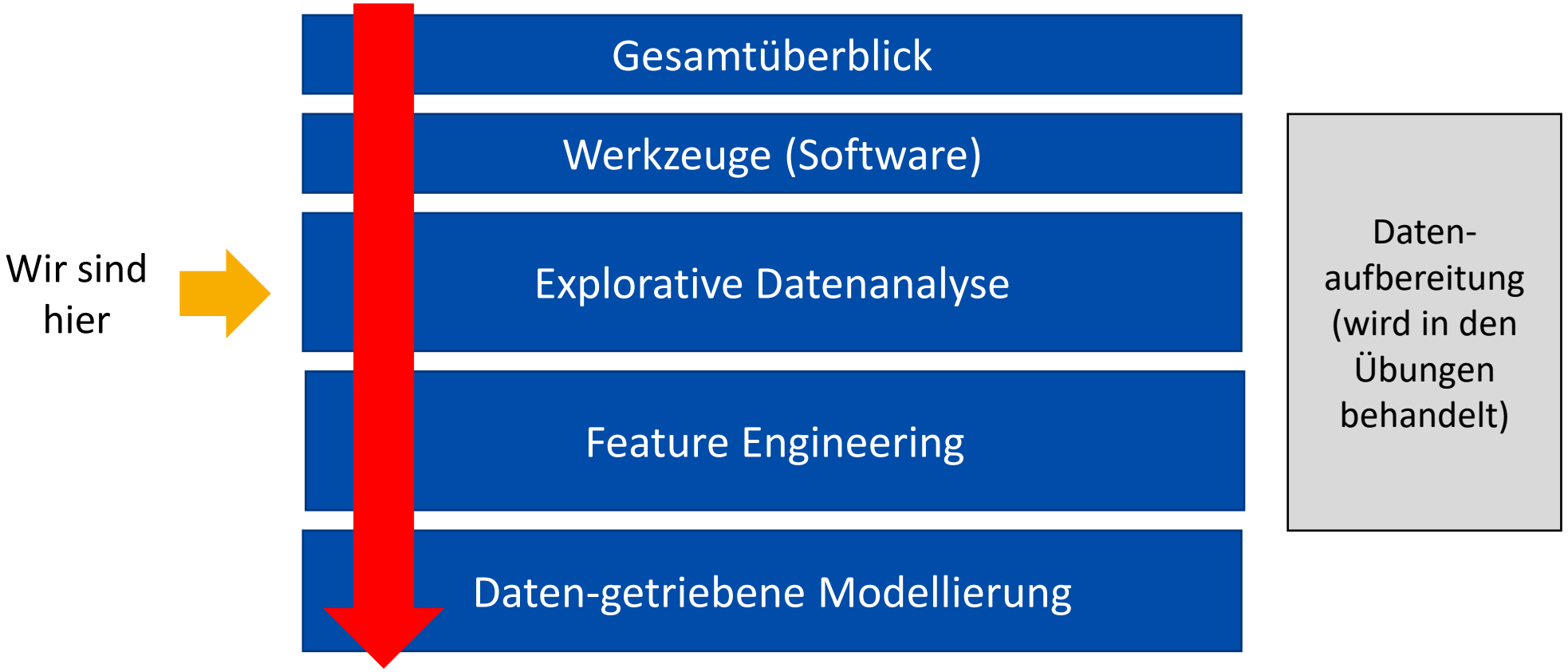
Einführung in Data Science

Unser Plan für heute:

1. Wiederholung

2. Dimensionsreduktionsverfahren
   - Hauptkomponentenzerlegung (PCA)

Data Science



Curriculum

1. Grundbegriffe / Überblick
2. Zentrale Softwarebibliotheken

3. Univariate explorative Analyse (EDA) Visualisierung (Teil 1)
4. Visualisierung (Teil 2), Multivariate explorative Analyse (Teil 1)
5. Multivariate explorative Analyse (Teil 2)
6. Dimensionsreduktion (Teil 1): PCA
7. Dimensionsreduktion (Teil 2): MDS, Isomap
8. Clustering: K-Means, HCA
9. Clustervalidierung

10. Probeklausur

11. Feature Engineering, Datengetriebene Modellierung (Teil 1)
12. Datengetriebene Modellierung (Teil 2)

Überblick / Begriffe

Explorative Analyse (EDA)

Feature Engineering & Modellierung

Multivariate Explorative Analyse

Methoden der multivariaten explorativen Analyse,
die Sie bisher kennengelernt haben:

1. Multivariate deskriptive Statistik
   (hier vor allem: Visualisierungsarten)

2. Korrelationskoeffizienten
   (bivariate Analyse: Suche nach Zusammenhängen)

Nun:

3. Dimensionsreduktionsverfahren

M‘variate Explorative Analyse | Dimensionsreduktion

In der Praxis:

- Daten sind häufig hochdimensional:
  Jeder Datenpunkt (jedes Objekt) hat oft viele Merkmale (Features)

Beispiel (kennen Sie aus der Übung):

	country	continent	year	lifeExp	pop	gdpPercap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710

Merkmale: Land, Kontinent, Jahr, Lebenserwartung, Populationsgröße, Bruttoinlandsprodukt/Kopf

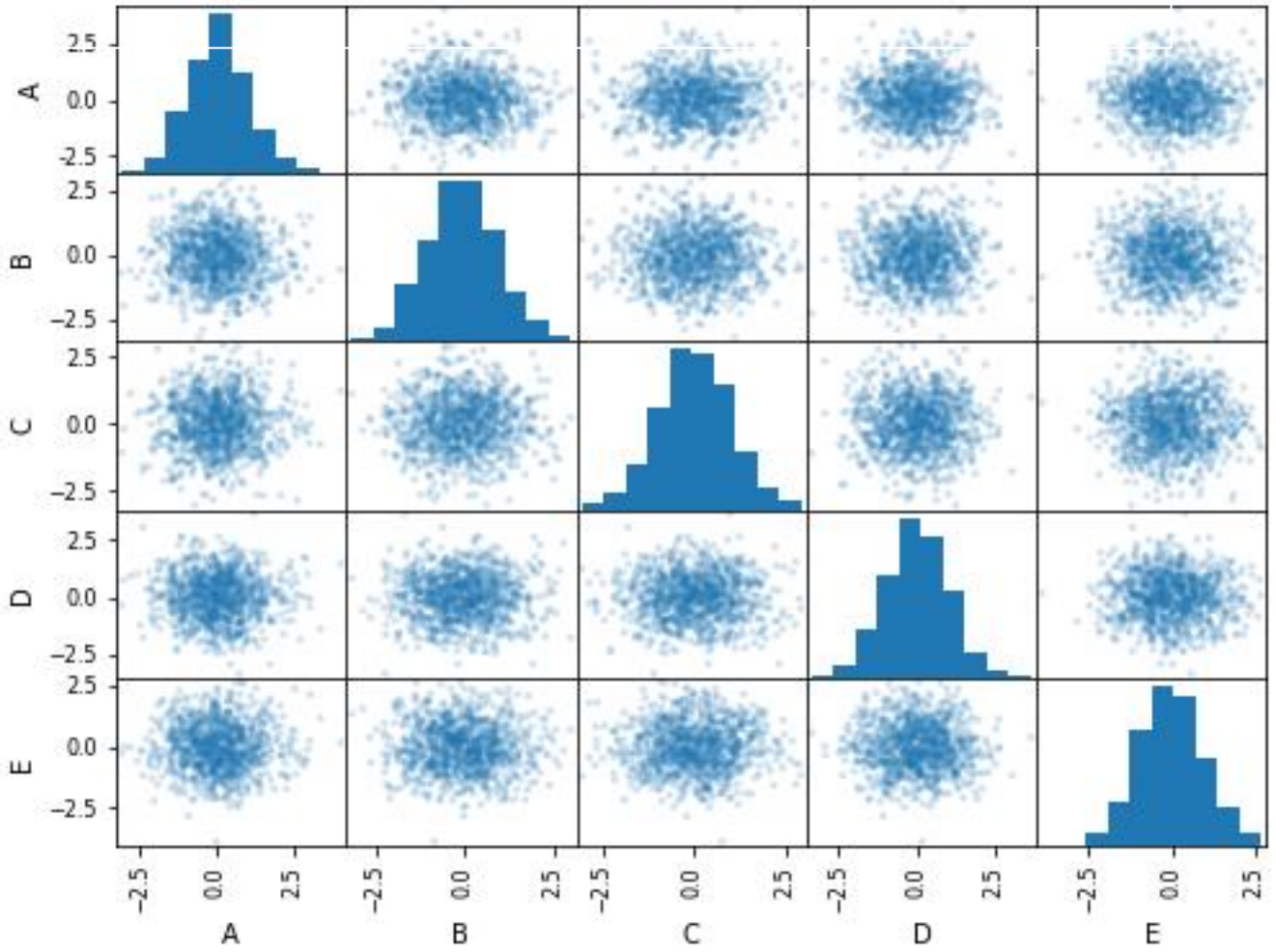
Erkundung (z.B. durch Visualisierung) hochdimensionaler Daten ist anspruchsvoll.

M‘variate Explorative Analyse | Dimensionsreduktion

Erkundung hochdimensionaler Daten durch Visualisierung durch viele Scatterplots:

Beispiel:

1000 Datenpunkte mit je 5 Features (also: 5-dimensionaler Merkmalsraum)

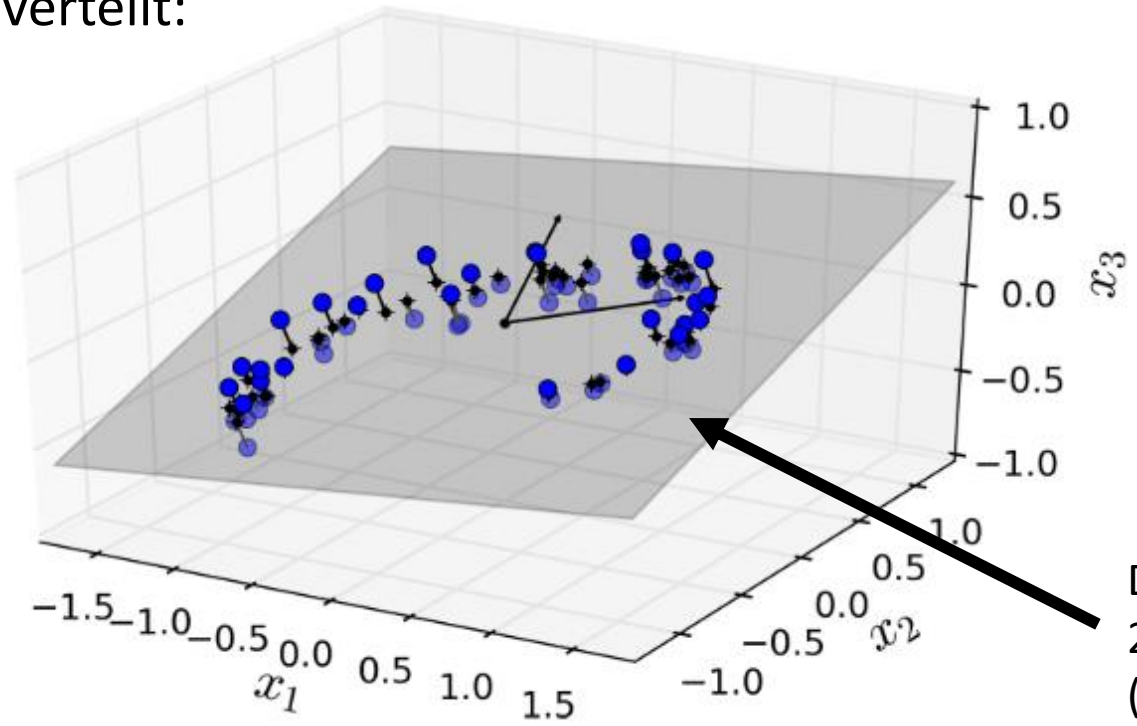


M‘variate Explorative Analyse | Dimensionsreduktion

Dimensionsreduktionstechniken

- reduzieren die Anzahl der Merkmale (Dimensionen)
  (oft mit dem Ziel, sowenig Informationen dabei zu verlieren)

Viele Daten sind nicht uniform im Merkmalsraum (Feature-Raum) verteilt:



Manche Merkmale sind miteinander stark korreliert

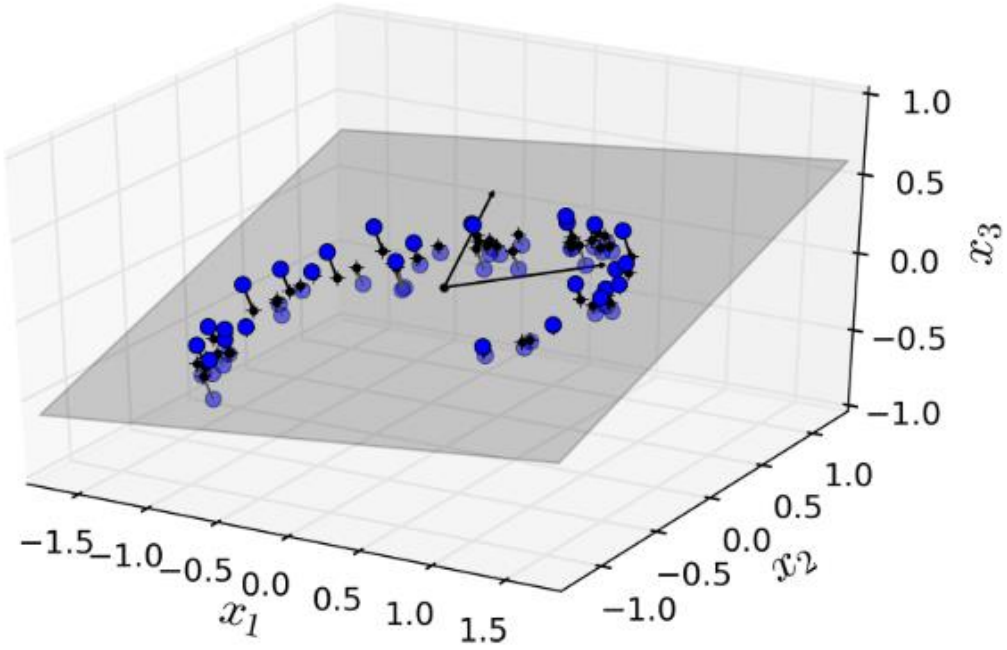
→ Datenpunkte liegen oft in einem Unterraum

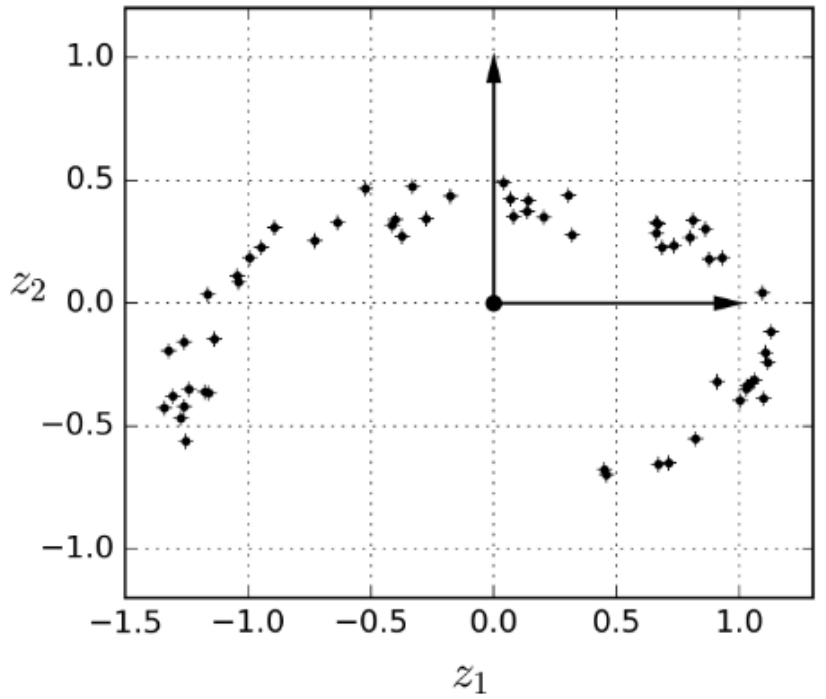
Daten liegen hier ungefähr in einem 2-dimensionalen Unterraum (= graue Fläche).

M‘variate Explorative Analyse | Dimensionsreduktion

Dimensionsreduktion durch Projektion jedes Punktes auf einen Unterraum

3-dimensionaler Merkmalsraum → 2-dimensionaler Merkmalsraum





alte Merkmale (Features)

neue, transformierte Merkmale

Dimensionsreduktion | PCA

Hauptkomponentenzerlegung (*principal component analysis, **PCA***)

- zählt zu den bekanntesten Dimensionsreduktionsverfahren
- 1901 von Karl Pearson publiziert[1]

Kernideen

- PCA findet neue Achsen (Komponenten) für die Daten, so dass die Daten bezüglich dieser Achsen eine *möglichst große Varianz* aufweisen.
- Achsen, bezüglich derer die Daten kaum (oder keine) Varianz aufweisen, können später weggelassen werden (→ Dimensionsreduktion).
- Neue Achsen werden durch *orthogonale Transformation* erzeugt (d.h. Vektorlängen und Winkel bleiben erhalten, präziser: das innere Produkt bleibt erhalten).

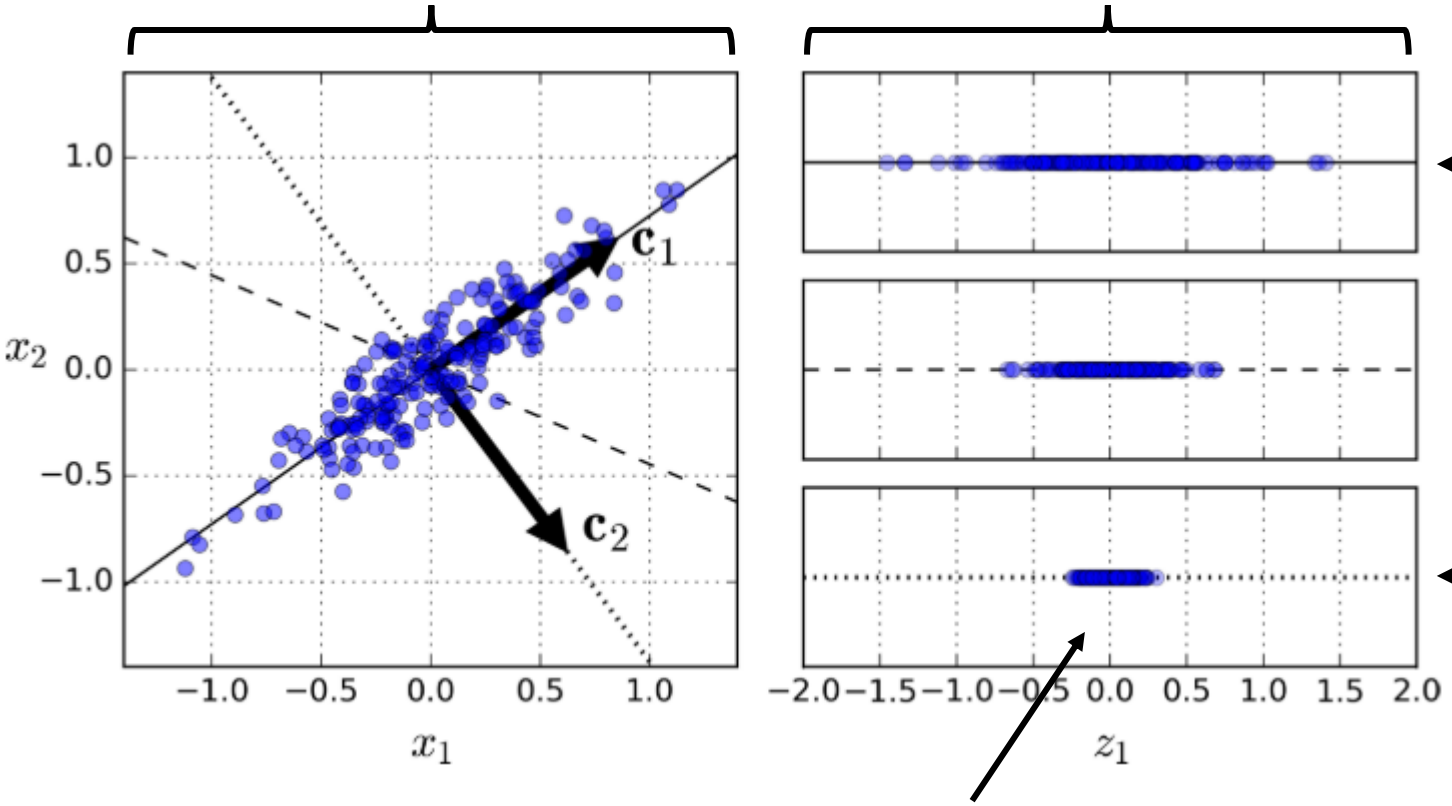
1) Pearson K, Philosophical Magazine 2, 559-572 (1901)

Dimensionsreduktion | PCA

Beispiel:

Daten im 2d-Merkmalsraum

Daten projiziert auf verschiedene Achsen

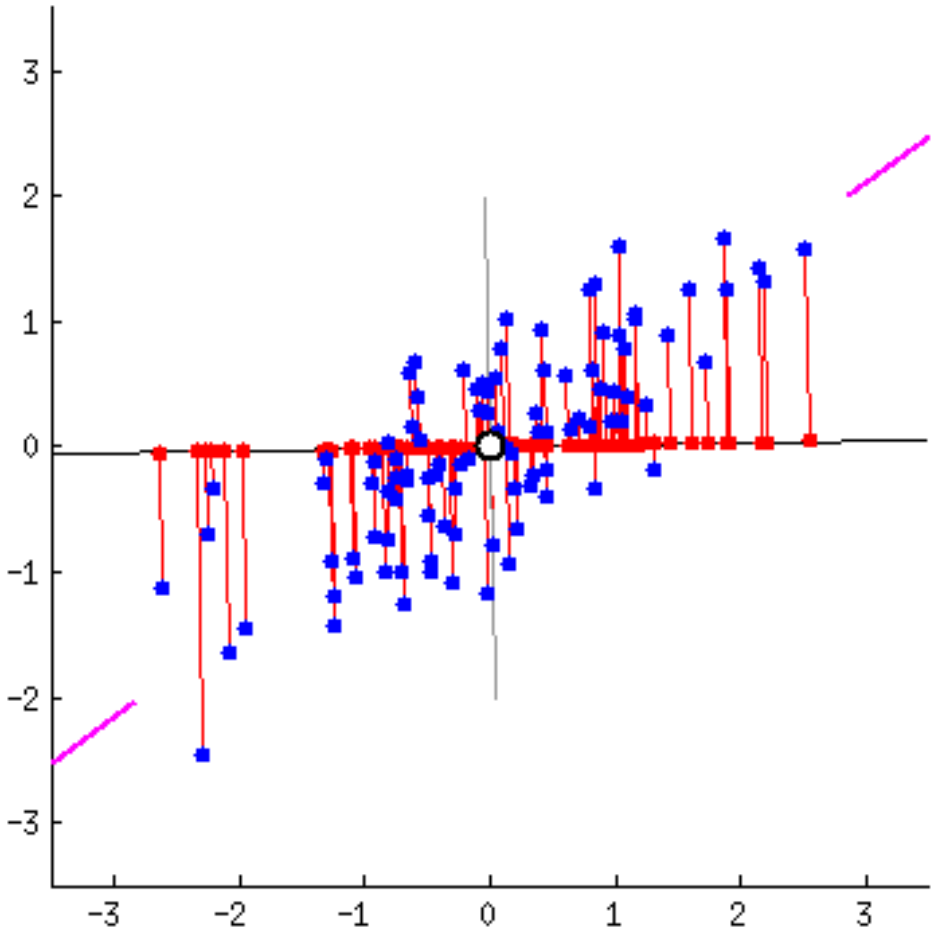


Varianz der Daten ist auf dieser Achse ($\mathbf{c}_1$) am größten: Dies ist die 1. Achse der PCA

PCA ist eine orthogonale Transformation: Achsen stehen senkrecht aufeinander. Dies ist also die 2. Achse ($\mathbf{c}_2$) der PCA.

In diesem Beispiel unterscheiden sich die Datenpunkte auf dieser Achse kaum voneinander (→ geringe Varianz). Wir könnten die Daten also 1-dimensional (mithilfe der 1. Achse) darstellen.

Dimensionsreduktion | PCA



Quelle: https://stats.stackexchange.com/a/140579

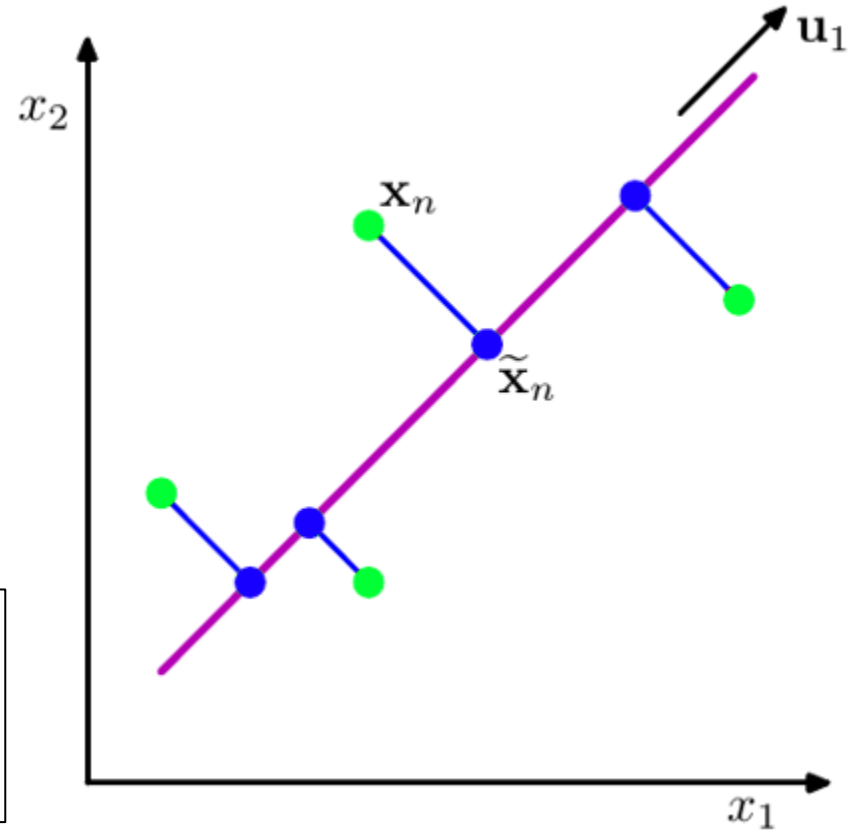
Dimensionsreduktion | PCA

Formale Betrachtung

$\{\mathbf{x}_n\}$: *N* Datenpunkte mit *n*=1,…,*N* mit *D* Merkmalen

Mittlerer Vektor der Daten $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$

Einheitsvektor $\mathbf{u}_1$ (mit: $\mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1 = 1$)

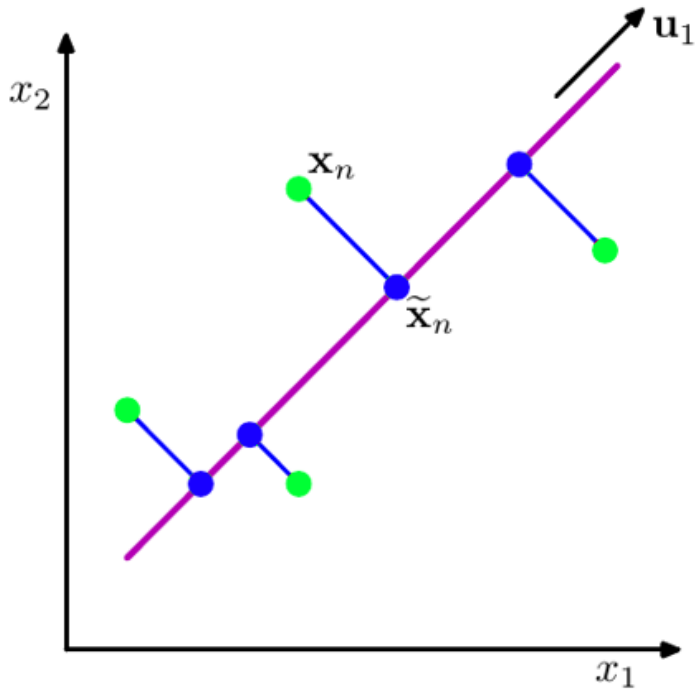


Wir suchen die Richtung $\mathbf{u}_1$, so dass die Varianz der Projektion der Daten auf $\mathbf{u}_1$ maximiert wird.

Projektion eines Datenpunktes $\mathbf{x}_n$ auf $\mathbf{u}_1$: $\mathbf{u}_1^T\mathbf{x}_n$

Dimensionsreduktion | PCA

Wir suchen die Richtung $\mathbf{u_1}$, so dass die Varianz der Projektion der Daten auf $\mathbf{u_1}$ maximiert wird.



Varianz der Projektion der Datenpunkte auf $\mathbf{u_1}$:

$$\mathrm{Var}(\mathbf{u}_1^T\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{u}_1^T\mathbf{x}_n - \mathbf{u}_1^T\bar{\mathbf{x}}\right)^2$$

$$= \mathbf{u}_1^T S\mathbf{u}_1$$

mit der Kovarianzmatrix $S = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$

$\mathrm{Var}(\mathbf{u}_1^T\mathbf{x})$ soll maximiert werden, also:

$$\max_{\mathbf{u}_1}\left(\mathbf{u}_1^T S\mathbf{u}_1\right) \text{ mit Nebenbedingung } \mathbf{u}_1^T\mathbf{u}_1 = 1$$

Dimensionsreduktion | PCA

$\mathrm{Var}(\mathbf{u}_1^T\mathbf{x})$ soll maximiert werden, also:

$$\max_{\mathbf{u}_1} \left(\mathbf{u}_1^T S \mathbf{u}_1\right) \text{ mit Nebenbedingung } \mathbf{u}_1^T\mathbf{u}_1 = 1$$

Dies ist ein Optimierungsproblem mit Nebenbedingung.

Lösungsstrategie: Umwandlung in ein unbeschränktes Optimierungsproblem durch Einführung eines Lagrange-Multiplikators und differenzieren der resultierenden Lagrange-Funktion.

$$\max_{\mathbf{u}_1} \underbrace{\left(\mathbf{u}_1^T S \mathbf{u}_1 + \lambda_1[1 - \mathbf{u}_1^T\mathbf{u}_1]\right)}_{\mathcal{L}}$$

Ableiten, auf null setzen und lösen: $\nabla_{\mathbf{u}_1,\lambda_1}\mathcal{L} \stackrel{!}{=} \mathbf{0}$

Dimensionsreduktion | PCA

$$\nabla_{\mathbf{u}_1,\lambda_1} \mathcal{L} \stackrel{!}{=} \mathbf{0}$$

Wir betrachten zunächst die Ableitung in λ_1 - Richtung:

$$\frac{\partial}{\partial \lambda_1} \left(\mathbf{u}_1^T S \mathbf{u}_1 + \lambda_1 [1 - \mathbf{u}_1^T \mathbf{u}_1] \right) \stackrel{!}{=} 0$$

$$\frac{\partial}{\partial \lambda_1} \lambda_1 [1 - \mathbf{u}_1^T \mathbf{u}_1] \stackrel{!}{=} 0$$

$$[1 - \mathbf{u}_1^T \mathbf{u}_1] \stackrel{!}{=} 0$$

$$\iff \mathbf{u}_1^T \mathbf{u}_1 = 1$$ (die Nebenbedingung ist also erfüllt)

Wir betrachten nun die Ableitungen in $\mathbf{u_1}$ - Richtung:

Dimensionsreduktion | PCA

$$\nabla_{\mathbf{u}_1} \left(\mathbf{u}_1^T S \mathbf{u}_1 + \lambda_1 [1 - \mathbf{u}_1^T \mathbf{u}_1] \right) \stackrel{!}{=} \mathbf{0}$$

$$S\mathbf{u}_1 - \lambda_1 \mathbf{u}_1 = \mathbf{0}$$

$$S\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

$\mathbf{u}_1$ ist also Eigenvektor von *S* mit Eigenwert λ_1.

↓ multiplizieren mit $\mathbf{u}_1^T$ von links

$$\underbrace{\mathbf{u}_1^T S \mathbf{u}_1}_{} = \lambda_1 \qquad \text{weil} \qquad \mathbf{u}_1^T \mathbf{u}_1 = 1$$

Varianz der Projektion
der Datenpunkte auf **$\mathbf{u}_1$**

Zur Erinnerung: Wir wollen die Varianz maximieren.

Die größte Varianz entspricht dem größten Eigenwert.
Wir wählen also **$\mathbf{u}_1$** als den Eigenvektor zum größten Eigenwert λ_1.

Dimensionsreduktion | PCA

- Bis jetzt: Richtung (Vektor $\mathbf{u}_1$) identifiziert, in der die Varianz der Daten maximal wird, die sogenannte Hauptkomponente (*principal component*).
- Jetzt: weitere Richtungen $\mathbf{u}_j$ finden, die senkrecht aufeinander und senkrecht auf $\mathbf{u}_1$ stehen), in denen die Varianz der Daten maximiert wird.

Induktionsbeweis

Induktionsannahme:

Seien $\mathbf{u}_1, \ldots, \mathbf{u}_M$ Eigenvektoren der Kovarianzmatrix *S* zu den *M* größten Eigenwerten $\lambda_1, \ldots, \lambda_M$.

Induktionsschritt:

Wir suchen $\mathbf{u}_{M+1}$ so dass

1. die Varianz in dieser Richtung maximal ist,
2. $\mathbf{u}_{M+1}$ orthogonal zu $\mathbf{u}_1, \ldots, \mathbf{u}_M$ ist
3. $\mathbf{u}_{M+1}$ normiert ist: $\mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1} = 1$

Dimensionsreduktion | PCA

Wir suchen $\mathbf{u}_{M+1}$ so dass

1. die Varianz in dieser Richtung maximal ist,
2. $\mathbf{u}_{M+1}$ orthogonal zu $\mathbf{u}_1, \ldots, \mathbf{u}_M$ ist
3. $\mathbf{u}_{M+1}$ normiert ist: $\mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1} = 1$

} Nebenbedingungen

Dies ist eine Optimierung unter Nebenbedingungen, die wir mithilfe von Lagrange-Multiplikatoren (wie auf den Folien davor) formulieren:

$$\mathcal{L} = \underbrace{\mathbf{u}_{M+1}^{\mathrm{T}} S \mathbf{u}_{M+1}}_{} + \underbrace{\lambda_{M+1}[1 - \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1}]}_{} + \underbrace{\sum_{i=1}^{M} \eta_i \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_i}_{}$$

Varianz in Richtung $\mathbf{u}_{M+1}$
→ Bedingung (1)

$\mathbf{u}_{M+1}$ sei normiert
→ Bedingung (3)

$\mathbf{u}_{M+1}$ sei orthogonal zu $\mathbf{u}_1, \ldots, \mathbf{u}_M$
→ Bedingung (2)

mit den Lagrange Multiplikatoren: $\lambda_{M+1}, \eta_1, \ldots, \eta_M$

Das Optimierungsproblem lautet nun: $\max_{\mathbf{u}_{M+1}} (\mathcal{L})$

Lösen durch Ableiten und setzen auf Null: $\nabla_{\mathbf{u}_{M+1}, \lambda_{M+1}, \eta_i} \mathcal{L} \stackrel{!}{=} 0$

Dimensionsreduktion | PCA

$$\mathcal{L} = \mathbf{u}_{M+1}^{\mathrm{T}} S \mathbf{u}_{M+1} + \lambda_{M+1}[1 - \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1}] + \sum_{i=1}^{M} \eta_i \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_i$$

$$\nabla_{\mathbf{u}_{M+1}, \lambda_{M+1}, \eta_i} \mathcal{L} \stackrel{!}{=} \mathbf{0}$$

Wir betrachten zunächst die Ableitung nach λ_{M+1}:

$$\frac{\partial}{\partial \lambda_{M+1}} \mathcal{L} = 1 - \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1} \stackrel{!}{=} 0 \iff \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1} = 1$$

implementiert also die Nebenbedingung (3) der vorherigen Folie: $\mathbf{u}_{M+1}$ ist normiert.

Nun betrachten wir die Ableitungen nach η_i :

$$\frac{\partial}{\partial \eta_i} \mathcal{L} = \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_i \stackrel{!}{=} 0$$

implementiert also die Nebenbedingung (2) der vorherigen Folie: $\mathbf{u}_{M+1}$ ist orthogonal zu $\mathbf{u}_1$ bis $\mathbf{u}_M$.

Auf der nächsten Folie betrachten wir die Ableitung nach $\mathbf{u}_{M+1}$.

Dimensionsreduktion | PCA

$$\mathcal{L} = \mathbf{u}_{M+1}^{\mathrm{T}} S \mathbf{u}_{M+1} + \lambda_{M+1}[1 - \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_{M+1}] + \sum_{i=1}^{M} \eta_i \mathbf{u}_{M+1}^{\mathrm{T}} \mathbf{u}_i$$

$$\nabla_{\mathbf{u}_{M+1}} \mathcal{L} \stackrel{!}{=} 0 \quad \Longleftrightarrow \quad 2S\mathbf{u}_{M+1} - 2\lambda_{M+1}\mathbf{u}_{M+1} + \sum_{i=1}^{M} \eta_i \mathbf{u}_i \stackrel{!}{=} \mathbf{0} \qquad (*)$$

Bestimmung von $\eta_j, j = 1 \ldots, M$:

$$2\mathbf{u}_j^{\mathrm{T}} S \mathbf{u}_{M+1} - 2\lambda_{M+1} \underbrace{\mathbf{u}_j^{\mathrm{T}} \mathbf{u}_{M+1}}_{\text{= 0 wegen Orthogonalität}} + \sum_{i=1}^{M} \eta_i \underbrace{\mathbf{u}_j^{\mathrm{T}} \mathbf{u}_i}_{= \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{sonst} \end{cases}} = 0$$

$\mathbf{u}_j^{\mathrm{T}}$ an Gleichung (*) von links multiplizieren

$$2\mathbf{u}_j^{\mathrm{T}} S \mathbf{u}_{M+1} + \eta_j = 0$$

$$2\mathbf{u}_{M+1}^{\mathrm{T}} S^{\mathrm{T}} \mathbf{u}_j + \eta_j = 0$$

Transponieren

$$2\mathbf{u}_{M+1}^{\mathrm{T}} S \mathbf{u}_j + \eta_j = 2\mathbf{u}_{M+1}^{\mathrm{T}} \lambda_j \mathbf{u}_j + \eta_j = \eta_j = 0 \quad \text{d.h.} \quad \boxed{\eta_j = 0, j = 1, \ldots, M} \qquad (**)$$

$S^{\mathrm{T}} = S$ weil *S* symmetrisch ist

weil $\boldsymbol{u}_j$ Eigenvektor von *S* ist

Dimensionsreduktion | PCA

Bestimmung von λ_{M+1} :

Aus (*) und (**) ergibt sich: $S\mathbf{u}_{M+1} = \lambda_{M+1}\mathbf{u}_{M+1}$

Das bedeutet: $\mathbf{u}_{M+1}$ ist ebenfalls Eigenvektor von *S*.

Zur Maximierung der Varianz **wählen wir den größten aus den verbliebenen Eigenwerten** λ_i mit dem assoziierten Eigenvektor $\mathbf{u}_i$ aus und nennen beide $(\lambda_{M+1}, \mathbf{u}_{M+1})$.

Damit ist der Induktionsschritt abgeschlossen.

Auf Deutsch: Die Eigenwertzerlegung der Kovarianzmatrix *S* liefert uns Eigenwerte und Eigenvektoren. Wir sortieren nach Größe der Eigenwerte und erhalten die Richtungen der PCA über die zu den Eigenwerten assoziierten Eigenvektoren von *S*.

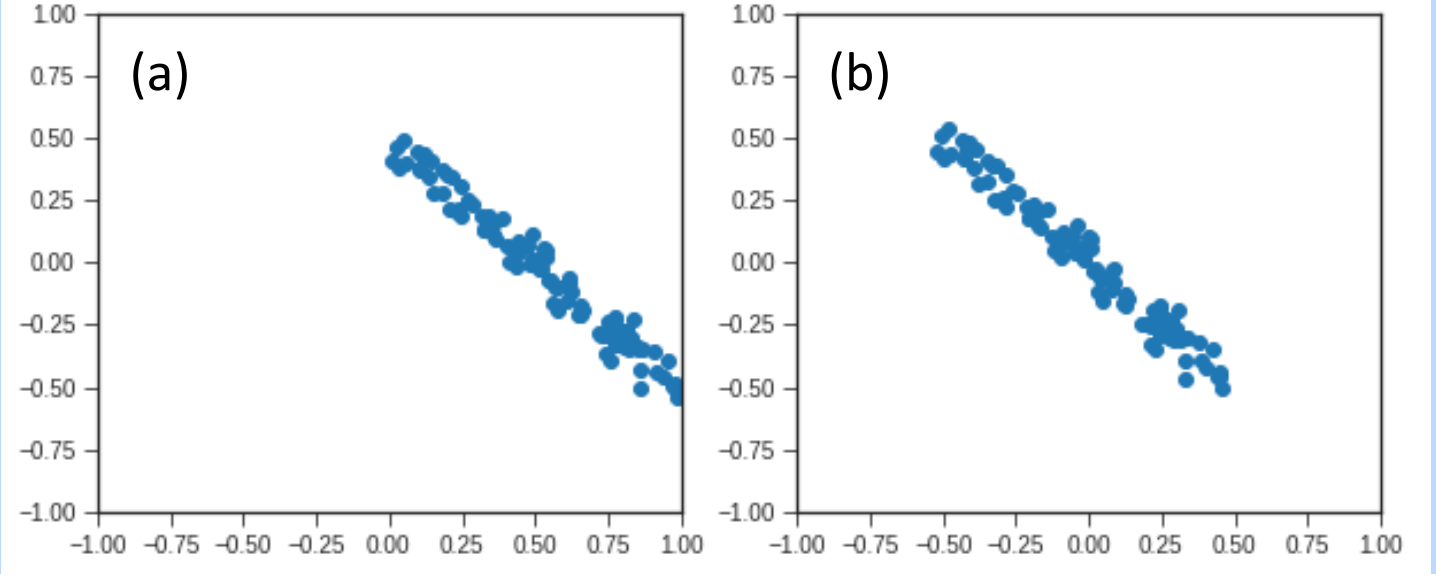
Dimensionsreduktion | PCA

F

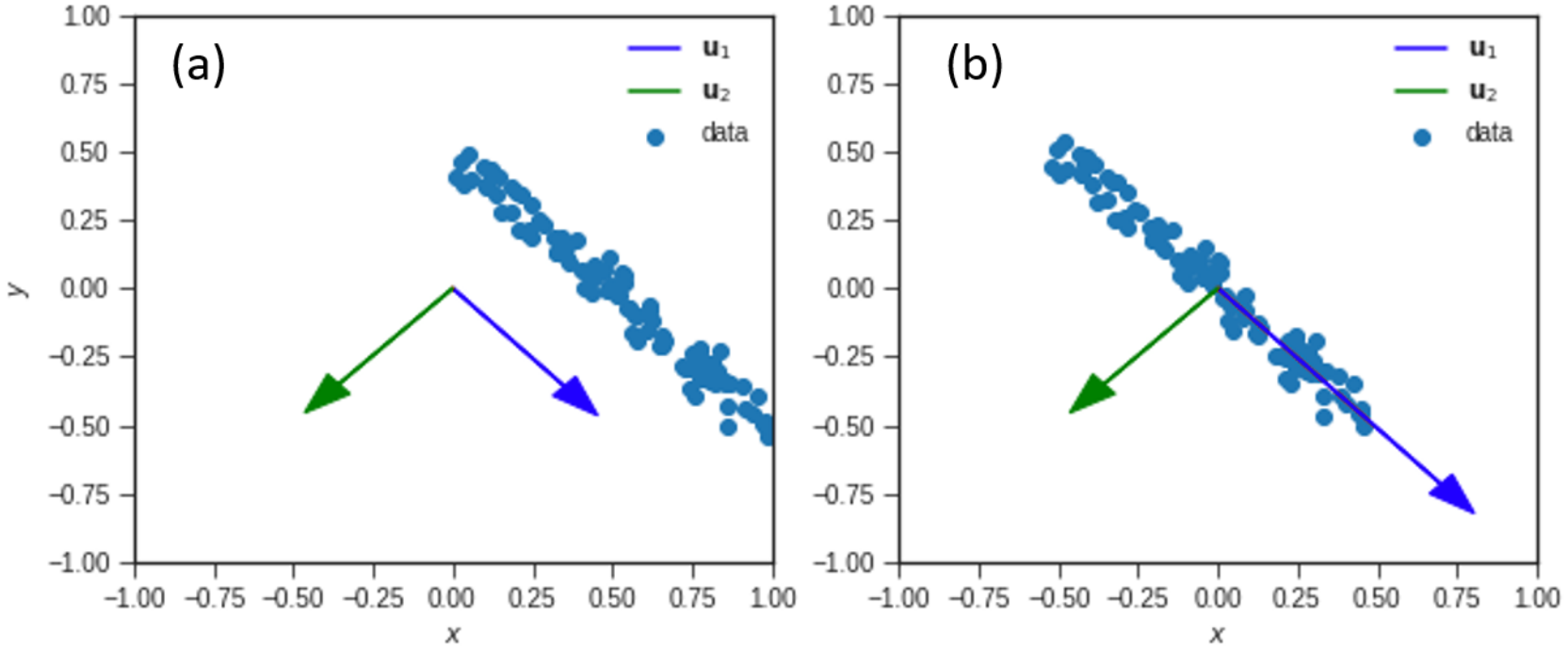
Aktivität

PCA kann als Transformation interpretiert werden, die die Achsen des Koordinatensystems so rotiert, dass die Varianz der Projektion der Daten auf der ersten Achse maximiert wird.

1. Diskutieren Sie mit Ihrem Banknachbarn die beiden unten dargestellten Fälle. Welche(n) Unterschied(e) erkennen Sie?



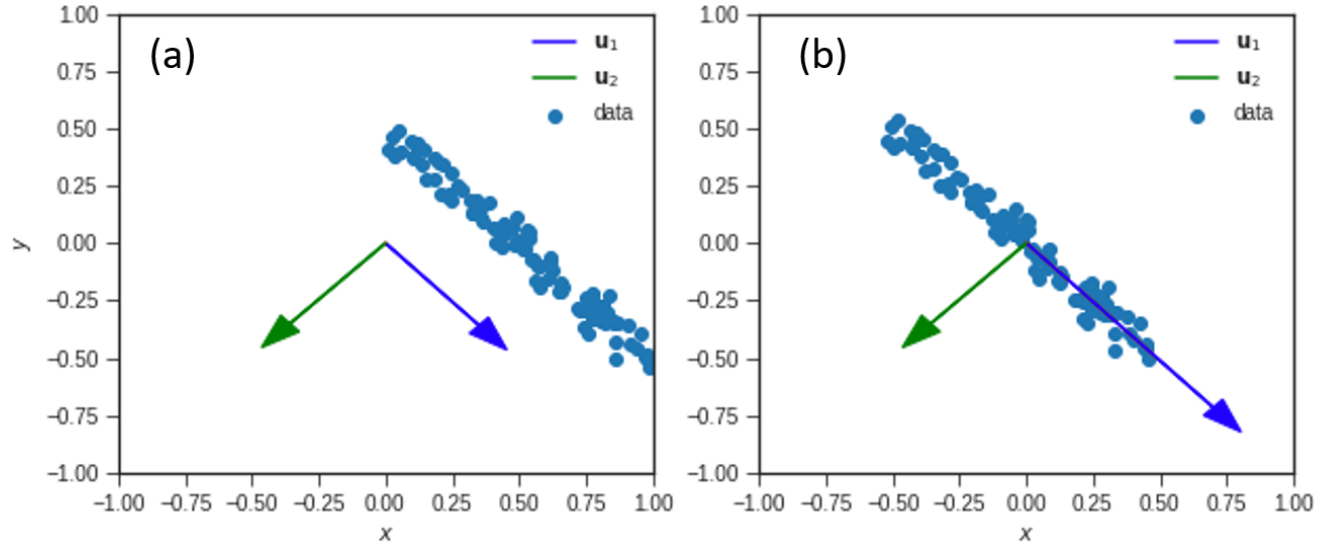
Dimensionsreduktion | PCA

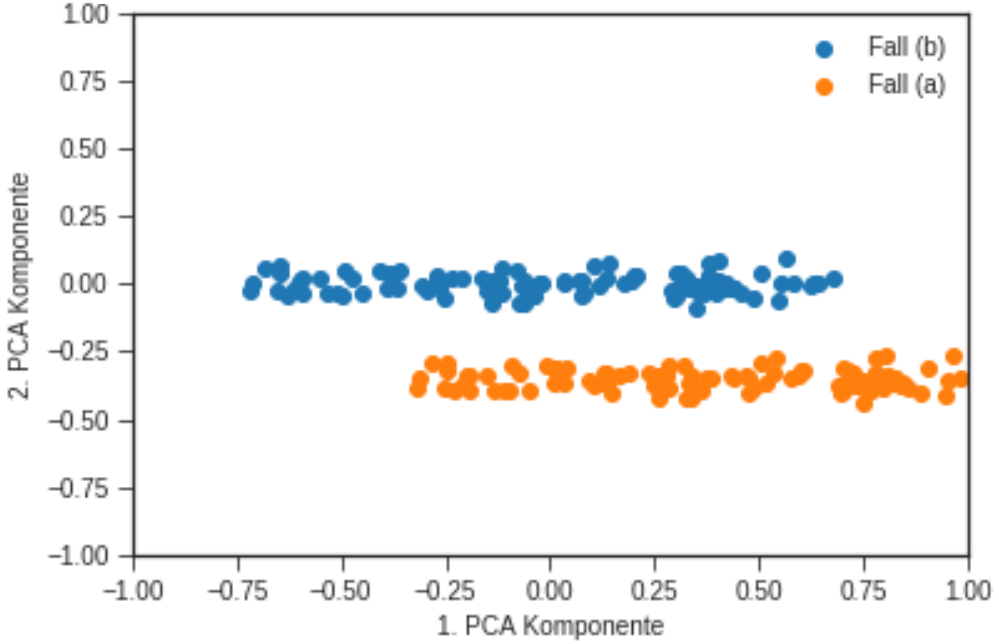


(a) Datenpunkte sind nicht zentriert, d.h. ihr Schwerpunkt (Mittelwert) liegt nicht bei (0, 0) (Koordinatenursprung).

(b) Datenpunkte sind zentriert: Mittelwerte beider Komponenten liegen im Ursprung (0, 0) des Koordinatensystems. Jedes Merkmal wurde vor Anwendung der PCA auf Mittelwert 0 normiert.

Dimensionsreduktion | PCA





Fall (b): Punkte projiziert auf $\mathbf{u}_2$ schwanken nur wenig um 0 herum: Kleine Varianz ist direkt ersichtlich.

Fall (a): Punkte projiziert auf $\mathbf{u}_2$ haben einen (negativen) Offset.
→ unnötig für unsere neuen Koordinaten

Dimensionsreduktion | PCA

Mittelwert der Daten beeinflusst PCA-Koordinaten:

- Daten mit Schwerpunkt (Mittelwert), der nicht dem Koordinatenursprung entspricht, haben nach PCA-Transformation in den neuen Koordinaten typischerweise Offsets
  → meist nicht gut interpretierbar
- PCA-Richtungen ändern sich nicht unter Änderung des Schwerpunkts der Daten. Grund: Kovarianzmatrix ist invariant unter Transformation der Mittelwerte.
- Manche Software-Implementierungen nutzen für die PCA nicht die Kovarianzmatrix. Dort kann die PCA von den Mittelwerten der Daten abhängen.

Empfehlung

- **Zentrieren** Sie jedes Merkmal auf Mittelwert 0 vor Anwendung der PCA:

$$\tilde{x}_i = x_i - \bar{x}$$

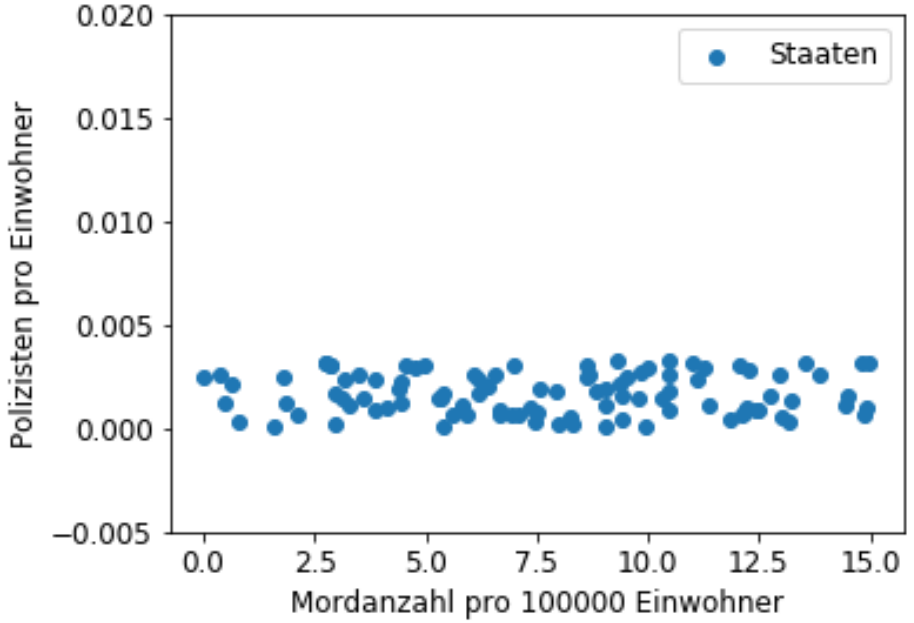
Dimensionsreduktion | PCA

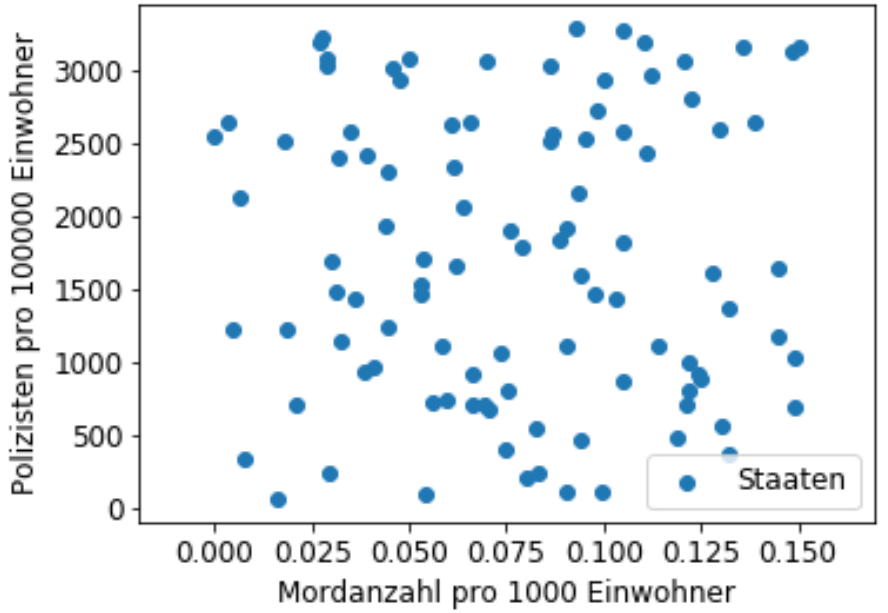
F

Aktivität

Unten sehen Sie dieselben (fiktiven) Daten in zwei Abbildungen dargestellt. Diskutieren Sie mit Ihrem Banknachbarn:

1. Was unterscheidet die beiden Abbildungen genau?
2. Wo liegt die erste Achse der PCA? Wo liegt die 2. Achse der PCA?





Dimensionsreduktion | PCA

Skalierung der Daten beeinflusst PCA:

- Die Skalen, in der ein Merkmal angegeben wird (Beispiel: Mordanzahl pro 100000 Einwohner vs Mordanzahl pro 1000 Einwohner) beeinflusst die Varianz dieses Merkmals.
- Die Ergebnisse der PCA ist abhängig von den Varianzen der Merkmale.

Empfehlung

- **Skalieren** Sie jedes Merkmal auf eine Varianz von 1, damit die Ergebnisse der PCA nicht von der (eventuell) willkürlichen Wahl der Skalen der Merkmale abhängig ist.

Reskalieren des Merkmals X auf Varianz 1: $\tilde{x}_i = \frac{x_i}{\sigma}, \forall i$

Dimensionsreduktion | PCA

Standardisierung (*z-scoring*)

- Transformation eines Merkmals *X*, so dass es Mittelwert 0 und Varianz 1 aufweist. (also: Zentrieren und Skalieren)

Sei $\bar{X}$ der Mittelwert und σ die Standardabweichung von *X*. Das standardisierte Merkmal (*z-score*) *Z* wird erzeugt durch:

$$Z = \frac{X - \bar{X}}{\sigma}$$

In der Praxis sind Daten oft in Matrizen (bzw. DataFrames) organisiert, in denen jede Spalte einem Merkmal entsprechen. In diesem Fall standardisieren Sie jedes Merkmal, indem Sie jede Spalte separat auf Mittelwert 0 und Varianz 1 transformieren.

Dimensionsreduktion mit PCA

- Grundannahme der Dimensionsreduktion mittels PCA:
  Die Richtungen mit der größten Varianz beinhalten die meiste Information.
- Dimensionsreduktion mit PCA heißt: Wir verwerfen die PCA-Komponenten (Richtungen), die nur wenig Varianz repräsentieren.
- PCA liefert uns neue Achsen mit Richtungsvektoren $\mathbf{u}_1, \mathbf{u}_2, \ldots$
- Projektion eines standardisierten Datenpunktes $\mathbf{x}_i$ auf Richtungsvektoren liefert die Komponenten der PCA-Koordinaten dieses Datenpunktes, also:

$$\mathbf{z}_i = \begin{pmatrix} —\mathbf{u}_1^{\mathrm{T}}— \\ —\mathbf{u}_2^{\mathrm{T}}— \\ \vdots \end{pmatrix} \mathbf{x}_i \qquad i = 1, \ldots, N$$

Zur Erinnerung:

N Datenpunkte
D Merkmale

Dimensionsreduktion mit PCA

Welche PCA-Komponenten verwerfen wir?

- Im Wesentlichen: Heuristiken und Ad-Hoc-Entscheidungen
- Nützlich für die Entscheidung: *Proportion of Variance Explained Plot*

Proportion of Variance Explained (PVE)
(= Bruchteil der Gesamtvarianz der Merkmale, die über die PCA-Komponenten repräsentiert wird)

Varianz der *j*-ten Komponente: $\mathrm{Var}(Z_j) = \lambda_j$

Vgl. Sie mit vorherigen Folien: Die Varianz entspricht dem Eigenwert.

Gesamtvarianz über alle Merkmale: $\mathrm{Var}_{\mathrm{total}} = \sum_{j=1}^{D} \mathrm{Var}(Z_j) = \sum_{j=1}^{D} \lambda_j$

Zur Erinnerung:
N Datenpunkte;
D Merkmale

Dimensionsreduktion | PCA

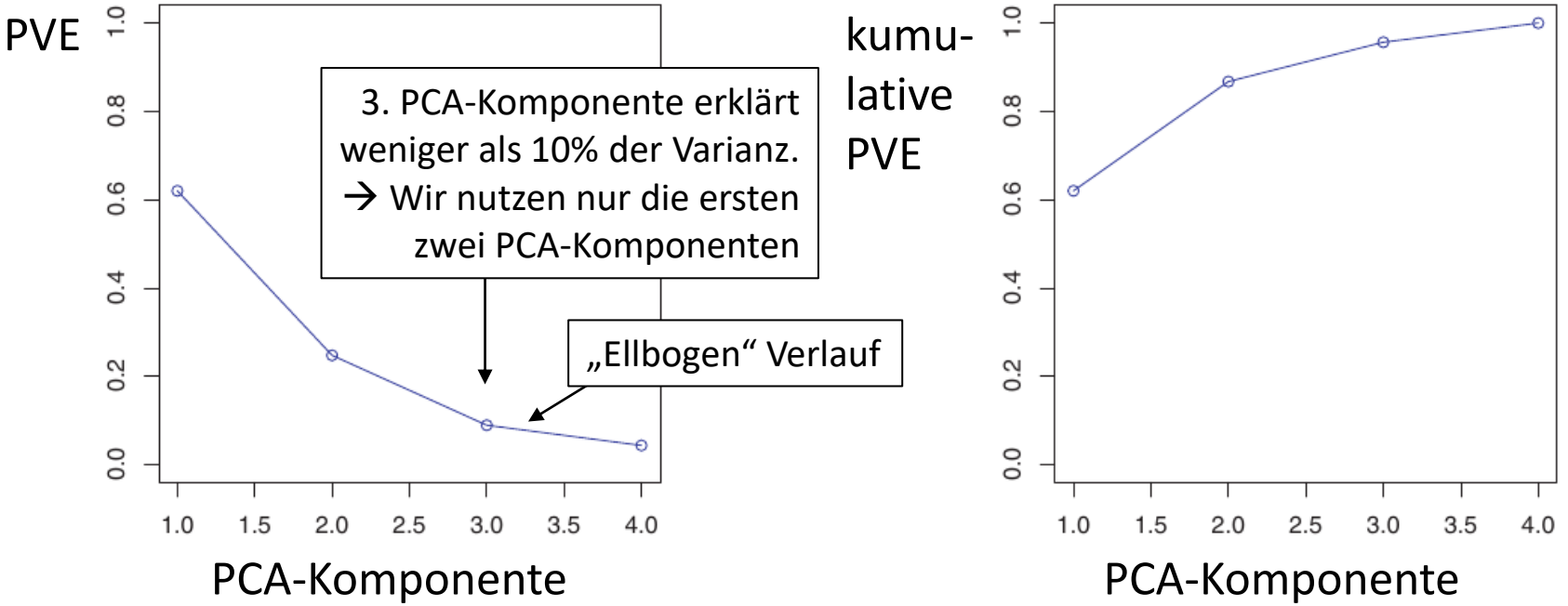
Proportion of Variance Explained (PVE)

(= Bruchteil der Gesamtvarianz der Merkmale, die über die PCA-Komponenten repräsentiert wird)

$$\mathrm{PVE}(j) = \frac{\mathrm{Var}(Z_j)}{\mathrm{Var}_{\mathrm{total}}} = \frac{\lambda_j}{\sum_{i=1}^{D} \lambda_i}$$

j-te PCA-Komponente

Beispiel



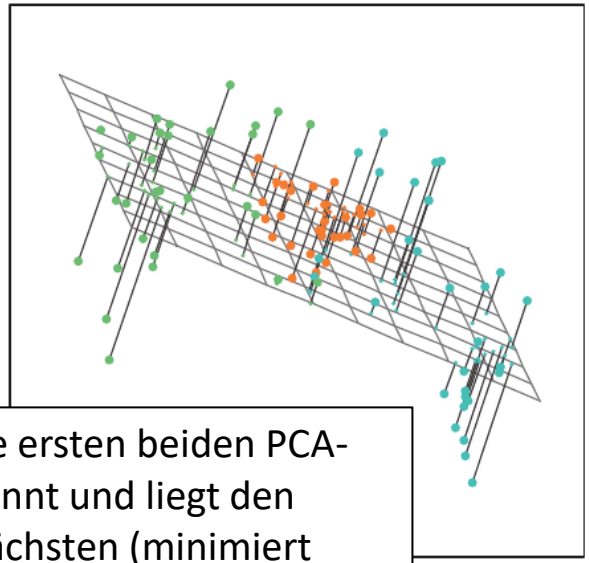
Dimensionsreduktion | PCA

Interpretationen der PCA

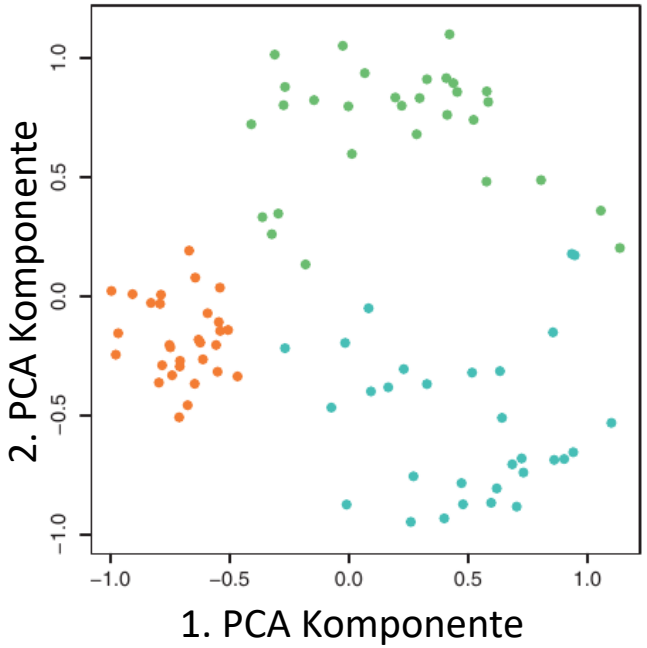
a) Transformation in ein Koordinatensystem, auf dessen Achsen die größte, zweitgrößte, … Varianz in den Daten entfällt

b) Erste *d*-Hauptkomponenten (Richtungen) der PCA spannen einen *d*-dimensionalen Raum auf, der den Datenpunkten am nächsten liegt[1].

Beispiel



Ebene wird durch die ersten beiden PCA-Richtungen aufgespannt und liegt den Datenpunkten am nächsten (minimiert Summe der quadratischen Abweichungen).



1) Im Sinne einer Minimierung quadratischer Abstände der Datenpunkte zum aufgespannten Raum.