding to the largest eigenvalue!

PCA - maths version

How to get the other components?

Same process but constraint on orthogonality: new component has to be orthogonal to all previous ones:

$$\max_{a_i} a_i^\top C a_i$$
$$s.t.\ \ \|a_i\| = 1,\ a_i^\top a_j = 0\ \forall\ j = 1, ..., i-1$$

...

Gives as a solution the *i*th largest eigenvector.

PCA - algorithm

Assumes centered data!

1. calculate sample covariance matrix Σ from data
2. calculate eigenvalues and eigenvectors for Σ
   - eigenvectors are the *principal components*, directions of maximum variance in data
   - eigenvalues give *ratio of variance explained by the data*: So if eigenvalues were 3, 0.5 and 0.2, the first component would explain 81% of the variance in data.
3. To project the data into m most influential directions:
   - Take m most significant components into a matrix:

$$\mathrm{W} = \begin{bmatrix} \mathrm{pc}_1 & \mathrm{pc}_2 & \cdots & \mathrm{pc}_m \end{bmatrix}$$

   - Calculate transformed data:

$$\underbrace{\hat{\mathrm{X}}}_{n \times m} = \underbrace{\mathrm{X}}_{n \times d} \underbrace{\mathrm{W}}_{d \times m}$$

PCA - other stuff

Not covered here but introduced in exercises:

- projecting data back to original space
- quality of transformation
  - correlation circle
  - squared cosine measure

Not covered at all:

- Extensions of PCA, such as non-linearity via kernel-PCA, or scale-invariant extension.

Exercises and homework

- Notebook to complete at `celestium.eu/teaching/dana`
- Open with:
  - google colab: go to `colab.research.google.com` and upload there the notebook you downloaded from my website
  - or with jupyter notebook from your own computer if you have python 3
- Return the notebook **at latest** at the beginning of the next exercise session
  - send to `hamed.benazha@centrale-marseille.fr`
  - Send the completed notebook with everything run
  - No need to send any of the accompanying material; I have them already
- Your work will be graded