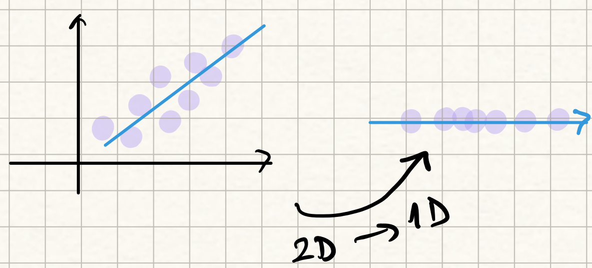
Determinants and Eigenvectors

13.03.2025

What is our motivation for learning determinants and eigenvectors?

Principal Component Analysis (PCA)



Goal

- ↓ dimensions of d. set
- Preserv. as much info as. psble.

NON-SINGULAR TRANSFORMATION

Non-singular matrix → Linear Transform. → Covers all the plane

SINGULAR TRANSFORMATION

Singular Matrix → Linear Transform. → A line or a dot! Doesn't cover all the plane

Rank of the lin. transf. MX = # dimensions of the result (the image)

Determinant of the lin. transf. MX = Area of the image

Non-singular lin transf. mx → Det $\neq 0$

Singular lin. transf. mx → Det $= 0$

Determinant of a product

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ -3 & 3 \end{bmatrix}$$

det = 5   det = 3   det = 15

$5 \times 3 = 15$

$\det(A \cdot B) = \det(A) \cdot \det B$

Determinant of an inverse matrix

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}$$

det = 5   det = 1/5

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

- We are going to use determinants and eigenvectors in order to perform a Principal Component Analysis (PCA), which will help us to reduce the dimensions of a dataset as much as possible without losing significant amount of data.

What are bases?

① Define the fundamental directions of a vector space

② Allow any point in that space to be reached through linear combinations of those vectoss.

③ Must be linearly independent

④ Must span the entire vector space

⑤ Must be the minimum number of vectors that span the entire space

What's the relationship between the bases and the dimension of the vector space?

* If we have more vectors than the dimension of the space we're trying to span, we will always have a linearly dependent group.

Ex: If we have a 2D plane, we can span all of it with 2 linearly independent vectors. If we add a 3rd vector, it will always be able to reached by the other two.

Let $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $v_3 = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$

Two dimensions
Three vectors

$$\alpha \cdot v_1 + \beta v_2 = v_3$$

$$\alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

$$\begin{aligned} -\alpha + 2\beta &= -5 \\ \alpha + \beta &= 3 \end{aligned} \Rightarrow \begin{aligned} \beta &= -2/3 \\ \alpha &= 11/3 \end{aligned} \Rightarrow v_3 = \frac{11 \cdot v_1 - 2 v_2}{3}$$

Basis is a set of linearly independent vectors that can "build" any other vector in a vector space through linear combinations. It's like the core set of directions we need to navigate the space.

Span is the set of all possible vectors that can be created by taking linear combinations of a given set of vectors. It's the "area" or "volume" covered by those vectors.

What is an eigenbasis?

A special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \Rightarrow$$

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A \cdot v_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \Rightarrow A \cdot v_1 = \lambda_1 \cdot v_1$$

$$\lambda_1 = 2$$

Eigenvalues

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A \cdot v_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \Rightarrow A \cdot v_2 = \lambda_2 \cdot v_2$$

Eigenvectors

$$\lambda_2 = 3$$

$$v_3 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad A \cdot v_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \Rightarrow A \cdot v_3 \neq \lambda_3 \cdot v_3$$

If $A \cdot v_1 = \lambda_1 \cdot v_1$ $\Rightarrow$ LESS WORK!

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{8 multiplications}} = \underbrace{2 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{2 multiplications}}$$

A special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

A eigenbasis is a special basis where the linear transformation acts by simply stretching the basis vectors. This means the output parallelogram's sides are parallel to the input parallelogram's sides.

Even if $A \cdot v_3 \neq \lambda_3 \cdot v_3$ might be less work.

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \left[-3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] \Rightarrow$$

$$A \cdot v_3 = A \cdot \left[-3 \cdot v_1 + 2 \cdot v_2 \right]$$

$$A \cdot v_3 = -3 \cdot \underbrace{A \cdot v_1}_{\lambda_1 \cdot v_1} + 2 \cdot \underbrace{A \cdot v_2}_{\lambda_2 \cdot v_2} \Rightarrow$$

$$A \cdot v_3 = -3 \cdot 2 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \cdot 3 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow$$

$$= -6 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

- $A \cdot v = \lambda \cdot v \Rightarrow \lambda$ = Eigenvalue , v = Eigenvector
- Eigenvectors : Direction of the stretch.
- Eigenvalues : How much stretch?
- Eigenbasis : The set of a matrix's eigenvectors. Can be arranged as a matrix with one eigenvector in each column.
- They save work and characterize a transformation.

If λ is an eigenvalue, then

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$ for infinitely many (x,y)

$$\Rightarrow \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$ has infinitely many solutions

$$\Rightarrow \det \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} = 0 \Rightarrow (2-\lambda)\cdot(3-\lambda) - 1\cdot 0 = 0$$

$\lambda_1 = 2$

$\lambda_2 = 3$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 2 \cdot \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{matrix} 2x + y = 2x \\ 0x + 3y = 2y \end{matrix} \Rightarrow \begin{matrix} x = 1 \\ y = 0 \end{matrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{matrix} 2x + y = 3x \\ 0x + 3y = 3y \end{matrix} \Rightarrow \begin{matrix} x = 1 \\ y = 1 \end{matrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

* If matrix is not square ⇒ No eigenvalues / eigenvectors !

---

* $\det(A - \lambda I) = 0$ is called the characteristic equation. (det = determinant, A: transformat. matrix, λ: Eigenvalues, I: Identity matrix. We use it to find eigenvalues.
* Then we find the eigenvectors with this formula: $A \cdot v_i = \lambda_i \cdot v_i$
* Only square matrices can have eigenbases.

14.03.2025

If we have repeated eigenvalues, can we make assumptions about the # of eugen vectors?

No!

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & -0.5 \\ 0 & 0 & 2 \end{bmatrix} \qquad \det(A - \lambda I) = 0 \Rightarrow$$

$$(2-\lambda)(4-\lambda)(2-\lambda) + 0.(-0.5).0 + 0.(-1).0 - 0.4.0 + 0.(-1).2 + 2.(0.5).0 = 0$$

$\lambda_1 = 2$ $\quad \lambda_2 = 4$ $\quad \lambda_3 = 2$

$$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

2 EIGEN VALUES
3 EIGEN VECTORS

point diff. directions

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & -0.5 \\ 4 & 0 & 2 \end{bmatrix}$$

$\lambda_1 = 2$ $\quad \lambda_2 = 4$ $\quad \lambda_3 = 2$

$$\begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0.5 \\ 2 \end{bmatrix}$$

$[0 \;\; k \;\; 4k] \Rightarrow$ same direction

2 EIGEN VALUES
2 EIGEN VECTORS

$\Downarrow$

3 Dimensions 2 Eigen vectors
NO EIGENBASIS
(cannot cover the whole space)

2x2

$\lambda_1 \neq \lambda_2 \Rightarrow$ 2 eigenvectors

$\lambda_1 = \lambda_2$ → 1 eigenvector / → 2 eigenvectors

3x3

$\lambda_1 \neq \lambda_2 \neq \lambda_3 \Rightarrow$ 3 eigenvectors

$\lambda_1 = \lambda_2 \neq \lambda_3$ → 2 eigenvectors / → 3 eigenvectors

$\lambda_1 = \lambda_2 = \lambda_3$ → 1 eigenvector / → 2 eigenvectors / → 3 eigenvectors

Why do we perform dimensionality reduction?

Dimensionality Reduction → Same # of rows , less # of columns.

Why?

- Smaller datasets
- Easier to visualize

How?

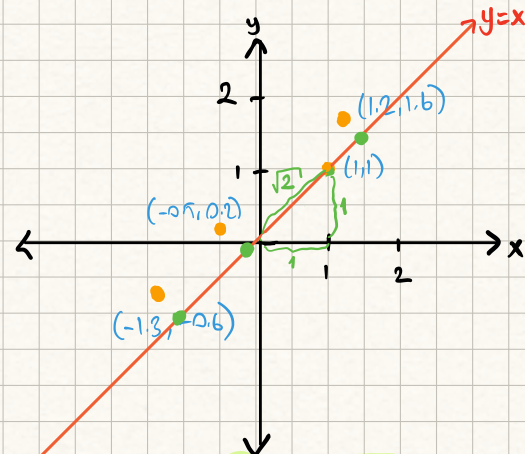
- Delete columns ? – Easy but probably loss of useful info..
- PCA – Harder but avoids losing too much useful info.

Why do we choose PCA to reduce dimensions?

What is projection?

Projection: Moving our data points into a vector space with fewer dimensions.

	x	y
V_1	1.0	1.0
V_2	1.2	1.6
V_3	-0.5	0.2
V_4	-1.3	-0.6



projecting: ● → ●

$(1,1) \rightarrow \sqrt{2}$

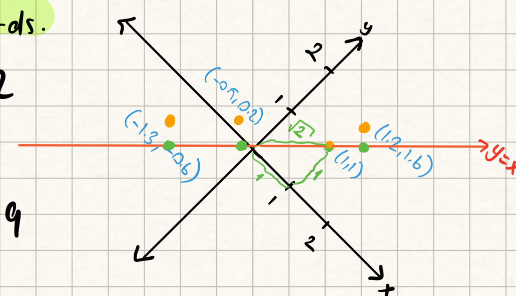
Final Coords.

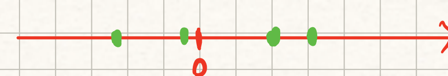
$V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \dfrac{1+1}{\sqrt{2}} = 1.4142$

$V_2 \begin{bmatrix} 1.2 \\ 1.6 \end{bmatrix} \Rightarrow \dfrac{1.2+1.6}{\sqrt{2}} = 1.9799$

$V_3 \begin{bmatrix} -0.5 \\ 0.2 \end{bmatrix} \Rightarrow \dfrac{-0.5+0.2}{\sqrt{2}} = -0.2121$

$V_4 \begin{bmatrix} -1.3 \\ -0.6 \end{bmatrix} \Rightarrow \dfrac{-1.3-0.6}{\sqrt{2}} = -1.344$





We reduce dimensions of our dataset by projecting our data points onto vectors from a vector space with fewer dimensions than our dataset.

What's the formula for projection?

To project a matrix A onto a vector v:

$$A_p = A \cdot \frac{v}{\|v\|_2}$$

$r \times 1 \quad r \times c \quad c \times 1$

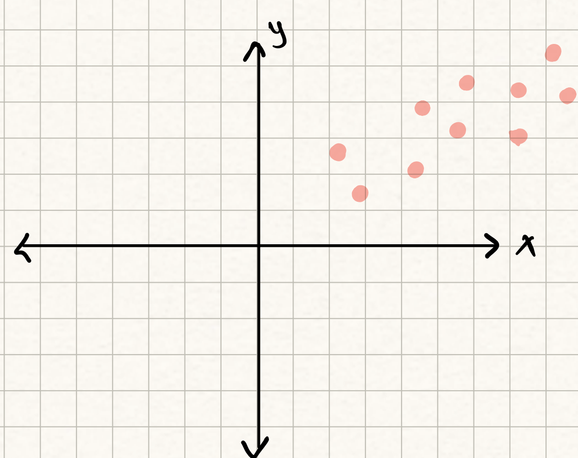
To project a matrix A onto vectors v_1 and v_2

$$A_p = A \cdot \underbrace{\begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix}}_{\text{matrix } V}$$

$r \times 2 \quad r \times c \quad c \times 2$

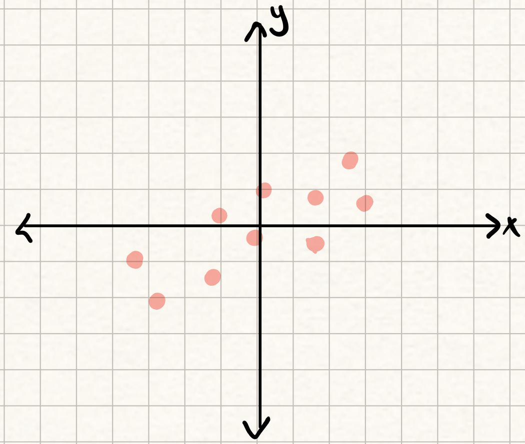
$$A_p = A \cdot V$$

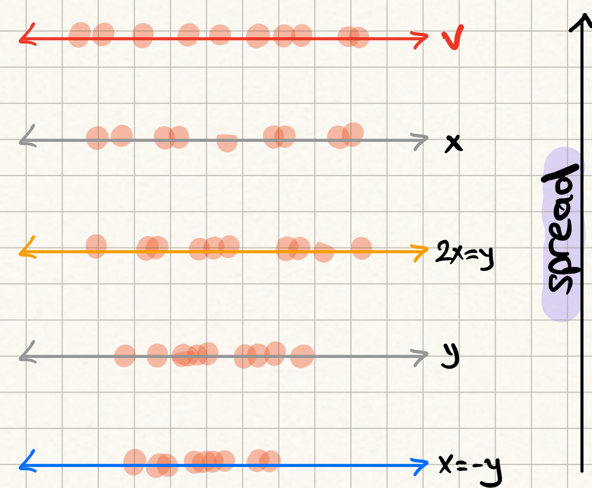
How do we pick the vectors to project onto?



This is the dataset we want to apply PCA.

Let's start with centering our dataset around the origin:





more spread → preserving more info.

so how do we find v ?

spread = variance = "The average squared distance from the mean"

$$var(x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)^2$$

* μ = mean

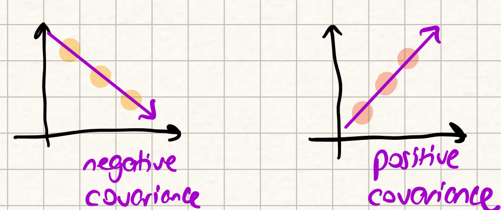
* $x_i - \mu_x$ gives coords for the centered data

Problem!

Both datasets have the same variance!

Solution :

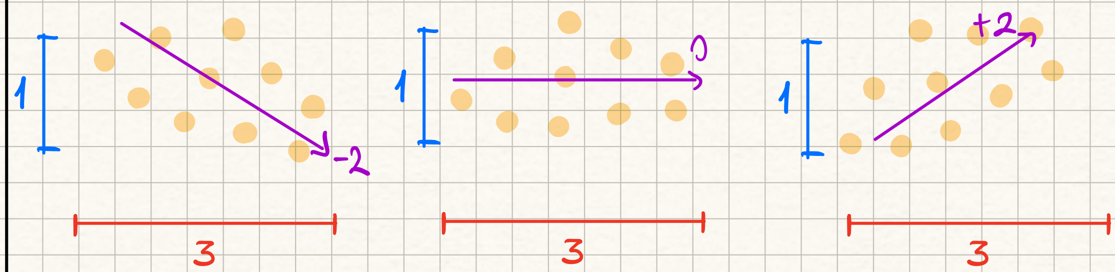
Covariance!



Covariance tells us the direction of the relationship between two variables.

$$Cov(x,y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x) \cdot (y_i - \mu_y)$$

· Our goal is to chose a projection matrix that will give us the maximum spread when we project data onto it, because maximum spread means minimum loss of information.



All datasets have the same variances on the x-axis (3) and y-axis (1).
However, their covariances differ.

$$\begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} var(x) & cov(x,y)^* \\ cov(y,x)^* & var(y) \end{bmatrix} \Rightarrow$$

* $cov(x,y) = cov(y,x)$

$$C = \begin{bmatrix} cov(x,x)^* & cov(x,y) \\ cov(y,x) & cov(y,y)^* \end{bmatrix}$$

* $var(a) = cov(a,a)$

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_y \end{bmatrix} \Rightarrow$$

$$C = \frac{1}{n-1} (A-\mu)^T (A-\mu) \Rightarrow$$

A covariance matrix is a square matrix that summarizes the variances and covariances of features in a dataset. The eugen vectors for the n number of largest eugen values for dataset is used to find the projection matrix that will give us the largest spread (and minimum data loss).

$$= \frac{1}{n-1}\left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} - \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_y \end{bmatrix}\right)^T \cdot \left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} - \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_x \end{bmatrix}\right)$$

$$= \frac{1}{n-1}\begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}^T \cdot \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}$$

$$= \frac{1}{n-1}\begin{bmatrix} x_1-\mu_x & x_2-\mu_x & \dots & x_n-\mu_x \\ y_1-\mu_y & y_2-\mu_y & \dots & y_n-\mu_y \end{bmatrix} \cdot \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}$$

$$(x_1-\mu_x)(x_1-\mu_x)+(x_2-\mu_x)(x_2-\mu_x)+\dots+(x_n-\mu_x)\cdot(x_n-\mu_x)$$

$$\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\mu_x)^2 = var(x)$$

$$(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\dots+(x_n-\mu_x)(y_n-\mu_y)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) = cov(x,y)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)(x_i-\mu_x) = cov(y,x)$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2 = var(y)$$

$$A-\mu = \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix} \qquad C = \frac{1}{n-1}(A-\mu)^T(A-\mu)$$

$$\underset{(2\times 8)}{A} = \begin{bmatrix} 10 & 5 \\ 12 & 3 \\ 6 & 9 \\ 6 & 4 \\ 5 & 11 \\ 14 & 2 \\ 8 & 1 \\ 3 & 13 \end{bmatrix} \Rightarrow A-\mu = \begin{bmatrix} 2 & -1 \\ 4 & -3 \\ -2 & 3 \\ -2 & -2 \\ -3 & 5 \\ 6 & -4 \\ 0 & -5 \\ -5 & 8 \end{bmatrix} \Rightarrow$$

$\mu_x = 8 \quad \mu_y = 6$

$$C = \frac{1}{8-1} \cdot \begin{bmatrix} 2 & 4 & -2 & -2 & -3 & 6 & 0 & 5 \\ -1 & -3 & 3 & -2 & 5 & -4 & -5 & 8 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 \\ 4 & -3 \\ -2 & 3 \\ -2 & -2 \\ -3 & 5 \\ 6 & -4 \\ 0 & -5 \\ -5 & 8 \end{bmatrix}$$

$$C = \begin{bmatrix} 14 & -11.86 \\ -11.86 & 19.71 \end{bmatrix}$$

1. Arrange data with a different feature in each column
2. Calculate column averages
3. Subtract each average from their respective column $(A-\mu)$
4. Apply: $\frac{1}{n-1} \cdot (A-\mu)^T \cdot (A-\mu)$

What are the 6 steps of PCA?

Covariance Matrix → PCA
Eigenvalues/Eigenvectors → PCA
Projections → PCA

We have n observations of 5 variables and our goal is to reduce our dataset to 2 variables.

① Create Matrix

$$X = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{5,1} \\ x_{2,1} & x_{2,2} & \cdots & x_{5,2} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,5} \end{bmatrix}$$

② Center the Data

$$X-\mu = \begin{bmatrix} x_{1,1}-\mu_1 & x_{1,2}-\mu_2 & \cdots & x_{1,5}-\mu_5 \\ x_{2,1}-\mu_1 & x_{2,2}-\mu_2 & \cdots & x_{2,5}-\mu_5 \\ \vdots & \vdots & & \\ x_{n,1}-\mu_1 & x_{n,2}-\mu_2 & \cdots & x_{n,5}-\mu_5 \end{bmatrix}$$

③ Calculate the Covarience Matrix

$$C = \frac{1}{n-1} \cdot (X-\mu)^T \cdot (X-\mu)$$

④ Calculate Eigenvectors / Eigenvalues

$$\det(A - \lambda I) = 0$$

largest ⟷ smallest

λ_1 λ_2 λ_3 λ_4 λ_5

v_1 v_2 v_3 v_4 v_5

⑤ Create Projection Matrix

$$V = \begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix}$$

⑥ Project the centered data onto the projection matrix

$$X_{PCA} = (X-\mu) \cdot V$$

No summary: Learn the 6-step process of PCA through the instructions on this page!

17.03.2025

What is a discrete dynamic System?

Discrete → B/c not continuous. We're looking at changes at specific time points (e.g. each day, each click)

Dynamic → B/c it changes

What is a Markov Matrix? (M)

1. All entries are non-negative (e.g. probabilities)

AND

2. The columns add up to 1 (e.g. probabilities)

What is the equilibrium state?

State vector : Current state of the system : X_0

Next state : $X_1 = M \cdot X_0$

$$\Rightarrow \quad X_{n+1} = M \cdot X_n$$

- Eventually $X_{n+1} \cong X_n$

$$\Rightarrow \quad \lim_{x \to \infty} X_n = X_{eq}$$

- So $M \cdot X_{eq} = X_{eq}$

$\Rightarrow$ X_{eq} = Eigenvector of the matrix M ! (Equilibrium vector)

What is a transition matrix?

Markov matrix in a discrete dynamic system = Transition Matrix (P)