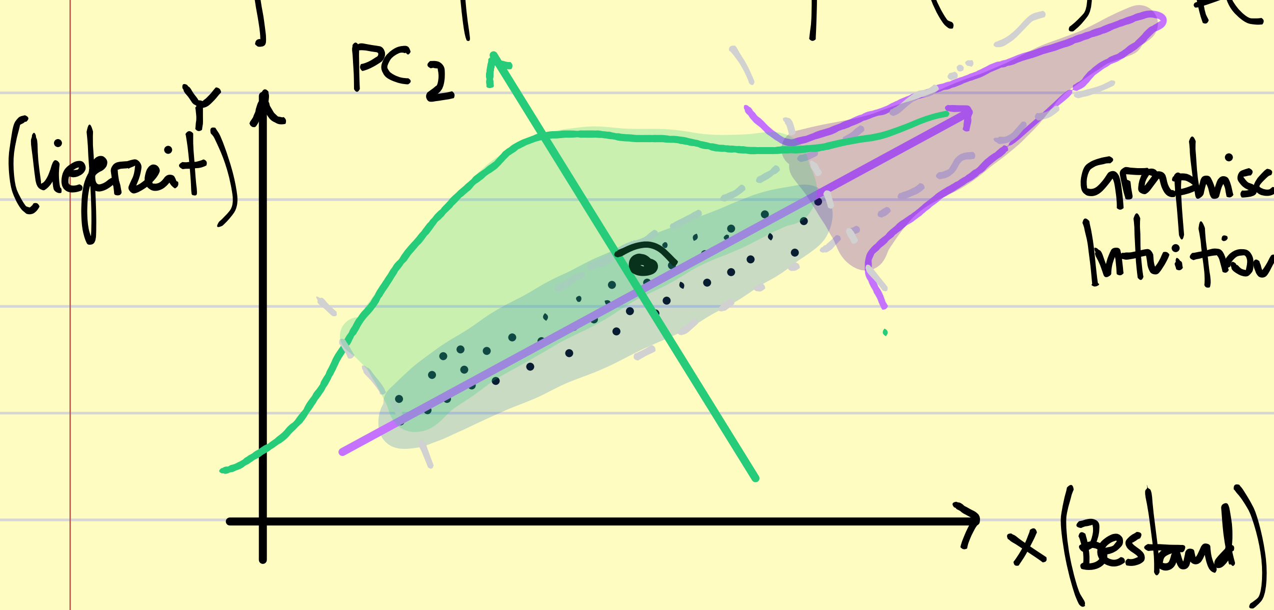
Wie viel Variabilität können wir aus mehrdimensionalen Daten herauslesen?

DIMENSIONALITÄTSREDUKTIONSANSÄTZE . HAUPTKOMPONENTENANALYSE

Principal Component Analysis (PCA)



$PC_1 = y = ax + b$

Graphische Intuition

LZ B

· Der Vektor → beschreibt die Daten am besten.

· Warum? Weil die Variabilität (Std Abweichung) der Daten in dieser Richtung minimal ist: die Gauß-Glocke ist sehr schmal in die Richtung.

· Diese Richtung in dem die Variabilität minimal ist, erklärt die Daten am besten und heißt HAUPTKOMPONENTE Nr 1. (PC_1).

· SENKRECHT zu PC_1 ist PC_2. PC_2 erklärt wenigere Variabilität (höhere Std Abweichung) von den Daten. →

· Die volle Variabilität der Daten X, Y wird mit PC_1 & PC_2 erklärt. In unserem Beispiel konnten wir 80% der Variab. der Daten mit NUR einem KPI (PC_1) erklären.

$\boxed{\text{Die EIGENVEKTOREN der KOVARIANZMATRIX (Normiert) sind die Hauptkomponenten eines Datasets.}}$

Eigenvektoren einer Matrix (A) ermitteln:

$A\vec{v} = \lambda\vec{v}$ ($\vec{v}$: Eigenvektoren ; λ: Eigenwerte)

$\downarrow$

$\det(A - \lambda I) = 0 \rightarrow \lambda \rightarrow \vec{v}$

Beispiel. Schritt 1. Kennzahlen System

	DLZ	Bestand
KW1	...	...
KW2	...	...
KW3	...	...
KW4	...	...

Schritt 2. Normierung. ✓

Schritt 3. Kovarianzmatrix ✓ A

Schritt 4. $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ Eigenwerte & Eigenvektoren? PCA

$$A\vec{v} = \lambda\vec{v} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

$$\det[A - \lambda I] = 0 \rightarrow \det\left[\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right] = 0 \rightarrow$$

$$\rightarrow \det\left[\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right] = 0 \rightarrow \det\left[\begin{bmatrix} 0-\lambda & 1-0 \\ -2-0 & -3-\lambda \end{bmatrix}\right] = 0$$

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

$$\boxed{\det\begin{bmatrix} a & c \\ b & d \end{bmatrix} = a.d - cb}$$

$$\rightarrow \lambda^2 + 3\lambda + 2 = 0$$

$$\lambda = \frac{-3 \pm \sqrt{9-4.2}}{2} = \frac{-3\pm 1}{2} = \begin{cases} \lambda_1 = -1 \\ \lambda_2 = -2 \end{cases}$$

Eigenwerte

$$\boxed{ax^2+bx+c=0 \rightarrow x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}}$$

$$\boxed{\lambda_1 = -1}$$

$$A\vec{v_1} = \lambda_1 \vec{v_1} \rightarrow \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = (-1)\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \rightarrow$$

$$\rightarrow \begin{cases} 0.v_{11} + 1.v_{12} = (-1)v_{11} \\ -2.v_{11} - 3.v_{12} = (-1)v_{12} \end{cases} \quad \begin{matrix} v_{12} = -v_{11} \\ v_{12} = -v_{11} \end{matrix}$$

$$v_{11} = 1 \rightarrow v_{12} = -1 \rightarrow \boxed{\vec{v_1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}} \rightarrow \vec{v_1}^* = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

Normiert

$$\boxed{\lambda_2 = -2}$$

$$A\vec{v_2} = \lambda_2 \vec{v_2} \rightarrow \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = (-2)\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} \rightarrow$$

$$\rightarrow \begin{cases} 0.v_{21} + v_{22} = -2v_{21} \\ -2.v_{21} - 3v_{22} = -2v_{22} \end{cases} \quad \begin{matrix} v_{22} = -2v_{21} \\ v_{22} = -2v_{21} \end{matrix}$$

$$v_{21} = 1 \rightarrow v_{22} = -2 \rightarrow \vec{v_2} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \rightarrow \vec{v_2}^* = \begin{bmatrix} 1/\sqrt{3} \\ -2/\sqrt{3} \end{bmatrix}$$

Übung. Gegeben wird ein KPI-System mit 2 Kennzahlen. Durchlaufzeit (X) und Output (Y).

(9) a) bitte ermitteln Sie die Kovarianzmatrix nach Normierung

(9) b) bitte ermitteln Sie die Eigenwerte und Eigenvektoren der Kov. Matrix und

(2) c) bitte interpretieren Sie die Ergebnisse.

DLZ(X) : [17, 14, 12, 13, 9, 7]

OUTPUT(Y): [200, 250, 270, 240, 310, 330]