Prof. Marc Alexa, Ugo Finnendahl,
Dimitrios Bogiokas, Gabriel Dogadov

Wissenschaftliches Rechnen - Übung 3.2

Singulärwertzerlegung

04.12.2023 bis 08.12.2023

Aufgabe 1: Singulärwertzerlegung

Die Singulärwertzerlegung (SVD) ist, in der vollen Form, die Faktorisierung einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ in

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_k \end{bmatrix} \begin{bmatrix} — & \mathbf{v}_1^\mathsf{T} & — \\ & \vdots & \\ — & \mathbf{v}_m^\mathsf{T} & — \end{bmatrix},$$

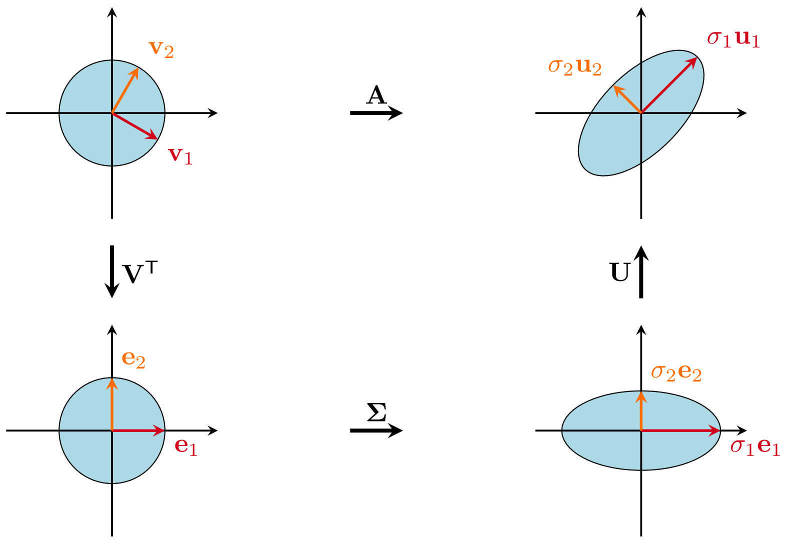
wobei $\mathbf{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times m}$, $\mathbf{V} \in \mathbb{R}^{m\times m}$ und $k = \min\{n, m\}$.

1. Welche Eigenschaften haben die Matrizen $\mathbf{U}$, $\boldsymbol{\Sigma}$ und $\mathbf{V}^\mathsf{T}$ in der vollen Form der SVD? Füllen Sie dazu die folgende Tabelle aus. Welche Eigenschaften und Konventionen sind bei den Singulärwerten $\sigma_1, \dots, \sigma_k$ zu beachten? Unter welchen Bedingungen besitzt eine Matrix eine SVD?

Matrix	Inhalt	Eigenschaft	Geometrische Interpretation
$\mathbf{U}$	linke Singulärvektoren (in den Spalten)	orthogonal	Drehung und/oder Spiegelung
$\boldsymbol{\Sigma}$	Singulärwerte (auf der Diagonalen)	(rechteckig) diagonal	Skalierung
$\mathbf{V}^\mathsf{T}$	rechte Singulärvektoren (in den Zeilen)	orthogonal	Drehung und/oder Spiegelung

Lösung

Die Singulärwerte sind nicht-negativ und der Größe nach geordnet: $\sigma_1 \geq \dots \geq \sigma_k \geq 0$. Jede Matrix besitzt eine Singulärwertzerlegung.



Lösung Ende

2. Für die folgende Matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ werden vier Zerlegungen $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}$ vorgestellt. Entscheiden Sie, ob es sich bei diesen um eine korrekte, den Konventionen konforme SVD handelt oder nicht.

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

a) $\mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{V}^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Keine SVD, da $\mathbf{U}$ nicht orthogonal ist.

b) $\mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$, $\mathbf{V}^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Keine SVD, da $\mathbf{\Sigma}$ einen negativen Eintrag enthält.

c) $\mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, $\mathbf{V}^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Keine SVD, da die Singulärwerte nicht korrekt geordnet sind.

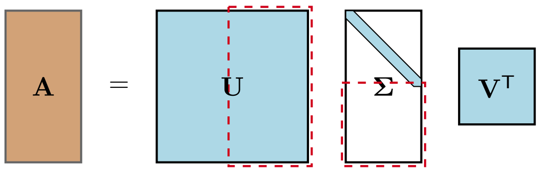
d) $\mathbf{U} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{V}^\top = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Eine korrekte SVD.

3. Sei $\mathbf{A} \in \mathbb{R}^{n \times m}$ mit $n \neq m$. Was ist der Unterschied zwischen der vollen und der reduzierten Form der SVD für die Fälle $n > m$ und $m > n$?

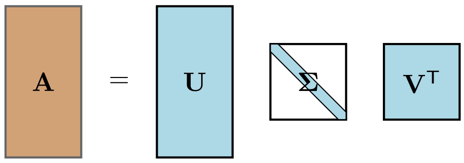
Lösung

Falls $n > m$, dann enthält die Matrix $\mathbf{U}$ zusätzliche $n - m$ Singulärvektoren, welche mit dem Nullblock der Matrix $\mathbf{\Sigma}$ multipliziert werden. Diese sind Eigenvektoren von $\mathbf{A}\mathbf{A}^\top$ zum Eigenwert Null.



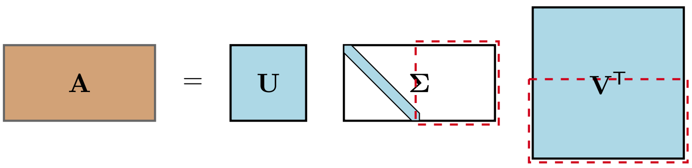
$\mathbf{A} \in \mathbb{R}^{n \times m}$
$\mathbf{U} \in \mathbb{R}^{n \times n}$
$\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$
$\mathbf{V}^\top \in \mathbb{R}^{m \times m}$

Die reduzierte Form sieht dann wie folgt aus:



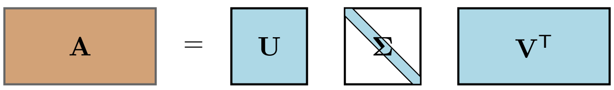
$\mathbf{A} \in \mathbb{R}^{n \times m}$
$\mathbf{U} \in \mathbb{R}^{n \times m}$
$\mathbf{\Sigma} \in \mathbb{R}^{m \times m}$
$\mathbf{V}^\top \in \mathbb{R}^{m \times m}$

Falls $m > n$, dann enthält die Matrix $\mathbf{V}$ zusätzliche $m - n$ Singulärvektoren, welche mit dem Nullblock der Matrix $\mathbf{\Sigma}$ multipliziert werden. Diese sind Eigenvektoren von $\mathbf{A}^\top\mathbf{A}$ zum Eigenwert Null.



$\mathbf{A} \in \mathbb{R}^{n \times m}$
$\mathbf{U} \in \mathbb{R}^{n \times n}$
$\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$
$\mathbf{V}^\top \in \mathbb{R}^{m \times m}$

Die reduzierte Form sieht dann wie folgt aus:



$\mathbf{A} \in \mathbb{R}^{n \times m}$
$\mathbf{U} \in \mathbb{R}^{n \times n}$
$\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$
$\mathbf{V}^\top \in \mathbb{R}^{n \times m}$

Lösung Ende

4. Stellen Sie einen Zusammenhang zwischen der SVD einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ und der Eigenzerlegung ihrer Gram-Matrizen $\mathbf{A}\mathbf{A}^\top$ sowie $\mathbf{A}^\top\mathbf{A}$ her.

Lösung

Sei $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ eine volle SVD. Dann gilt für die Gram-Matrizen:

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

Damit gilt:

- Spalten von $\mathbf{U}$: Eigenvektoren von $\mathbf{A}\mathbf{A}^\top$
- Spalten von $\mathbf{V}$: Eigenvektoren von $\mathbf{A}^\top\mathbf{A}$
- Singulärwerte $\sigma_1, \ldots, \sigma_k$: Wurzeln der Eigenwerte von $\mathbf{A}^\top\mathbf{A}$ bzw. $\mathbf{A}\mathbf{A}^\top$

Hierbei ist erwähnenswert, dass $\mathbf{A}\mathbf{A}^\top$ und $\mathbf{A}^\top\mathbf{A}$ symmetrisch und positiv semidefinit sind. Damit sind ihre Eigenwerte reell und nicht-negativ. Außerdem besitzen sie, bis auf zusätzliche Nullen, die gleichen Eigenwerte.

Lösung Ende

5. Wie kann mit der SVD der Rang und der Kern einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ bestimmt werden?

Lösung

Der Kern ist die Lösungsmenge des homogenen LGS $\mathbf{A}\mathbf{x} = \mathbf{0}$.

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{0},\ \ \mathbf{x} \neq \mathbf{0} \\ \Leftrightarrow \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{x} &= \mathbf{0} \\ \Leftrightarrow \mathbf{\Sigma}\mathbf{V}^\top\mathbf{x} &= \mathbf{0} \\ \Leftrightarrow \mathbf{\Sigma}\mathbf{y} &= \mathbf{0} \\ \Leftrightarrow (\sigma_1 y_1, \ldots, \sigma_n y_n)^\top &= \mathbf{0} \end{aligned}$$

Dabei gilt, dass die Singulärwerte geordnet sind: $\sigma_1 \geq \ldots \geq \sigma_n$. Sei k die größte Zahl, sodass $\sigma_k > 0$. Es ist klar, dass alle Einträge $y_1, \ldots, y_k$ gleich Null sein müssen. Alle anderen Eintrage sind frei wählbar, zum Beispiel so, dass ein Eintrag auf Eins und alle anderen Einträge auf Null gesetzt sind. Da $\mathbf{x} = \mathbf{V}\mathbf{y}$, bilden die letzten Spalten von $\mathbf{V}$ (die zu einem Singulärwert $\sigma_i = 0$ gehören) eine Basis des Kerns. Der Rang entspricht der Anzahl an Singulärwerten ungleich Null.

Was passiert, falls die Matrix nicht quadratisch ist? Dann existieren genauso viele Singulärwerte wie das Minimum aus Zeilen und Spalten. Falls die Matrix mehr Zeilen als Spalten hat, dann gilt die analoge Aussage wie im quadratischen Fall. Falls sie mehr Spalten hat, sind die zusätzlichen rechten Singulärvektoren (in der vollen Form der SVD), also die, die zu keinem Singulärwert gehören (bzw. können sie ebenso zum Singulärwert Null gehören, da sie Eigenvektoren von $\mathbf{A}^\top\mathbf{A}$ zum Eigenwert Null sind), ebenso im Kern.

Lösung Ende

Die (Moore-Penrose) Pseudoinverse ist eine Verallgemeinerung der Inversen auf singuläre und nicht-quadratische Matrizen beliebigen Ranges. Für eine Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ ist die eindeutig bestimmte Matrix $\mathbf{A}^+ \in \mathbb{R}^{m\times n}$ ihre Pseudoinverse, falls $\mathbf{A}^+$ die folgenden Bedingungen erfüllt:

1) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, 2) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$, 3) $\mathbf{A}\mathbf{A}^+$ ist symmetrisch, 4) $\mathbf{A}^+\mathbf{A}$ ist symmetrisch.

5. Wie kann man die Pseudoinverse einer Matrix mithilfe der SVD berechnen?

Lösung

Man berechnet $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^\top$ mit

$$\mathbf{\Sigma}^+ = \begin{bmatrix} \sigma_1^+ & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_k^+ \end{bmatrix} \quad \text{mit} \quad \sigma_i^+ = \begin{cases} 1/\sigma_i & \text{falls } \sigma_i \neq 0 \\ 0 & \text{falls } \sigma_i = 0. \end{cases}$$

Die Pseudoinverse einer regulären Matrix ist damit ihre Inverse.

Lösung Ende

6. Wie kann man die Psudoinverse nutzen, um lineare Gleichungssysteme $\mathbf{Ax} = \mathbf{b}$ zu lösen? Welche Lösung(en) erhält man, falls das LGS keine, genau eine bzw. unendlich viele Lösungen hat?

Lösung

Man kann das LGS durch $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ lösen. Man erhält, egal welche Eigenschaften $\mathbf{A}$ und $\mathbf{b}$ haben, immer eine Lösung. Falls $\mathbf{Ax} = \mathbf{b}$ ...

- genau eine Lösung hat, wird die Lösung auch errechnet.
- keine Lösung hat, entspricht die gefundene Lösung dem Ergebnis der Normalengleichung.
- unendlich viele Lösungen hat, wird die Lösung mit der kleinsten (euklidischen) Norm $\|\mathbf{x}\|$ bestimmt.

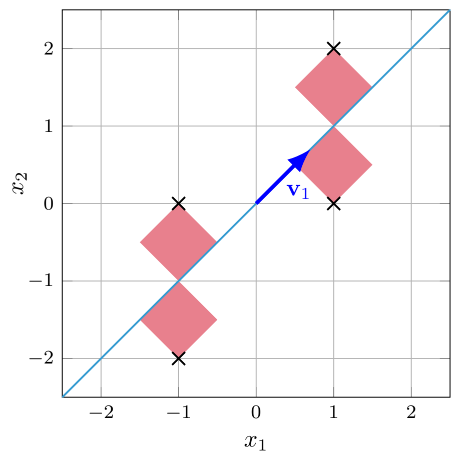
Die Pseudoinverse ist numerisch sehr stabil, aber dessen Berechnung hat eine hohe Laufzeit.

Lösung Ende

Aufgabe 2: Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA) ist ein statistisches Verfahren, bei dem ein zentrierter Datensatz $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ mithilfe eines k-dimensionalen linearen Unterraumes des $\mathbb{R}^d$ bestmöglich approximiert werden soll, wobei üblicherweise gilt, dass $k \ll d$. Gesucht ist eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ (bei der wir ohne Beschränkung der Allgemeinheit annehmen können, dass sie orthonormal ist), dessen Spann möglichst nah an allen Punkten ist, also folgende Fehlerfunktion minimiert: $\mathcal{E}(\mathbf{V}) = \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{V}\mathbf{V}^\mathsf{T}\mathbf{x}_i\|^2$ mit $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$. Dabei ist $\mathbf{V}\mathbf{V}^\mathsf{T}\mathbf{x}_i$ die orthogonale Projektion des Punktes $\mathbf{x}_i$ in den Spaltenraum von $\mathbf{V}$.

1. Gegeben ist ein Basisvektor $\mathbf{v}_1$ mit dem von ihm aufgespannten linearen Unterraum $\mathrm{Span}(\mathbf{v}_1)$ und eine Menge von Punkten. Zeichnen Sie die Fehler ein, die von PCA minimiert wird.



2. Wozu wird PCA im Kontext der Gesichtserkennung angewandt?

Lösung

PCA ist kein Algorithmus zur Gesichtserkennung, sondern dient lediglich zur Dimensionsreduktion. Dabei werden die mittelwertbefreiten Gesichtsbilder auf einen möglichst klein dimensionalen linearen Unterraum projiziert und dann mit dem Nearest-Neighbor Algorithmus einer Person zugeordnet. Dies ist ein notwendiger Schritt, da (Gesichts)Bilder eine hohe Dimensionalität aufweisen.

Lösung Ende

3. Erklären Sie jeden Schritt des Eigenface-Algorithmus zur Gesichtserkennung.

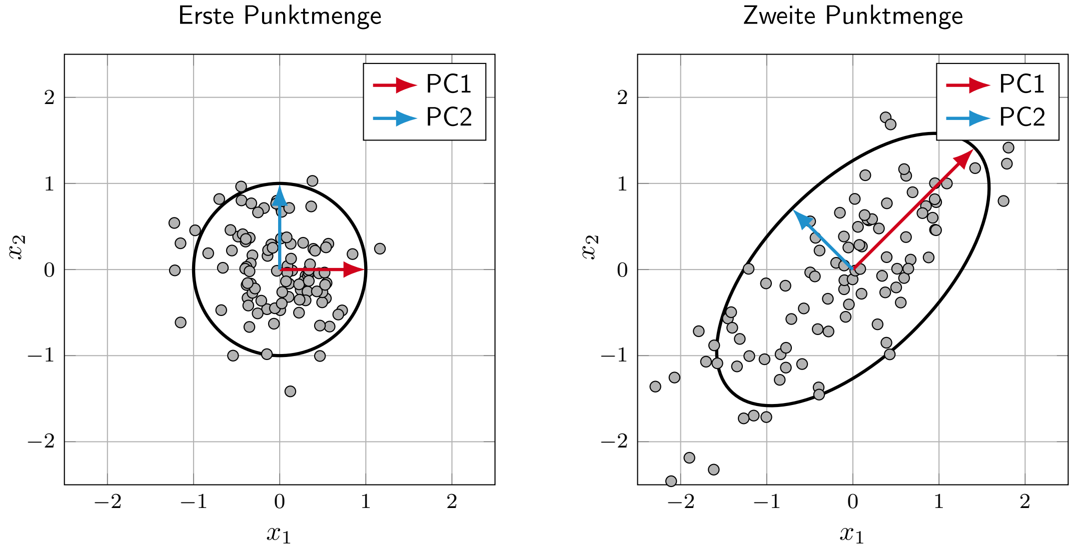
Lösung

Gegeben sind eine Menge von m Gesichtsbildern $\mathbf{g}_1, \ldots, \mathbf{g}_m$(in Graustufe) mit Breite w und Höhe h als Vektoren des $\mathbb{R}^n$ mit $n = wh$.

a) Berechnung des Durchschnitts der Gesichtsbilder: $\overline{\mathbf{g}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{g}_i$.
b) Mittelwertbefreiung: $\tilde{\mathbf{g}_i} = \mathbf{g}_i - \overline{\mathbf{g}}$.
c) Aufstellen der Datenmatrix:
$$\mathbf{X} = \begin{bmatrix} — & \tilde{\mathbf{g}_1}^\mathsf{T} & — \\ & \vdots & \\ — & \tilde{\mathbf{g}_m}^\mathsf{T} & — \end{bmatrix}$$
Jede Zeile der Datenmatrix entspricht einem mittelwertbefreitem Gesicht.
d) Berechnung der SVD $\mathbf{X} = \mathbf{U\Sigma V}^\mathsf{T}$.
e) Analyse des Spektrums der Singulärwerte: Kleine Singulärwerte bzw. ihre zugehörigen Singulärvektoren können vernachlässigt werden, da sie kaum Informationen zu den Daten beinhalten. Man wählt die $k \ll n$ größten Singulärwerte aus.
f) Sei $\mathbf{V}_k$ die Matrix, die die ersten k Spalten von $\mathbf{V}$ beinhaltet. Wir projizieren jeden Datenpunkt in den Unterraum von $\mathbf{V}_k$: $\mathbf{v}_i = \mathbf{V}_k^\mathsf{T}\tilde{\mathbf{g}_i}$.
g) Für jeden neuen Datenpunkt $\mathbf{g}^* \in \mathbb{R}^n$: Man befreit ihn vom Mittelwert der Trainingsdaten und projiziert den ebenso in den Raum von $\mathbf{V}_k$: $\mathbf{v}^* = \mathbf{V}_k^\mathsf{T}(\mathbf{g}^* - \overline{\mathbf{g}})$. Nun vergleicht man die Projektion des neuen Punktes mit denen der ursprünglichen Daten $\{\mathbf{v}_i\}$ mit einer geeigneten Abstandsfunktion.

Lösung Ende

4. Im Folgenden sind zwei zentrierte Punktemengen im $\mathbb{R}^2$ abgebildet. Zeichnen Sie die erste und zweite Hauptkomponente, falls möglich, ein. Was können Sie über das Verhältnis der Singulärwerte der zugehörigen Datenmatrix sagen?



Lösung

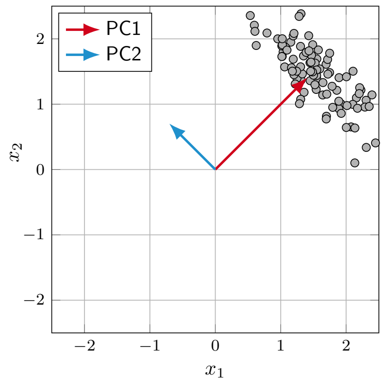
Im ersten Bild sind die Hauptkomponenten nicht eindeutig bestimmt, da die Singulärwerte (fast) gleich groß sind (sie können beliebig rotiert werden). Im zweiten Bild gilt $\sigma_1 \gg \sigma_2$.

Lösung Ende

5. Da beim PCA Algorithmus angenommen wird, dass die Datenpunkte zentriert sind, werden die Beobachtungen vom Mittelwert befreit. Was kann passieren, falls dieser Schritt ausgelassen wird?

Lösung

Die erste Hauptkomponente zeigt, falls genügend Punkte vorhanden sind, in die Mitte der Punktwolke (oder in die entgegengesetzte Richtung).



Lösung Ende