Wissenschaftliches Rechnen - Großübung 3.2

Themen: Singulärwertzerlegung, Hauptkomponentenanalyse

Ugo & Gabriel

6. Dezember 2022

Aufgabe 1: Singulärwertzerlegung

1. Die Singulärwertzerlegung ist eine Faktorisierung einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ in $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$. Im Folgenden betrachten wir ausschließlich die volle Form.

   a) Was steht in den Spalten von $\mathbf{U}$ und $\mathbf{V}$? Welche Eigenschaft haben diese Matrizen?

   b) Wie sieht $\mathbf{\Sigma}$ aus?

   c) Was für eine Art von Transformation beschreiben $\mathbf{U}, \mathbf{V}$ und $\mathbf{\Sigma}$?

   d) Welche Eigenschaft muss eine Matrix haben, damit sie wie oben beschrieben faktorisiert werden kann?

   Lösung

   a) In den Spalten von $\mathbf{U} \in \mathbb{R}^{n\times n}$ stehen die linken Singulärvektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$, welche Eigenvektoren von $\mathbf{A}\mathbf{A}^\mathsf{T}$ sind. In den Spalten von $\mathbf{V} \in \mathbb{R}^{m\times m}$ stehen die rechten Singulärvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_m$, welche Eigenvektoren von $\mathbf{A}^\mathsf{T}\mathbf{A}$ sind. Beide Matrizen sind orthogonal.

   b) $\mathbf{\Sigma} \in \mathbb{R}^{n\times m}$ ist (fast) eine Diagonalmatrix. Auf ihrer Diagonalen stehen die Singulärwerte $\sigma_1, \ldots, \sigma_k$ mit $k = \min\{n, m\}$, welche der Wurzeln der Eigenwerte von $\mathbf{A}\mathbf{A}^\mathsf{T}$ bzw. $\mathbf{A}^\mathsf{T}\mathbf{A}$ entsprechen. Da die Gram-Matrizen positiv semidefinit sind, sind die Singulärwerte nicht-negativ und per Konvention ordnet man diese absteigend: $\sigma_1 \geq \ldots \geq \sigma_k$.

   c) $\mathbf{U}$ und $\mathbf{V}$ sind Rotationen und/oder Spiegelungen, $\mathbf{\Sigma}$ ist eine Skalierung, wobei sich die Dimension des Raumes ändern kann.

   d) Alle Matrizen haben eine Singulärwertzerlegung.

   Lösung Ende

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

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

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

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

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

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

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

Keine SVD, da die Singulärwerte nicht korrekt geordnet sind.

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

Eine korrekte SVD.

3. Sei $\mathbf{A} \in \mathbb{R}^{n\times n}$ quadratisch und $\mathbf{A} = \mathbf{U\Sigma V}^\mathsf{T}$ eine Singulärwertzerlegung:

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

Entscheide, ob die folgenden Aussagen im Allgemeinen gelten oder nicht.

a) $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ Korrekt

b) $\mathbf{A}\mathbf{u}_i = \sigma_i \mathbf{v}_i$ Falsch

c) $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{v}_i$ Falsch

d) Falls $\sigma_i > 0$ und $\sigma_{i+1} = \ldots = \sigma_n = 0$, dann ist $\mathbf{v}_{i+1}, \ldots, \mathbf{v}_n$ eine Basis des Kerns von $\mathbf{A}$ Korrekt

e) Falls $\sigma_i > 0$ und $\sigma_{i+1} = \ldots = \sigma_n = 0$, dann ist $\mathbf{u}_{i+1}, \ldots, \mathbf{u}_n$ eine Basis des Kerns von $\mathbf{A}^\mathsf{T}$ Korrekt

f) $\det \mathbf{A} = \prod_{i=1}^n \sigma_i$ Falsch

g) Falls $\mathbf{A}$ symmetrisch ist, gilt $\mathbf{U} = \mathbf{V}$ Falsch

h) Falls $\mathbf{A}$ orthogonal ist, gilt $\sigma_1 = \sigma_2 = \ldots = 1$ Korrekt

i) Falls $\mathbf{A}$ eine Diagonalmatrix ist, gilt $\mathbf{U} = \mathbf{V} = \mathbf{I}$ Falsch

j) $\mathbf{A}^2 = \mathbf{U\Sigma}^2\mathbf{V}^\mathsf{T}$ Falsch

4. Wie kann man anhand der Singulärwerte den Rang einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ bestimmen?

Lösung

Der Rang entspricht der Anzahl der Singulärwerte größer als Null.

Lösung Ende

5. Die SVD einer Matrix ist nicht eindeutig bestimmt. Gib eine Matrix und zwei unterschiedliche SVDs dazu an.

Lösung

Wir machen es uns mal ganz einfach:

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{V}^\mathsf{T}}$$

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}}_{\mathbf{V}^\mathsf{T}}$$

Lösung Ende

6. Gibt es Matrizen, welche unendlich viele SVDs haben? Falls ja, unter welchen Bedingungen?

Lösung

Ja, solche Matrizen gibt es, beispielsweise die Identität, die man in $\mathbf{I} = \mathbf{UIU}^\mathsf{T}$ faktorisieren kann, wobei $\mathbf{U}$ eine beliebige orthogonale Matrix ist.

Eine Matrix hat unendlich viele SVDs, falls es einen doppelten Singulärwert gibt, denn dann kann man sich die Basis des Spannes der zugehörigen Singulärvektoren (um eine orthogonale Transformation) frei wählen.

Lösung Ende

7. Erinnerung: Die Kondition einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist definiert als

$$\kappa(\mathbf{A}) = \frac{\max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|}{\min_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|}.$$

Zeigen Sie, dass die Kondition von $\mathbf{A}$ bezüglich der ℓ^2-Norm berechnet werden kann durch $\kappa(\mathbf{A}) = \frac{\sigma_1}{\sigma_n}$.

Lösung

$$\begin{aligned} \kappa(\mathbf{A}) &= \frac{\max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|}{\min_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|} \\ &= \frac{\max_{\|\mathbf{x}\|=1} \|\mathbf{U\Sigma V}^\mathsf{T}\mathbf{x}\|}{\min_{\|\mathbf{x}\|=1} \|\mathbf{U\Sigma V}^\mathsf{T}\mathbf{x}\|} \\ &= \frac{\max_{\|\mathbf{x}\|=1} \|\mathbf{\Sigma V}^\mathsf{T}\mathbf{x}\|}{\min_{\|\mathbf{x}\|=1} \|\mathbf{\Sigma V}^\mathsf{T}\mathbf{x}\|} \\ &= \frac{\max_{\|\mathbf{y}\|=1} \|\mathbf{\Sigma y}\|}{\min_{\|\mathbf{y}\|=1} \|\mathbf{\Sigma y}\|} \\ &= \frac{\sigma_1}{\sigma_n} \end{aligned}$$

Lösung Ende

8. Sei $\mathbf{A} \in \mathbb{R}^{n\times m}$ eine Matrix und $\mathbf{A}^+$ ihre Pseudoinverse. Entscheide für die folgenden Aussagen, ob sie gelten oder im Allgemeinen falsch sind.

   a) $\mathbf{A}^+\mathbf{A} = \mathbf{I}$ Falsch.

   b) $\mathbf{AA}^+ = \mathbf{I}$ Falsch.

   c) $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$ Korrekt.

   d) $\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$ Korrekt

9. Was ist die Pseudoinverse eines (Spalten-)Vektors $\mathbf{v} \in \mathbb{R}^n$?

Lösung

Man kann einen Spaltenvektor wie folgt zerlegen:

$$\mathbf{v} = \mathbf{U\Sigma V}^\mathsf{T}$$

mit

$$\mathbf{U} = \left(\frac{\mathbf{v}}{\|\mathbf{v}\|}, \mathbf{u}_2, \dots, \mathbf{u}_n\right), \quad \mathbf{\Sigma} = \begin{bmatrix} \|\mathbf{v}\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{V} = (1)$$

wobei $\mathbf{u}_2, \dots, \mathbf{u}_n$ beliebige zu $\mathbf{v}$ und zu einander orthogonale sowie normierte Vektoren sind (ja, die volle Form ist unnötig aufwendig).

Die Pseudoinverse ist dann $\mathbf{v}^+ = \mathbf{V\Sigma}^+\mathbf{U}^\mathsf{T}$ mit

$$\mathbf{\Sigma}^+ = \begin{bmatrix} 1/\|\mathbf{v}\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Damit ist $\mathbf{v}^+ = \frac{\mathbf{v}^\mathsf{T}}{\|\mathbf{v}\|^2}$.

Lösung Ende

Aufgabe 2: Hauptkomponentenanalyse

Ziel der Hauptkomponentenanalyse ist es für eine Menge an Datenpunkten $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ eine reduzierte Basis mit $k \ll n$ Basisvektoren zu finden, sodass die Daten bezüglich dieser Basis möglichst exakt dargestellt werden können. Für einen Basisvektor (die sog. erste Hauptkomponente) lautet das Optimierungsproblem:

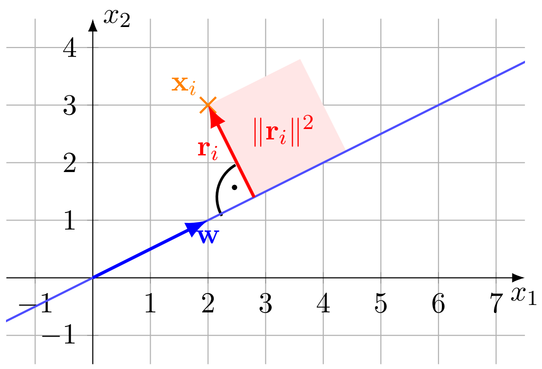
$$\min_{\mathbf{w}} \quad \sum_{i=1}^{n} \|\mathbf{x}_i - \langle \mathbf{x}_i, \mathbf{w} \rangle \mathbf{w}\|^2 \quad \text{s.t.} \quad \|\mathbf{w}\| = 1$$

Dabei nehmen wir an, dass die Daten zentiert sind, d.h. $\sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$.

1. Welche geometrische Bedeutung hat die Fehlerfunktion, die minimiert wird?

Lösung

Wir minimieren die Summe der quadratischen Rekonstruktionsfehler aller Punkte, wenn man sie in den linearen Unterraum von $\mathbf{w}$ projiziert:



Lösung Ende

2. Das Optimierungsproblem lässt sich als Maximierungsproblem umschreiben. Wie lautet für dieses die Zielfunktion?

Lösung

$$\max_{\mathbf{w}} \quad \mathbf{w}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{X} \mathbf{w} \quad \text{s.t.} \quad \|\mathbf{w}\| = 1$$

Die Matrix $\mathbf{X}$ enthält die Punkte in ihren Zeilen:

$$\mathbf{X} = \begin{bmatrix} — & \mathbf{x}_1^{\mathsf{T}} & — \\ & \vdots & \\ — & \mathbf{x}_n^{\mathsf{T}} & — \end{bmatrix}$$

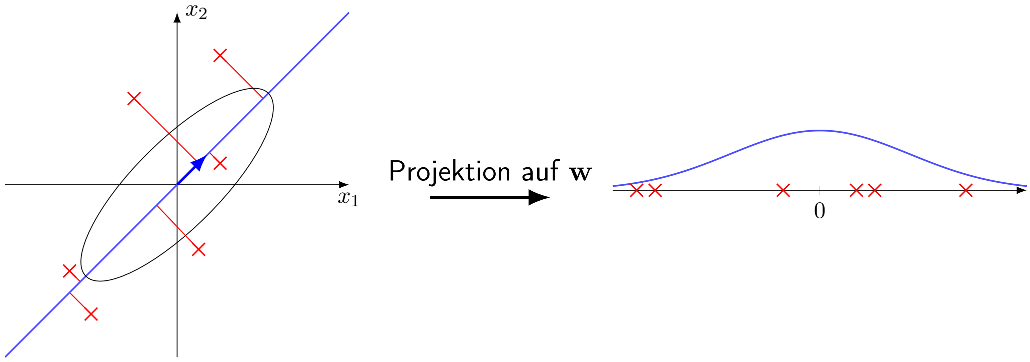
Lösung Ende

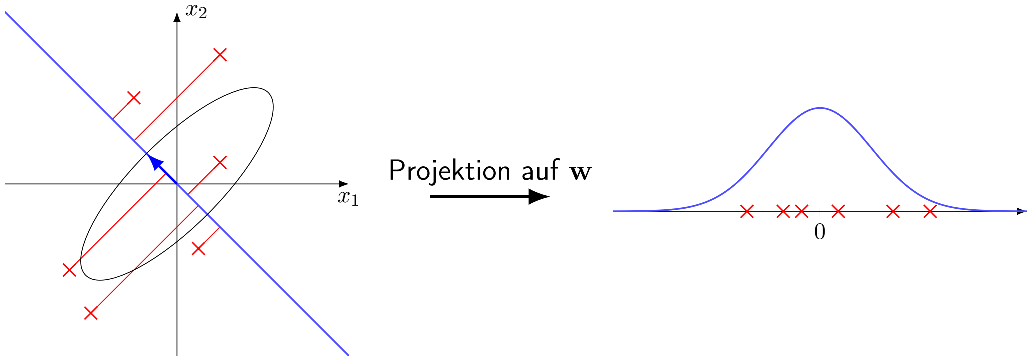
3. Welche Bedeutung hat die Zielfunktion des Maximierungsproblems?

Lösung

Wir maximieren die Varianz der Projektion auf den Vektor $\mathbf{w}$.

Hier eine coole Visualisierung, die den Rekonstruktionsfehler mit der Varianz der Projektion verknüpft:





Je kleiner der Fehler, umso größer die Varianz. Die Zielfunktion $\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w}$ ist (bis auf den konstanten Faktor n) die empirische Varianz der Daten in Richtung von $\mathbf{w}$.

Lösung Ende

4. Zeigen Sie, dass die Lösung des Problems ein Eigenvektor einer bestimmten Matrix ist. Welchen Zusammenhang zu Singulärvektoren hat die Lösung?

Lösung

Nach dem letzten Blatt ist die Lösung dieses Optimierungsproblems der Eigenvektor von $\mathbf{X}^\mathsf{T}\mathbf{X}$ zum größten Eigenwert λ_1. Dies entspricht dem ersten rechten Singulärvektor $\mathbf{v}_1$ von $\mathbf{X}$.

Lösung Ende

5. Die Lösung der obigen Optimierungsprobleme ist die erste Hauptkomponente. Wie erhält man die zweite Hauptkomponente? Welche Eigenschaft hat diese?

Lösung

Die zweite Hauptkomponente ist die Richtung mit kleinstem Fehler bzw. größter Varianz, wenn man die Projektion der Daten in den linearen Unterraum des ersten Singulärvektors abzieht ($\mathbf{X}' = \mathbf{X} - \mathbf{X}\mathbf{v}_1\mathbf{v}_1^\mathsf{T}$).

Was passiert bei der Operation $\mathbf{X} - \mathbf{X}\mathbf{v}_1\mathbf{v}_1^\mathsf{T}$? Da die Hauptkomponenten eine Basis bilden, kann jeder Punkt $\mathbf{x}$ als Linearkombination dieser geschrieben werden:

$$\mathbf{x} = \alpha_1\mathbf{v}_1 + \ldots + \alpha_d\mathbf{v}_d$$

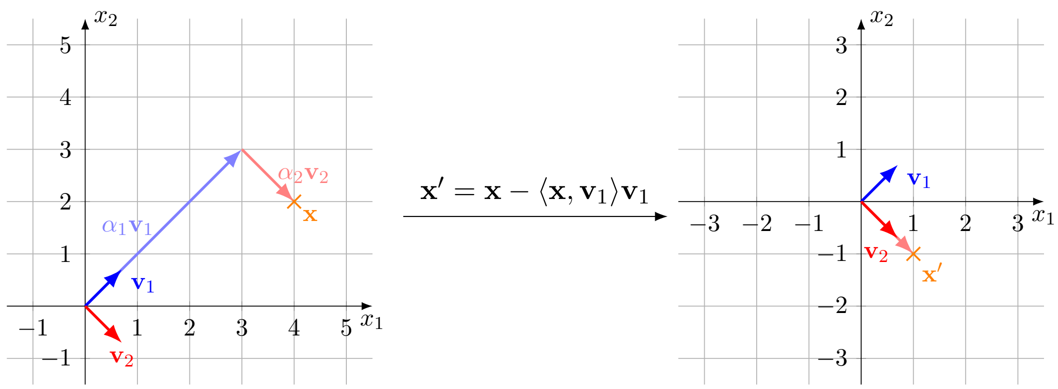
Dabei ist α_i der Anteil der i-ten Hauptkomponenten am Punkt $\mathbf{x}$. Dabei sollen die ersten α_i einen möglichst großen Wert haben und die letzten α_i einen möglichst kleinen (schließlich möchte man, dass die ersten Hauptkomponenten möglichst viel der Daten rekonstruieren können). Dies erreicht man, indem die Hauptkomponenten keinen Anteil zu einander haben, also orthogonal sind. Da die Hauptkomponenten zur Orthogonalität noch Einheitslänge haben, ergibt die Operation $\langle \mathbf{x}, \mathbf{v}_1 \rangle$ (also das Skalarprodukt) den Anteil an der ersten Hauptkomponente:

$$\langle \mathbf{x}, \mathbf{v}_1 \rangle = \alpha_1$$

Streckt man die erste Hauptkomponente um ihren Anteil (also das Skalarprodukt), erhält man $\langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 = \alpha_1 \mathbf{v}_1$. Zieht man das vom ursprünglichen Punkt ab, bleibt nur die Komponente aller anderen Hauptkomponenten übrig:

$$\mathbf{x}' = \mathbf{x} - \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 = \alpha_2 \mathbf{v}_2 + \ldots + \alpha_d \mathbf{v}_d$$

Dies liefert dasselbe Ergebnis, als hätte man eine SVD der Daten $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ berechnet, den ersten Singulärwert auf 0 gesetzt und die Daten dann wieder rekonstruiert $\mathbf{X}' = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^\mathsf{T}$. $\mathbf{X}' = \mathbf{X} - \mathbf{X}\mathbf{v}_1\mathbf{v}_1^\mathsf{T}$ entspricht also den Datenpunkten, nachdem man den Anteil der 1. Hauptkomponenten für jeden Punkt eliminiert.



Alternative Formulierung: Die zweite Hauptkomponente ist die Richtung der größten Varianz in den Daten, die orthogonal zur ersten Hauptkomponenten ist.

Die zweite Hauptkomponente ist der 2. rechte Singulärvektor $\mathbf{v}_2$ von $\mathbf{X}$, die dritte Hauptkomponente ist der 3. rechte Singulärvektor $\mathbf{v}_3$, usw.

Lösung Ende

6. Welche Bedeutung hat ein Singulärwert σ_i bzw. ein Eigenwert $\lambda_i = \sigma_i^2$?

Lösung

Sei $\mathbf{v}_i$ die i-te Hauptkomponente der Daten $\mathbf{X}$ und σ_i sowie $\lambda_i = \sigma_i^2$ der zugehörige Singulär- bzw. Eigenwert.

	Rekonstruktionsfehler	Varianz
Eigenwert λ_i von $\mathbf{X}^\mathsf{T}\mathbf{X}$	Fehler, falls man Hauptkomponente $\mathbf{v}_i$ weglässt	n mal Varianz der Projektion der Daten auf $\mathbf{v}_i$
Singulärwert σ_i von $\mathbf{X}$	Wurzel des Fehlers, falls man Hauptkomponente $\mathbf{v}_i$ weglässt	$\sqrt{n}$ mal Standardabweichung der Projektion der Daten auf $\mathbf{v}_i$

Lösung Ende

7. Sie führen eine Hauptkomponentenanalyse auf einem dreidimensionalen Datensatz durch und berechnen die Singulärwerte $\sigma_1 = 10$, $\sigma_2 = \sigma_3 = 0$. Falls Sie den Datensatz visualisieren: Was würden Sie feststellen?

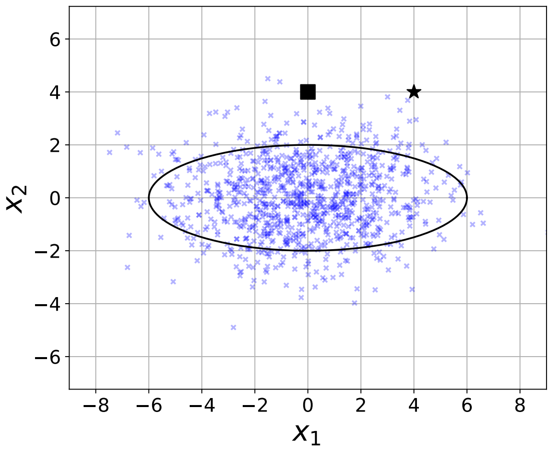
Lösung

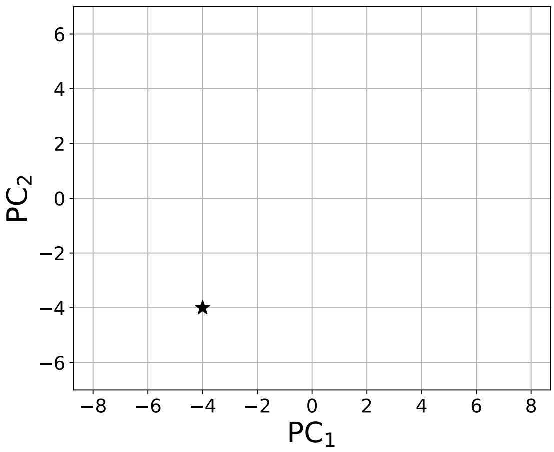
Da es nur einen Singulärwert ungleich Null gibt, heißt es, dass die Daten in einem eindimensionalen Unterraum leben. Damit liegen die Datenpunkte auf einer Gerade.

Lösung Ende

8. Die folgende Geschichte ist garantiert so passiert: Ihr Freund führt eine Hauptkomponentenanalyse auf einer zweidimensionalen zentrierten Punktemenge durch. Leider hat er das Ergebnis verloren und weiß nur, dass er einen Punkt, der die Form eines Sterns hat, in den Raum der PCs projiziert hat. Nun wendet er sich an Sie und zeigt Ihnen seinen Datensatz mit der statistischen Verteilung sowie den Stern und dessen Projektion in den von den PCs aufgespannten linearen Unterraum. Helfen Sie Ihrem Freund beim Wiederherstellen seiner Ergebnisse!

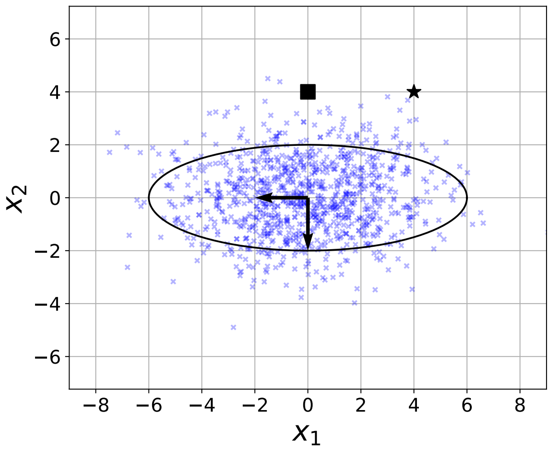
   Hinweis: Sie sollen genau dieselben Hauptkomponenten berechnen, die Ihr Freund auch berechnet hat.

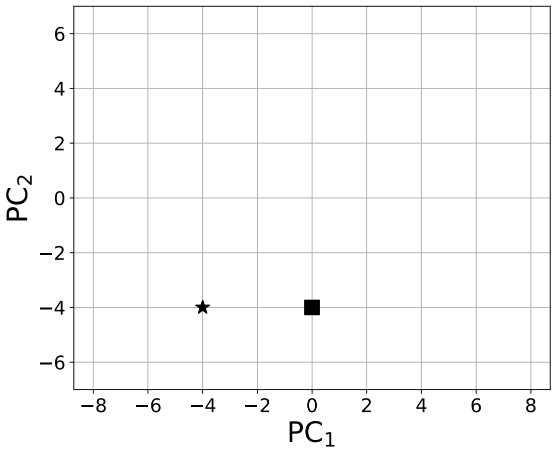




   a) Zeichnen Sie die Hauptkomponenten, die Ihr Freund berechnet hat, in das linke Bild.
   b) Wo landet das Quadrat, wenn man es in den Raum der Hauptkomponenten rein projiziert? Zeichnen Sie dessen Projektion in das rechte Bild.

Lösung





Lösung Ende