Hauptkomponentenanalyse

Annika Scheithe

Dr. Nikolic
Universität zu Köln

28. Juni 2019

Gliederung

- Hauptkomponenten

Gliederung

- Hauptkomponenten
- Spectral Clustering

Motivation

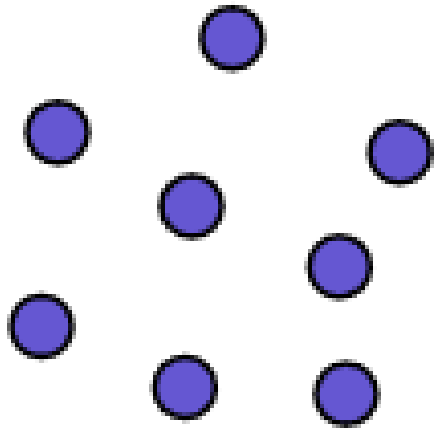


Abbildung: Angenommen wir haben mehrere Menschen mit unterschiedlichen politischen Einstellungen.

Motivation

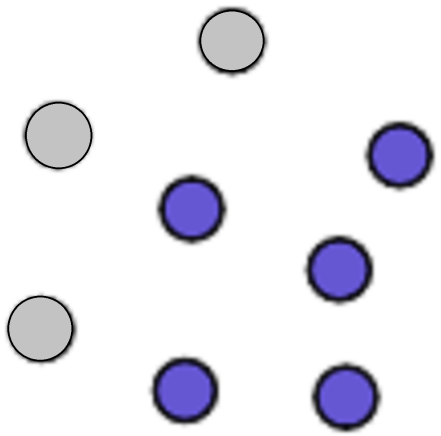


Abbildung: Grau könnte einer Gruppierung von Menschen mit ähnlicher Meinung sein.

Motivation

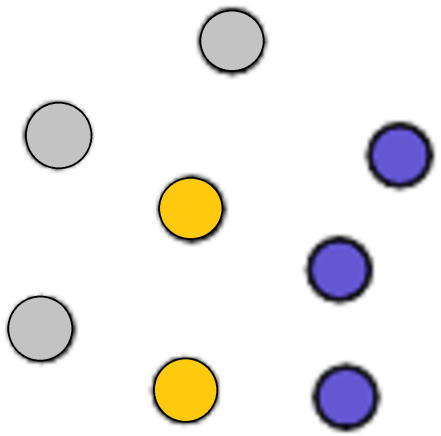


Abbildung: Gelb könnte noch eine andere Gruppierung von Menschen mit ähnlicher Meinung sein.

Motivation

	BGE	Naturschutz	Privatisierung	Artikel 13	...
Person A	10	10	4	0	...
Person B	10	9,5	2	0,5	...
Person C	6	9,5	4	0,5	...
Person D	8	9	2	0,5	...
Person E	4	10	6	0	...
Person F	2	10	8	0,5	...
Person G	6	9	6	0,5	...
Person H	2	10	2	0	...

Abbildung: Datensatz aus den Präferenzen der einzelnen Personen.

Motivation

	BGE	Privatisierung
Person A	10	4
Person B	10	2
Person C	6	4
Person D	8	2
Person E	4	6
Person F	2	8
Person G	6	6
Person H	2	2

Abbildung: Angenommen es existieren nur zwei Themen

Motivation

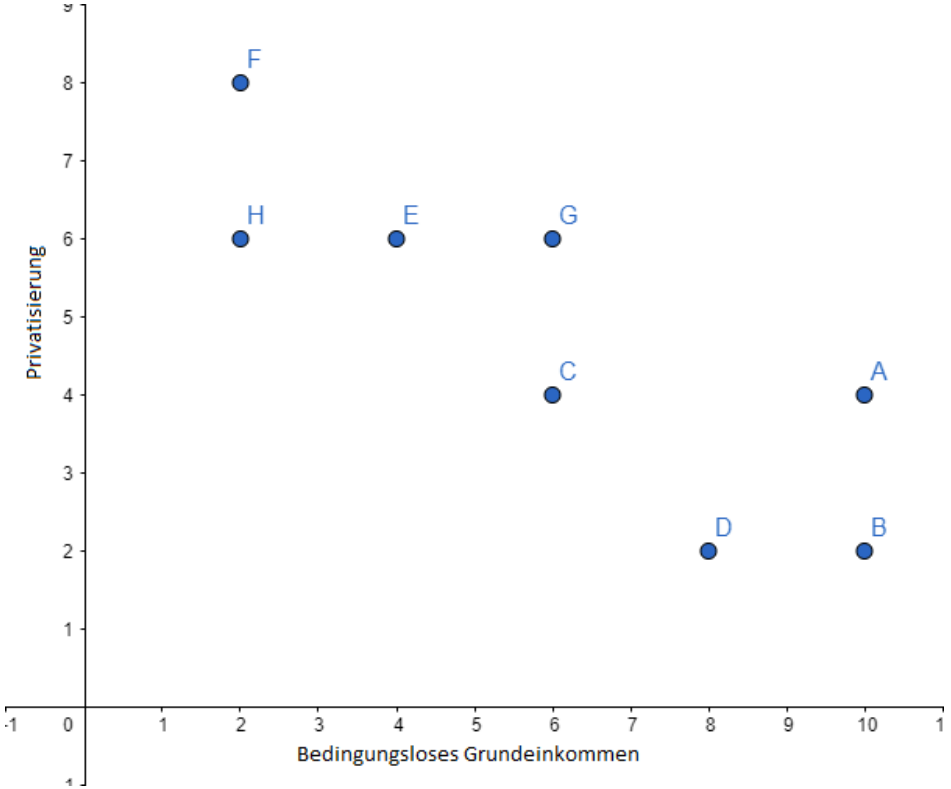


Abbildung: Daten in einem zweidimensionalen Koordinatensystem

Motivation

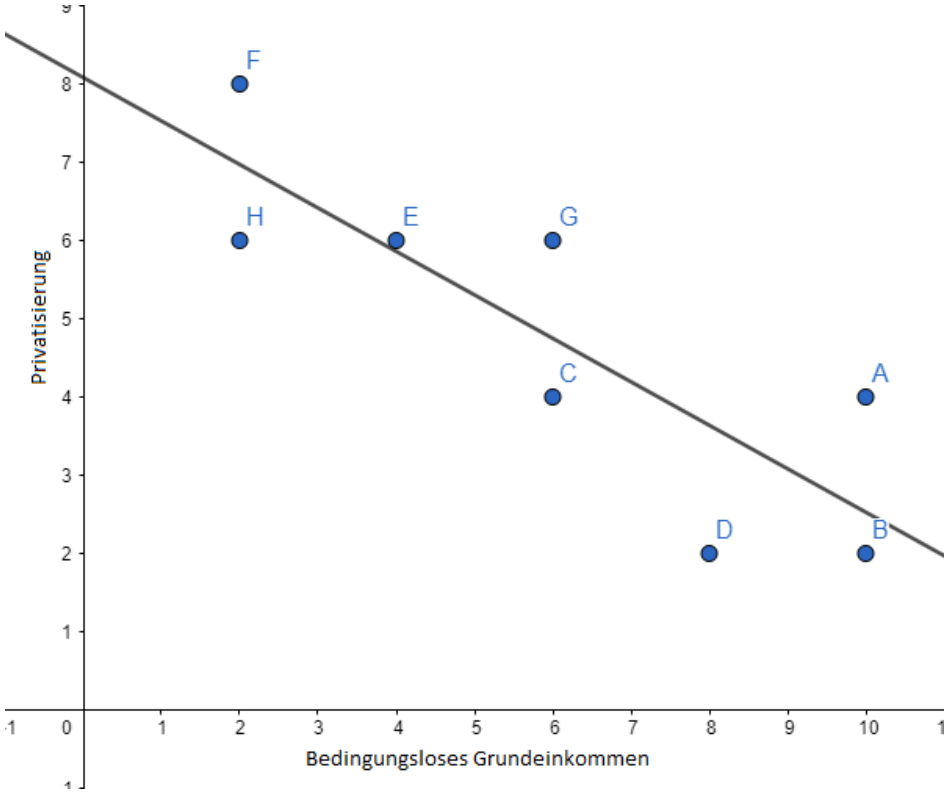


Abbildung: BGE und Privatisierung haben eine negative Korrelation

Motivation

	BGE	Privatisierung	Artikel 13
Person A	10	4	0
Person B	10	2	0,5
Person C	6	4	0,5
Person D	8	2	0,5
Person E	4	6	0
Person F	2	8	0,5
Person G	6	6	0,5
Person H	2	2	0

Abbildung: Angenommen es gibt nun drei Themen

Motivation

- Wir können jetzt zusätzlich bedingungsloses Grundeinkommen mit Artikel 13 vergleichen

Motivation

- Wir können jetzt zusätzlich bedingungsloses Grundeinkommen mit Artikel 13 vergleichen
- Oder Privatisierung mit Artikel 13

Motivation

- Wir können jetzt zusätzlich bedingungsloses Grundeinkommen mit Artikel 13 vergleichen
- Oder Privatisierung mit Artikel 13
- Oder wir können versuchen alle drei Themen in ein dreidimensionales Koordinatensystem zu zeichnen

Motivation

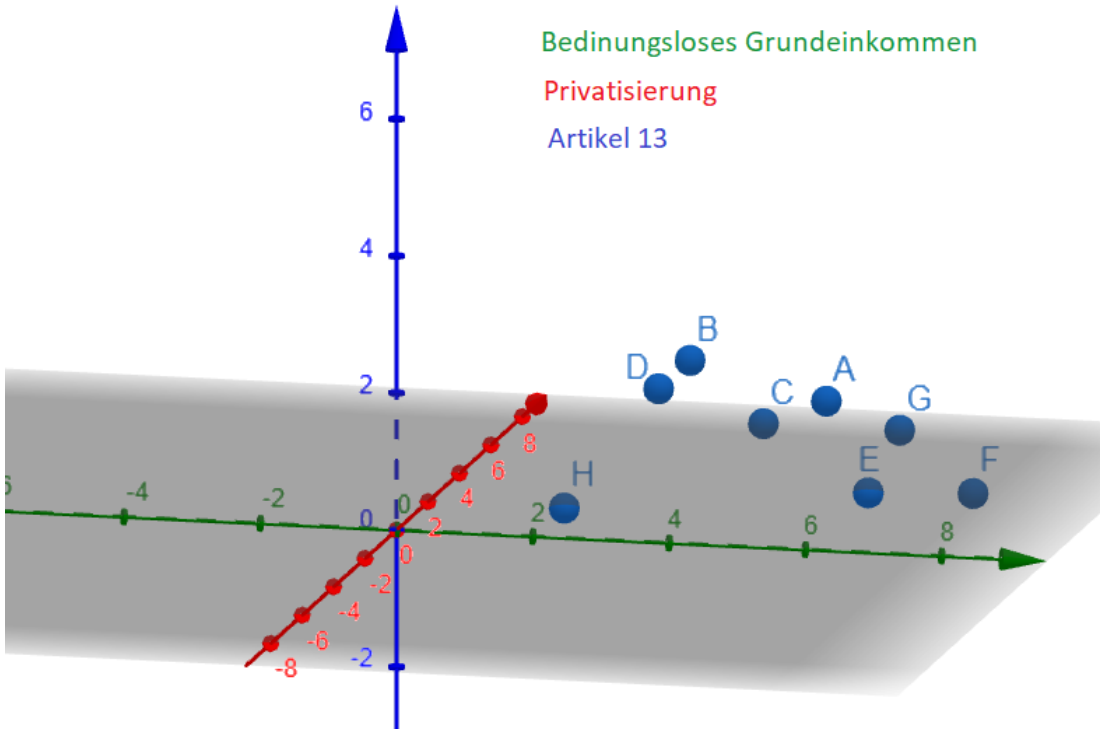


Abbildung: Daten in einem dreidimensionalen Koordinatensystem

Motivation

- Was machen wir, wenn wir nun vier oder mehr Themen einbringen wollen?

Motivation

- Was machen wir, wenn wir nun vier oder mehr Themen einbringen wollen?
- Wir zeichnen eine Hauptkomponentenanalyse!

Motivation

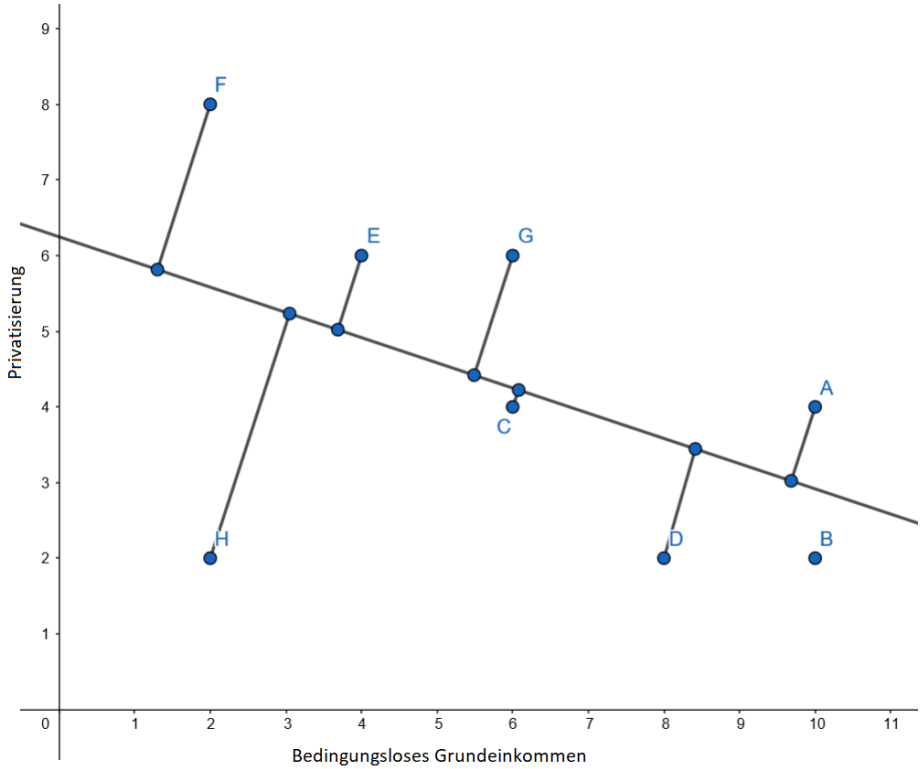


Abbildung: Größte Varianz hat B und F bzw. A und D

Motivation

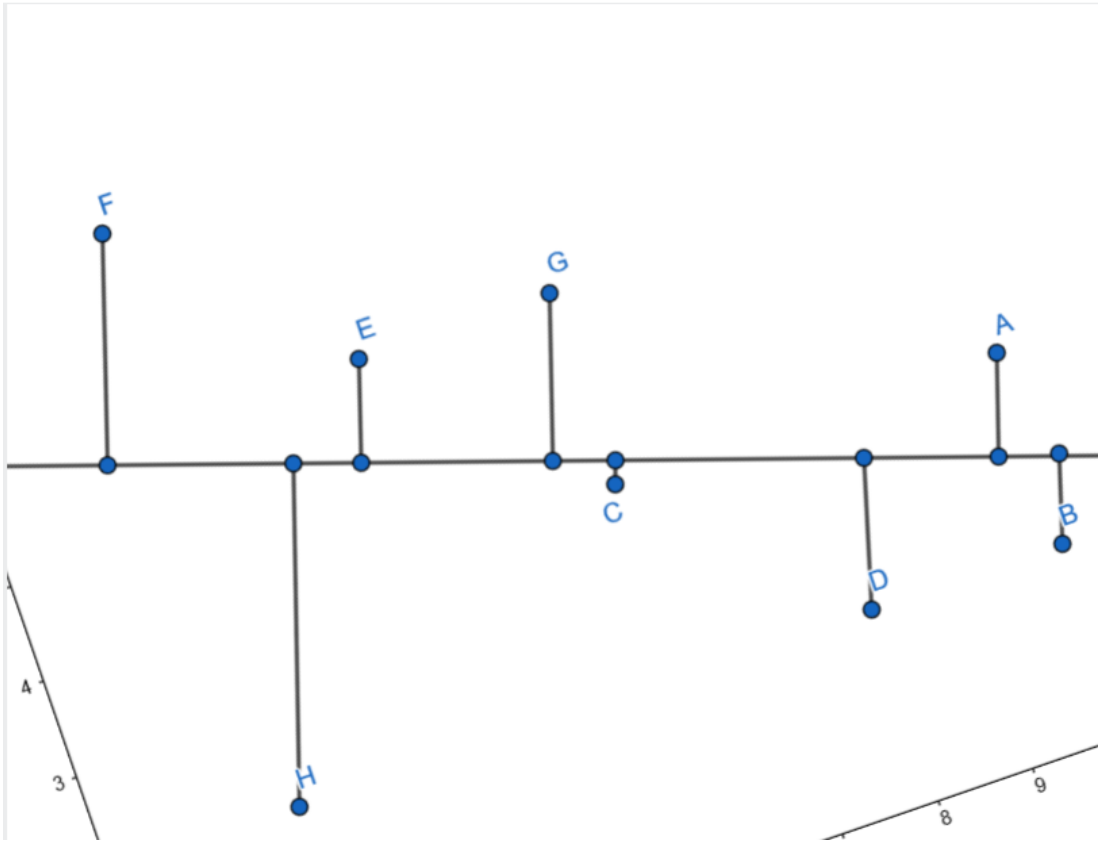


Abbildung: Drehung erzeugt zwei neue Achsen der Varianz

- Hauptkomponente 1 (PC1) ist nun die Achse, die die größte Varianz aufspannt

- Hauptkomponente 1 (PC1) ist nun die Achse, die die größte Varianz aufspannt
- Hauptkomponente 2 (PC2) ist die Achse, die die zweit größte Varianz aufspannt

Idee

- Strukturierung, Vereinfachung und Visualisierung hochdimensionaler Daten

Idee

- Strukturierung, Vereinfachung und Visualisierung hochdimensionaler Daten
- Datenpunkte aus einem p-dimensionalen Raum in einen q-dimensionalen Unterraum ($q \leq p$) zu projizieren

Idee

- Strukturierung, Vereinfachung und Visualisierung hochdimensionaler Daten
- Datenpunkte aus einem p-dimensionalen Raum in einen q-dimensionalen Unterraum ($q \leq p$) zu projizieren $\rightarrow$ dabei soll möglichst wenig Information verloren gehen und vorliegende Redundanz in Form von Korrelation in den Datenpunkten zusammengefasst werden

Modell

- Seien $x_1, x_2, \ldots, x_N$ Beobachtungen

Modell

- Seien $x_1, x_2, \ldots, x_N$ Beobachtungen
- Eine Darstellung erfolgt durch das lineare Modell von Rang q

$$f(\lambda) = \mu + V_q \lambda$$

Modell

- Seien $x_1, x_2, \ldots, x_N$ Beobachtungen
- Eine Darstellung erfolgt durch das lineare Modell von Rang q

$$f(\lambda) = \mu + V_q \lambda$$

- Rekronstruktionsfehler wird mit Hilfe der Methode des kleinsten Quadrates minimiert

$$\min_{\mu, \{\lambda_i\}, V_q} \sum_{i=1}^{N} \|x_i - \mu - V_q \lambda_i\|^2.$$

Modell

- Optimierung erfolgt durch

Modell

- Optimierung erfolgt durch

$$\hat{\mu} = \bar{x},$$

Modell

- Optimierung erfolgt durch

$$\hat{\mu} = \bar{x},$$

$$\hat{\lambda}_i = V_q^T (x_i - \bar{x})$$

Modell

- Optimierung erfolgt durch

$$\hat{\mu} = \bar{x},$$

$$\hat{\lambda}_i = V_q^T (x_i - \bar{x})$$

- zu

$$\min_{V_q} \sum_{i=1}^{N} \|(x_i - \bar{x}) - V_q V_q^T (x_i - \bar{x})\|^2.$$

Modell

- Optimierung erfolgt durch

$$\hat{\mu} = \bar{x},$$

$$\hat{\lambda}_i = V_q^T (x_i - \bar{x})$$

- zu

$$\min_{V_q} \sum_{i=1}^{N} \|(x_i - \bar{x}) - V_q V_q^T (x_i - \bar{x})\|^2.$$

mit $\bar{x} = 0$ folgt

$$\min_{V_q} \sum_{i=1}^{N} \|x_i - V_q V_q^T x_i\|^2.$$

Modell

- $p \times p$-Matrix $H_q = V_q V_q^T$ ist eine Projektionsmatrix

Modell

- $p \times p$-Matrix $H_q = V_q V_q^T$ ist eine Projektionsmatrix $\rightarrow$ bildet jeden Punkt x_i auf seine Rekonstruktion $H_q x_i$ von Rang q ab, die orthogonale Projektion von x_i auf den Halbraum von den Spalten von V_q

Wie erhält man die Hauptkomponenten?

- Die Beobachtungen in eine $N \times p$-Matrix X schreiben

Wie erhält man die Hauptkomponenten?

- Die Beobachtungen in eine $N \times p$-Matrix X schreiben
- Beispiel

$$X = \underbrace{\left.\begin{pmatrix} 10 & 4 & 0 & 10 \\ 10 & 2 & 0.5 & 9.5 \\ 6 & 4 & 0.5 & 9.5 \\ 8 & 2 & 0.5 & 9 \\ 4 & 6 & 0 & 10 \\ 2 & 8 & 0.5 & 10 \\ 6 & 6 & 0.5 & 9 \\ 2 & 6 & 0 & 10 \end{pmatrix}\right\}}_{p=4 Merkmale} \text{N} = 8 \text{ Beobachtungen}$$

Wie erhält man die Hauptkomponenten?

- Zentrierung der Daten durch den Mittelwert

$$X = \begin{pmatrix} 4 & -0.75 & -0.3215 & 0.375 \\ 4 & -2.75 & 0.1785 & -0.125 \\ 0 & 1.25 & 0.1785 & -0.125 \\ 2 & -0.75 & 0.1785 & -0.625 \\ -2 & 3.25 & -0.3215 & 0.375 \\ -4 & 5.25 & 0.1785 & 0.375 \\ 0 & 3.25 & 0.1785 & -0.625 \\ -4 & 3.25 & -0.3215 & 0.375 \end{pmatrix}$$

Wie erhält man die Hauptkomponenten?

- Wir konstruieren die Singulärwertzerlegung von X

Wie erhält man die Hauptkomponenten?

- Wir konstruieren die Singulärwertzerlegung von X

$$X = UDV^T$$

Wie erhält man die Hauptkomponenten?

- Wir konstruieren die Singulärwertzerlegung von X

$$X = UDV^T$$

- U ist eine $N \times p$ orthogonale Matrix ($U^T U = I_p$) mit Spalten u_j, genannt die linken singulären Vektoren

Wie erhält man die Hauptkomponenten?

- Wir konstruieren die Singulärwertzerlegung von X

$$X = UDV^T$$

- U ist eine $N \times p$ orthogonale Matrix ($U^T U = I_p$) mit Spalten u_j, genannt die linken singulären Vektoren
- V ist eine $p \times p$ orthogonale Matrix ($V^T V = I_p$) mit Spalten v_j, genannt die rechten singulären Vektoren

Wie erhält man die Hauptkomponenten?

- Wir konstruieren die Singulärwertzerlegung von X

$$X = UDV^T$$

- U ist eine $N \times p$ orthogonale Matrix ($U^T U = I_p$) mit Spalten u_j, genannt die linken singulären Vektoren
- V ist eine $p \times p$ orthogonale Matrix ($V^T V = I_p$) mit Spalten v_j, genannt die rechten singulären Vektoren
- D ist eine $p \times p$ Diagonalmatrix mit diagonalen Elementen $d_1 \geq d_2, \cdots, d_p \geq 0$ als Singulärwerte

Wie erhält man die Hauptkomponenten?

$$\boxed{X = UDV^T}$$

- Für jeden Rang q besteht die Lösung V_q aus den ersten q Spalten von V

Wie erhält man die Hauptkomponenten?

$$\boxed{X = UDV^T}$$

- Für jeden Rang q besteht die Lösung V_q aus den ersten q Spalten von V
- Die Spalten von UD werden als die **Hauptkomponenten** von X bezeichnet

Wie erhält man die Hauptkomponenten?

$$\boxed{X = UDV^T}$$

- Für jeden Rang q besteht die Lösung V_q aus den ersten q Spalten von V
- Die Spalten von UD werden als die **Hauptkomponenten** von X bezeichnet $\rightarrow$ Die N optimalen $\hat{\lambda}_i$ werden durch die ersten q Hauptkomponenten gegeben

Beispiel

- Singulärwertzerlegung von X

$$U = \begin{pmatrix} -0.299 & -0.577 & -0.63 & -0.098 \\ -0.425 & -0.215 & -0.034 & 0.36 \\ 0.077 & -0.235 & 0.145 & 0.244 \\ -0.176 & -0.233 & 0.427 & -0.236 \\ 0.330 & -0.236 & -0.333 & -0.295 \\ 0.580 & -0.248 & -0.047 & 0.666 \\ 0.199 & -0.615 & 0.488 & -0.270 \\ 0.457 & 0.126 & -0.215 & -0.376 \end{pmatrix}, D = \begin{pmatrix} 11.268 & 0 & 0 & 0 \\ 0 & 3.846 & 0 & 0 \\ 0 & 0 & 1.187 & 0 \\ 0 & 0 & 0 & 0.434 \end{pmatrix}$$

$$, V^T = \begin{pmatrix} -0.715 & 0.698 & -0.01 & 0.038 \\ -0.696 & -0.715 & -0.014 & 0.061 \\ -0.070 & -0.017 & 0.467 & -0.881 \\ 0.018 & 0.005 & 0.884 & 0.467 \end{pmatrix}$$

Beispiel

$$UD = \begin{pmatrix} \overbrace{-3.369}^{\text{PC1}} & -2.219 & -0.748 & -0.043 \\ -4.789 & -0.827 & -0.040 & 0.156 \\ 0.868 & -0.904 & 0.172 & 0.106 \\ -1.983 & -0.896 & 0.507 & -0.102 \\ 3.718 & -0.908 & -0.395 & -0.128 \\ 6.535 & -0.954 & -0.056 & 0.289 \\ 2.242 & -2.365 & 0.579 & -0.117 \\ 5.149 & 0.485 & -0.255 & -0.163 \end{pmatrix}$$

Beispiel

$$UD = \begin{pmatrix} & \overbrace{}^{\text{PC2}} & & \\ -3.369 & -2.219 & -0.748 & -0.043 \\ -4.789 & -0.827 & -0.040 & 0.156 \\ 0.868 & -0.904 & 0.172 & 0.106 \\ -1.983 & -0.896 & 0.507 & -0.102 \\ 3.718 & -0.908 & -0.395 & -0.128 \\ 6.535 & -0.954 & -0.056 & 0.289 \\ 2.242 & -2.365 & 0.579 & -0.117 \\ 5.149 & 0.485 & -0.255 & -0.163 \end{pmatrix}$$

Beispiel

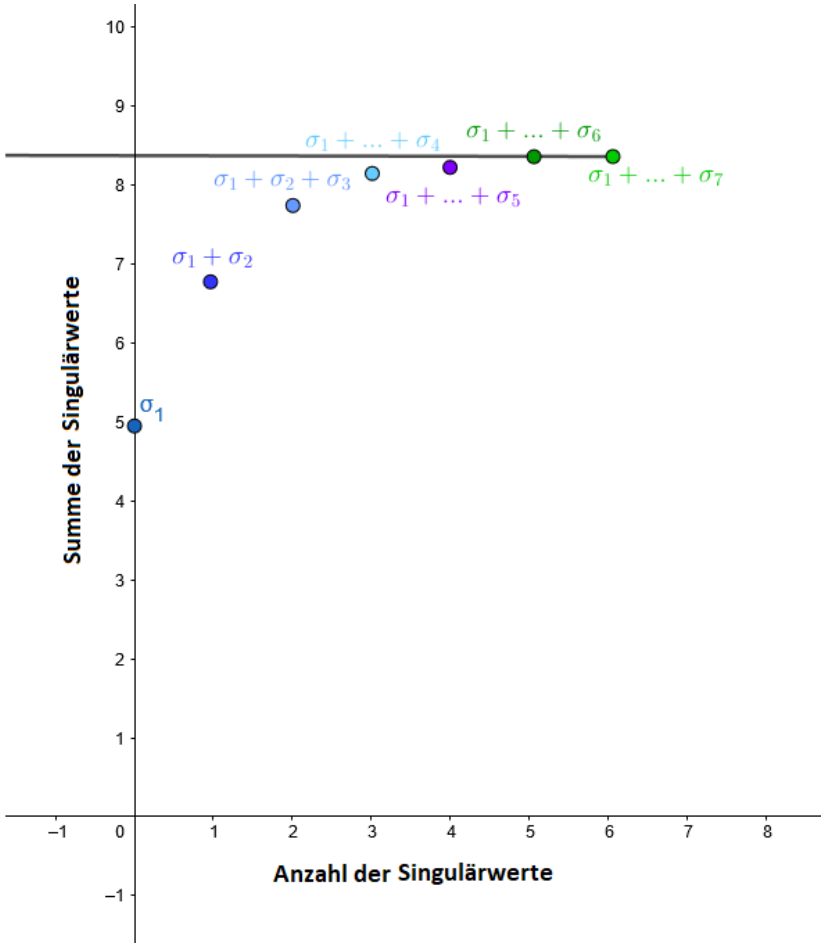
$$UD = \begin{pmatrix} & & \overbrace{}^{\text{PC3}} & \\ -3.369 & -2.219 & -0.748 & -0.043 \\ -4.789 & -0.827 & -0.040 & 0.156 \\ 0.868 & -0.904 & 0.172 & 0.106 \\ -1.983 & -0.896 & 0.507 & -0.102 \\ 3.718 & -0.908 & -0.395 & -0.128 \\ 6.535 & -0.954 & -0.056 & 0.289 \\ 2.242 & -2.365 & 0.579 & -0.117 \\ 5.149 & 0.485 & -0.255 & -0.163 \end{pmatrix}$$

Beispiel

$$UD = \begin{pmatrix} & & & \overbrace{}^{\text{PC4}} \\ -3.369 & -2.219 & -0.748 & -0.043 \\ -4.789 & -0.827 & -0.040 & 0.156 \\ 0.868 & -0.904 & 0.172 & 0.106 \\ -1.983 & -0.896 & 0.507 & -0.102 \\ 3.718 & -0.908 & -0.395 & -0.128 \\ 6.535 & -0.954 & -0.056 & 0.289 \\ 2.242 & -2.365 & 0.579 & -0.117 \\ 5.149 & 0.485 & -0.255 & -0.163 \end{pmatrix}$$

Wie viele Hauptkompenten braucht man?

Wie viele Hauptkompenten braucht man?



Beispiel

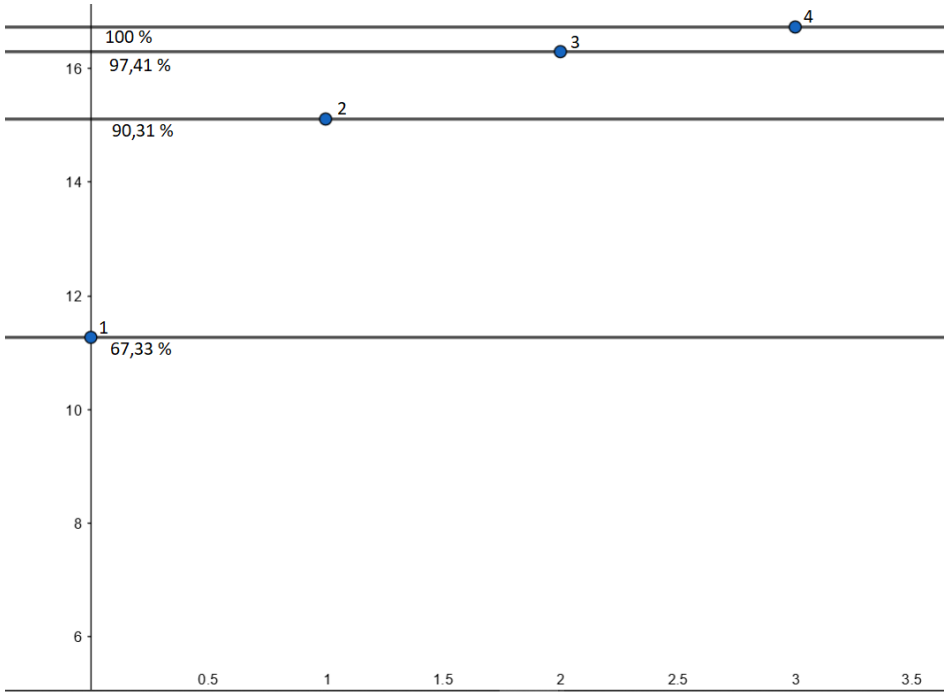


Abbildung: Die ersten beiden Hauptkomponenten überdecken 90 % der Daten

Beispiel

$$U_2 D_2 = \begin{pmatrix} -3.369 & -2.219 \\ -4.789 & -0.827 \\ 0.868 & -0.904 \\ -1.983 & -0.896 \\ 3.718 & -0.908 \\ 6.535 & -0.954 \\ 2.242 & -2.365 \\ 5.149 & 0.485 \end{pmatrix}$$

- Matrix der ersten zwei Hauptkomponenten

Beispiel

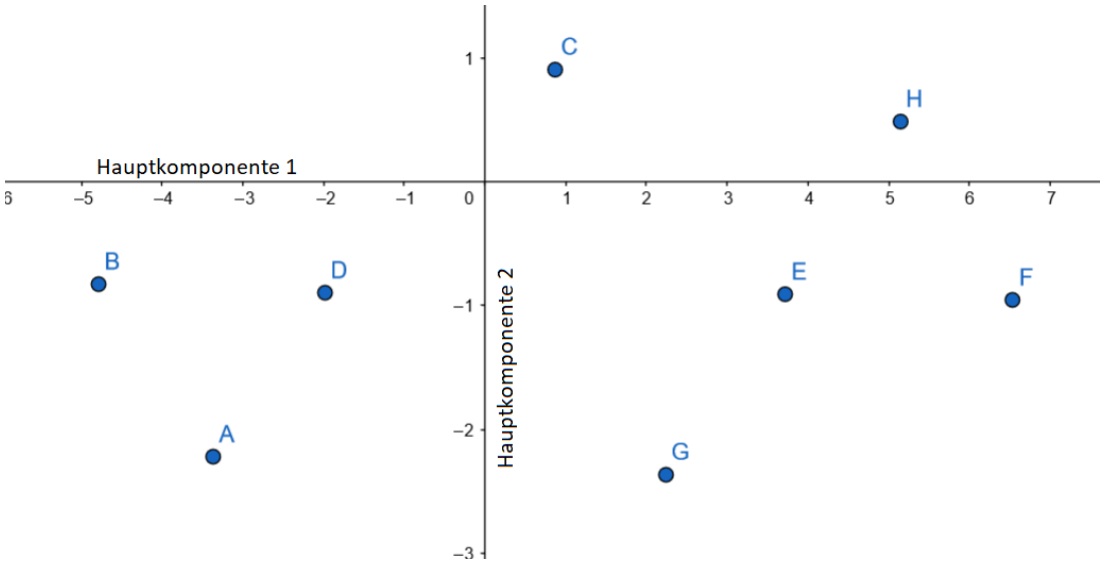


Abbildung: Einzeichnen der Hauptkomponenten in ein Koordinatensystem

Fazit

- Die Hauptkomponentenanalyse ist ein gutes Verfahren, um hochdimensionale Daten zu vereinfachen und zu visualisieren

Fazit

- Die Hauptkomponentenanalyse ist ein gutes Verfahren, um hochdimensionale Daten zu vereinfachen und zu visualisieren
- Die Annahmen der **Linearität** des Datensatzes kann problematisch sein

Fazit

- Die Hauptkomponentenanalyse ist ein gutes Verfahren, um hochdimensionale Daten zu vereinfachen und zu visualisieren
- Die Annahmen der **Linearität** des Datensatzes kann problematisch sein
- Die Zentrierung der Datens ist ein Muss vor der Hauptkomponentenanalyse, da sonst die HKA nicht die optimalen Hauptkomponenten findet

Fazit

- Die Hauptkomponentenanalyse ist ein gutes Verfahren, um hochdimensionale Daten zu vereinfachen und zu visualisieren
- Die Annahmen der **Linearität** des Datensatzes kann problematisch sein
- Die Zentrierung der Datens ist ein Muss vor der Hauptkomponentenanalyse, da sonst die HKA nicht die optimalen Hauptkomponenten findet
- Obwohl die Hauptkomponenten versuchen, die maximale Varianz zwischen den Merkmalen in einem Datensatz abzudecken, kann es sein, dass sie, wenn wir die Anzahl der Hauptkomponenten nicht sorgfältig auswählen, einige Informationen im Vergleich zur ursprünglichen Liste der Merkmale verpassen

Motivation

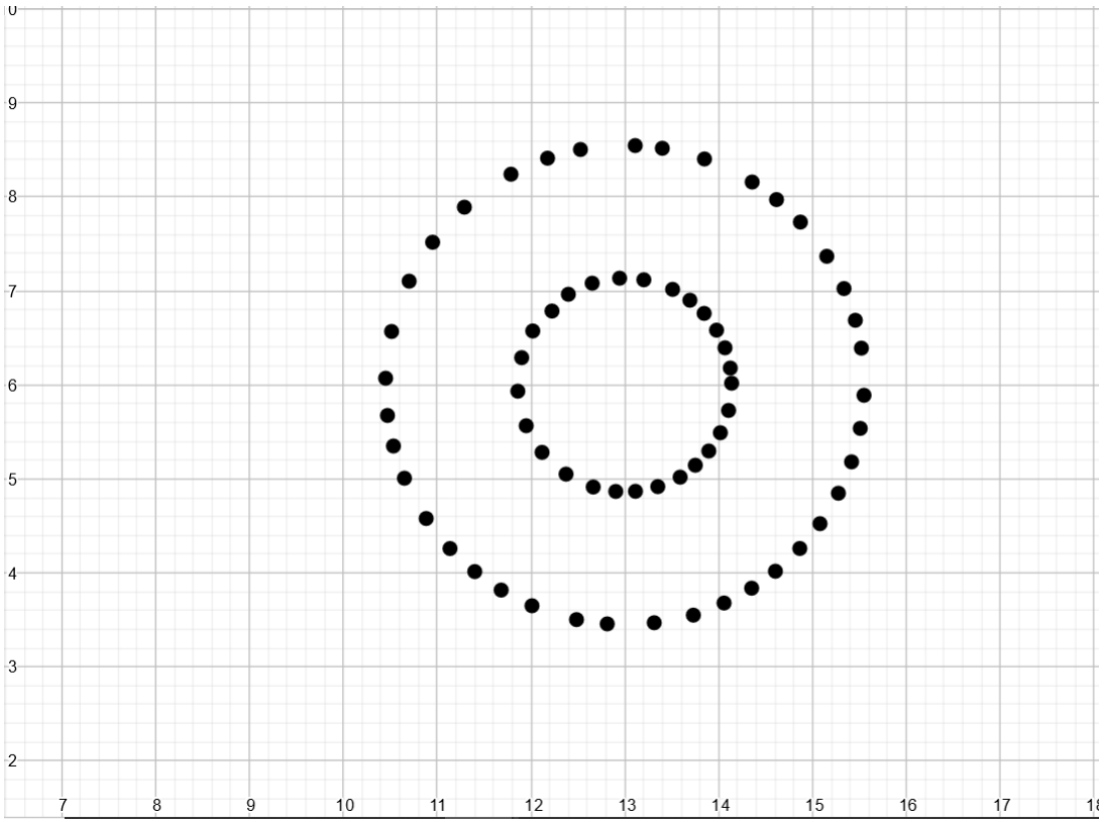


Abbildung: Beispiel eines Datensatzes

Motivation

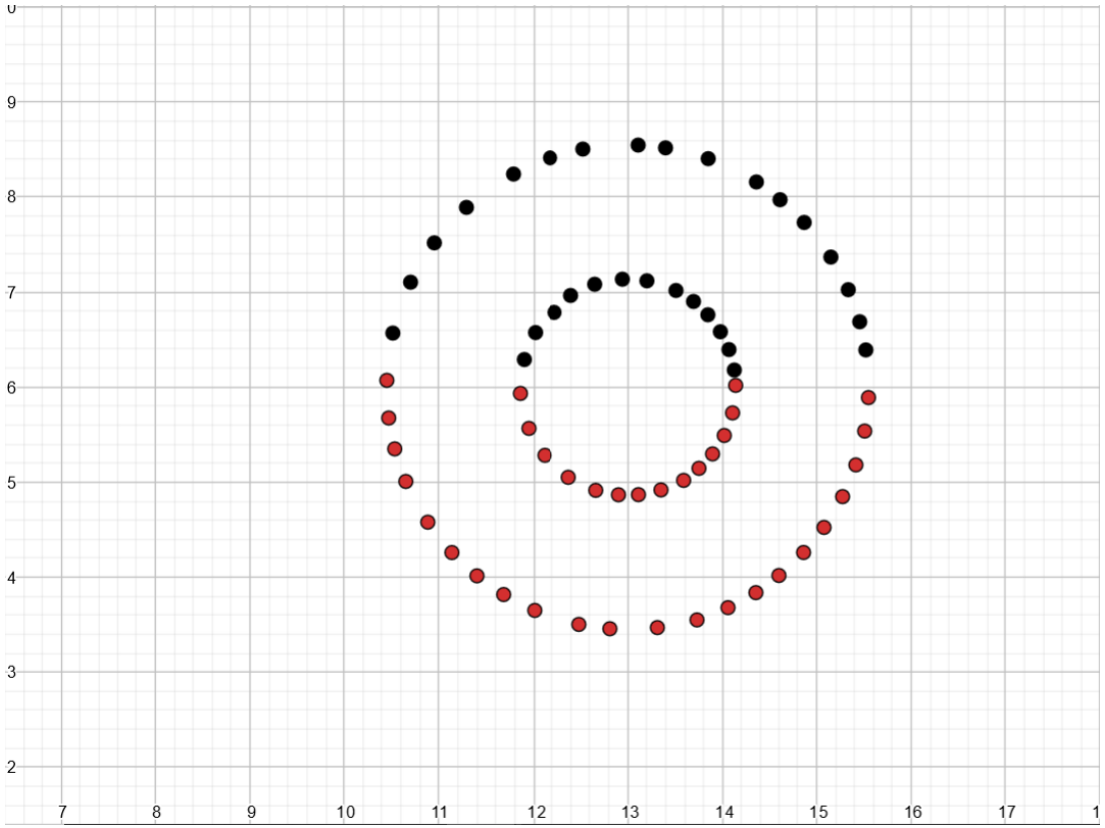


Abbildung: Cluster durch k-Mean Clustering

Motivation

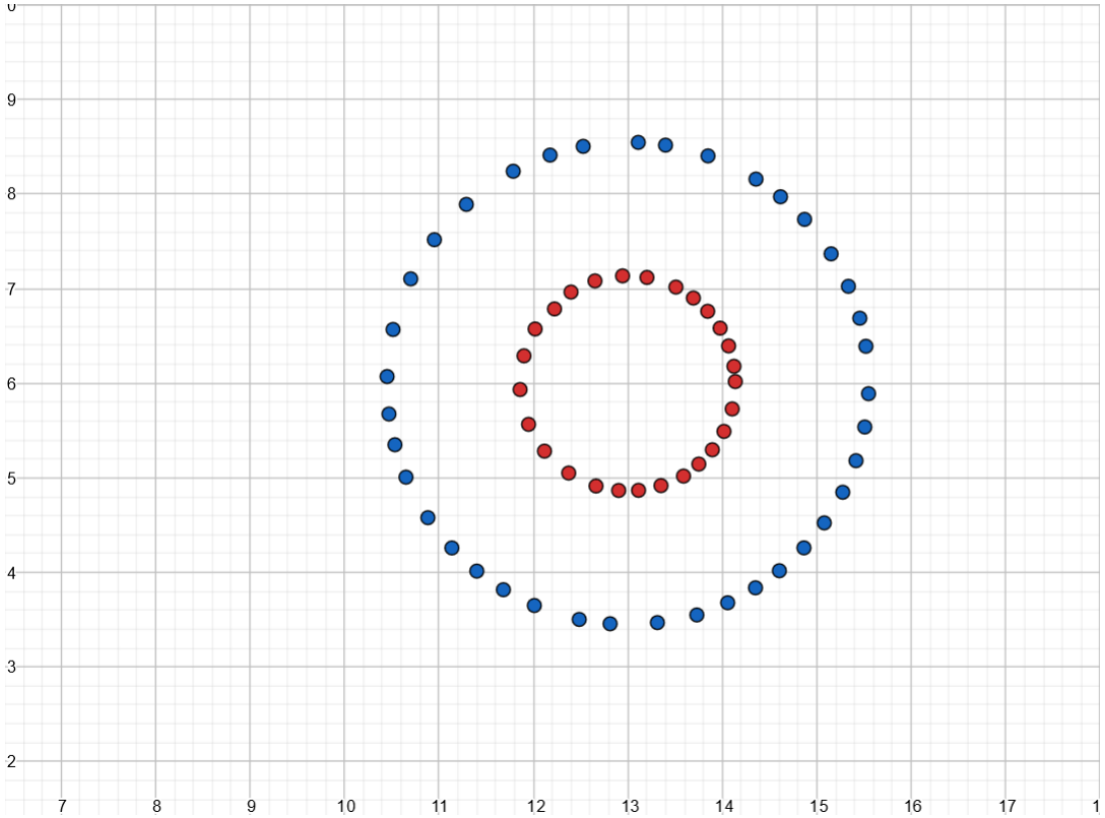


Abbildung: Cluster durch Spectral Clustering

Idee

- Verallgemeinerung von Standard-Clustering-Methoden

Idee

- Verallgemeinerung von Standard-Clustering-Methoden
- Erstellung von Ähnlichkeitsgraphen, die die lokalen Nachbarschaftsbeziehungen zwischen Beobachtungen darstellen

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren
- Beobachtungen werden in einem ungerichteten Ähnlichkeitsgraphen $\langle V, E \rangle$ dargestellt

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren
- Beobachtungen werden in einem ungerichteten Ähnlichkeitsgraphen $\langle V, E \rangle$ dargestellt
- Die N-Knoten v_i repräsentieren die Beobachtungen

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren
- Beobachtungen werden in einem ungerichteten Ähnlichkeitsgraphen $\langle V, E \rangle$ dargestellt
- Die N-Knoten v_i repräsentieren die Beobachtungen
- Paare von Knoten sind durch eine Kante verbunden, wenn ihre Ähnlichkeit positiv ist

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren
- Beobachtungen werden in einem ungerichteten Ähnlichkeitsgraphen $\langle V, E \rangle$ dargestellt
- Die N-Knoten v_i repräsentieren die Beobachtungen
- Paare von Knoten sind durch eine Kante verbunden, wenn ihre Ähnlichkeit positiv ist
- Kanten werden durch die $s_{ii'}$ gewichtet

Idee

- Ausgangspunkt ist eine $N \times N$-Matrix paarweiser Ähnlichkeiten $s_{ii'} \geq 0$ zwischen allen Beobachtungspaaren
- Beobachtungen werden in einem ungerichteten Ähnlichkeitsgraphen $\langle V, E \rangle$ dargestellt
- Die N-Knoten v_i repräsentieren die Beobachtungen
- Paare von Knoten sind durch eine Kante verbunden, wenn ihre Ähnlichkeit positiv ist
- Kanten werden durch die $s_{ii'}$ gewichtet
- Clustering wird zu einem Graph-Partitionierungs Problem

Konkreter

- Betrachte eine Reihe von N Punkten $x_i \in \mathbb{R}^p$

Konkreter

- Betrachte eine Reihe von N Punkten $x_i \in \mathbb{R}^p$
- Sei $d_{ii'}$ der euklidische Abstand zwischen x_i und x_i'

Konkreter

- Betrachte eine Reihe von N Punkten $x_i \in \mathbb{R}^p$
- Sei $d_{ii'}$ der euklidische Abstand zwischen x_i und x_i'
- Die **Ähnlichkeitsmatrix** wird dann aus $s_{ii'} = \exp(-d_{ii'}^2/c)$ mit $c > 0$ erhalten

Beispiel

- Die Abstandsmatrix D mit den Einträgen $(d^2_{ii'})$ ist gegeben durch

$$D = \begin{array}{c} \text{Beob.} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \\ \text{G} \\ \text{H} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} \end{array} \\ \begin{pmatrix} 0 & 4.5 & 16.5 & 9.3 & 40 & 80.3 & 21.3 & 68 \\ 4.5 & 0 & 20 & 4.3 & 52.5 & 100.3 & 32.3 & 80.5 \\ 16.5 & 20 & 0 & 8.3 & 8.5 & 32.3 & 4.3 & 20.3 \\ 9.3 & 4.3 & 8.3 & 0 & 33.3 & 73 & 30 & 53.3 \\ 40 & 52.5 & 8.5 & 33.3 & 0 & 9.3 & 5.3 & 4 \\ 80.3 & 100.3 & 32.3 & 73 & 9.3 & 0 & 21 & 4.3 \\ 21.3 & 32.3 & 4.3 & 30 & 5.4 & 21 & 0 & 12.3 \\ 68 & 80.5 & 20.3 & 53.3 & 4 & 4.3 & 12.3 & 0 \end{pmatrix} \end{array}$$

Beispiel

- Die Ähnlichkeitsmatrix mit $c = 10$ und Schwellenwert 0.09 ist gegeben durch

$$S = \begin{array}{c} \text{Beob.} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \\ \text{G} \\ \text{H} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} \end{array} \\ \begin{pmatrix} 1 & 0.6 & 0.2 & 0.4 & 0 & 0 & 0.12 & 0 \\ 0.6 & 1 & 0.14 & 0.65 & 0 & 0 & 0 & 0 \\ 0.2 & 0.14 & 1 & 0.44 & 0.43 & 0 & 0.65 & 0.13 \\ 0.4 & 0.65 & 0.44 & 1 & 0 & 0 & 0.14 & 0 \\ 0 & 0 & 0.43 & 0 & 1 & 0.4 & 0.6 & 0.67 \\ 0 & 0 & 0 & 0 & 0.4 & 1 & 0.12 & 0.65 \\ 0.12 & 0 & 0.65 & 0.14 & 0.6 & 0.12 & 1 & 0.3 \\ 0 & 0 & 0.13 & 0 & 0.67 & 0.65 & 0.3 & 1 \end{pmatrix} \end{array}$$

Beispiel

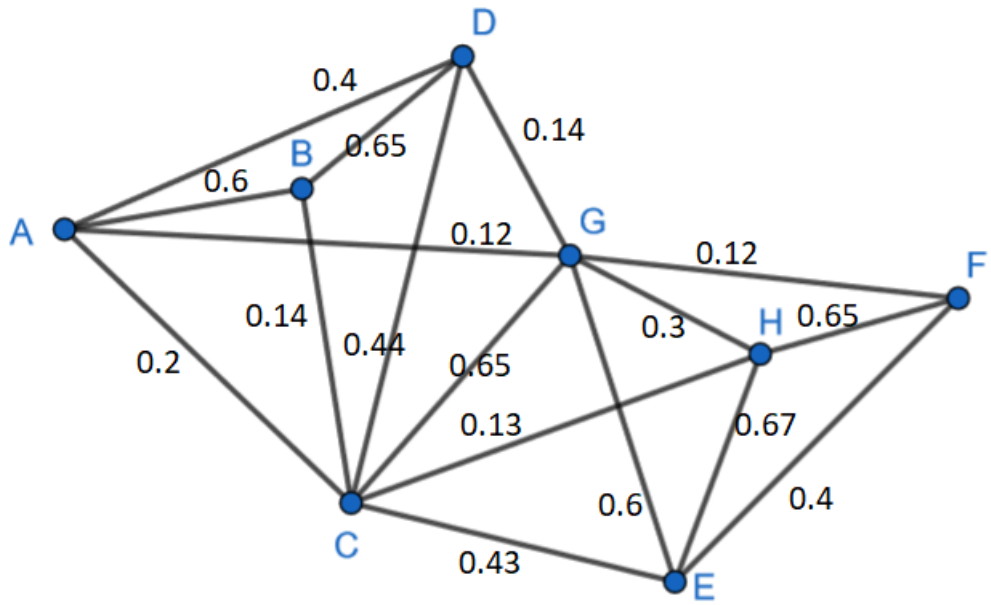


Abbildung: Zeichnen des Graphens durch die Ähnlichkeitsmatrix

k-Nearest Neighbor Graph

- Sei $\mathcal{N}_k$ eine metrische Menge von nahe beieinander liegenden Punktpaaren

k-Nearest Neighbor Graph

- Sei $\mathcal{N}_k$ eine metrische Menge von nahe beieinander liegenden Punktpaaren
- (i, i') befindet sich in $\mathcal{N}_k$, wenn Punkt i' zu den nächsten k-Nachbarn von i gehört oder umgekehrt

k-Nearest Neighbor Graph

- Sei $\mathcal{N}_k$ eine metrische Menge von nahe beieinander liegenden Punktpaaren
- (i, i') befindet sich in $\mathcal{N}_k$, wenn Punkt i' zu den nächsten k-Nachbarn von i gehört oder umgekehrt
- Verbindung aller symetrischen nächsten Nachbarn mit Kantengewichten $w_{ii'} = s_{ii'}$ ansonsten ist das Kantengewicht Null

k-Nearest Neighbor Graph

- Sei $\mathcal{N}_k$ eine metrische Menge von nahe beieinander liegenden Punktpaaren
- (i, i') befindet sich in $\mathcal{N}_k$, wenn Punkt i' zu den nächsten k-Nachbarn von i gehört oder umgekehrt
- Verbindung aller symetrischen nächsten Nachbarn mit Kantengewichten $w_{ii'} = s_{ii'}$ ansonsten ist das Kantengewicht Null
- Äquivalent setzen wir auf Null alle paarweisen Ähnlichkeiten, die nicht in $\mathcal{N}_k$ sind

Laplacian Graph

- Die Matrix der Kantengewichte $W = \{w_{ii'}\}$ wird als **Adjazenzmatrix** bezeichnet

Laplacian Graph

- Die Matrix der Kantengewichte $W = \{w_{ii'}\}$ wird als **Adjazenzmatrix** bezeichnet
- **Grad des Kontens** i ist $g_i = \sum_{i'} w_{ii'}$

Laplacian Graph

- Die Matrix der Kantengewichte $W = \{w_{ii'}\}$ wird als **Adjazenzmatrix** bezeichnet
- **Grad des Kontens** i ist $g_i = \sum_{i'} w_{ii'}$
- Sei G eine Diagonalmatrix mit Diagonaleinträgen g_i

Laplacian Graph

- Die Matrix der Kantengewichte $W = \{w_{ii'}\}$ wird als **Adjazenzmatrix** bezeichnet
- **Grad des Kontens** i ist $g_i = \sum_{i'} w_{ii'}$
- Sei G eine Diagonalmatrix mit Diagonaleinträgen g_i
- Der unnormalisierte Laplacian Graph wird nun definiert als

Laplacian Graph

- Die Matrix der Kantengewichte $W = \{w_{ii'}\}$ wird als **Adjazenzmatrix** bezeichnet
- **Grad des Kontens** i ist $g_i = \sum_{i'} w_{ii'}$
- Sei G eine Diagonalmatrix mit Diagonaleinträgen g_i
- Der unnormalisierte Laplacian Graph wird nun definiert als

$$L = G - W$$

Beispiel

$$W = \begin{array}{c} \text{Beob.} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \\ \text{G} \\ \text{H} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} \end{array} \\ \begin{pmatrix} 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0.12 & 0 \\ 0.6 & 0 & 0.14 & 0.65 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.44 & 0.43 & 0 & 0 & 0.13 \\ 0.4 & 0.65 & 0.44 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0.67 \\ 0 & 0 & 0 & 0 & 0.4 & 0 & 0.12 & 0.65 \\ 0 & 0 & 0 & 0.14 & 0.6 & 0.12 & 0 & 0 \\ 0 & 0 & 0.13 & 0 & 0.67 & 0.65 & 0 & 0 \end{pmatrix} \end{array}$$

Beispiel

$$G = \begin{pmatrix} 1.12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.39 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.49 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.67 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.17 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.86 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.45 \end{pmatrix}$$

Beispiel

$$L = \begin{pmatrix} 1.12 & -0.6 & 0 & -0.4 & 0 & 0 & -0.12 & 0 \\ -0.6 & 1.39 & -0.14 & -0.65 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -0.44 & -0.43 & 0 & 0 & -0.13 \\ -0.4 & -0.65 & -0.44 & 1.49 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.67 & -0.4 & -0.6 & -0.67 \\ 0 & 0 & 0 & 0 & -0.4 & 1.17 & -0.12 & -0.65 \\ 0 & 0 & 0 & -0.14 & -0.6 & -0.12 & 0.86 & 0 \\ 0 & 0 & -0.13 & 0 & -0.67 & -0.65 & 0 & 1.45 \end{pmatrix}$$

Laplacian Graph

- Die Eigenvektoren zu den Eigenwerten der Matrix L liefern die Basis Z eines neuen Raumes, in dem der Datensatz projiziert wird

Laplacian Graph

- Die Eigenvektoren zu den Eigenwerten der Matrix L liefern die Basis Z eines neuen Raumes, in dem der Datensatz projiziert wird
- Die Eigenwerte geben die „Wichtigkeit“ der Eigenvektoren an

Laplacian Graph

- Die Eigenvektoren zu den Eigenwerten der Matrix L liefern die Basis Z eines neuen Raumes, in dem der Datensatz projiziert wird
- Die Eigenwerte geben die „Wichtigkeit“ der Eigenvektoren an
- Möchte man den Datensatz auf einen k-dimensionalen Raum verkleinern, wählt man die k Eigenvektoren mit den kleinsten Eigenwerten

Laplacian Graph

- Die Eigenvektoren zu den Eigenwerten der Matrix L liefern die Basis Z eines neuen Raumes, in dem der Datensatz projiziert wird
- Die Eigenwerte geben die „Wichtigkeit“ der Eigenvektoren an
- Möchte man den Datensatz auf einen k-dimensionalen Raum verkleinern, wählt man die k Eigenvektoren mit den kleinsten Eigenwerten
- Nun wählt man eine Standard-Clustering-Methode, wie k-Mean Clustering, um die Spalten von Z zu clustern und ein Clustering des ursprünglichen Datensatzes zu erhalten

Konkreter

Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

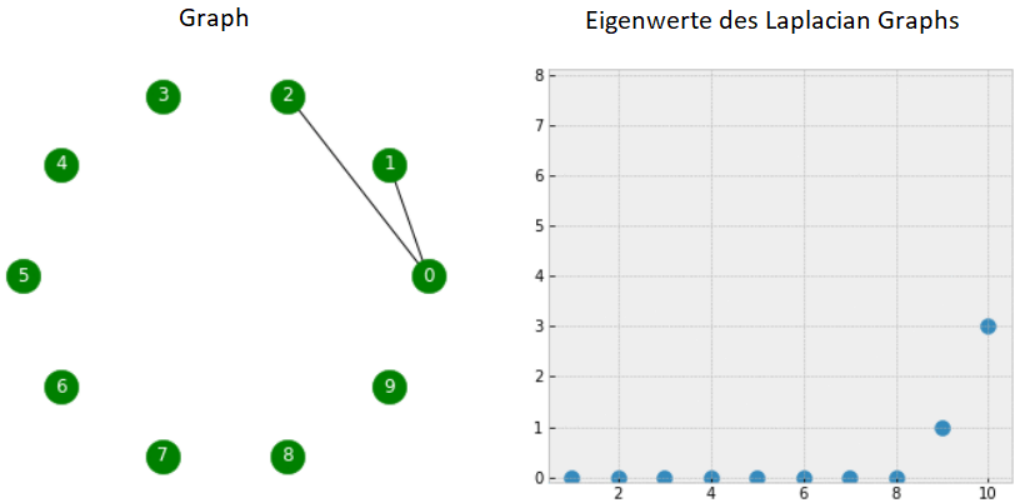


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

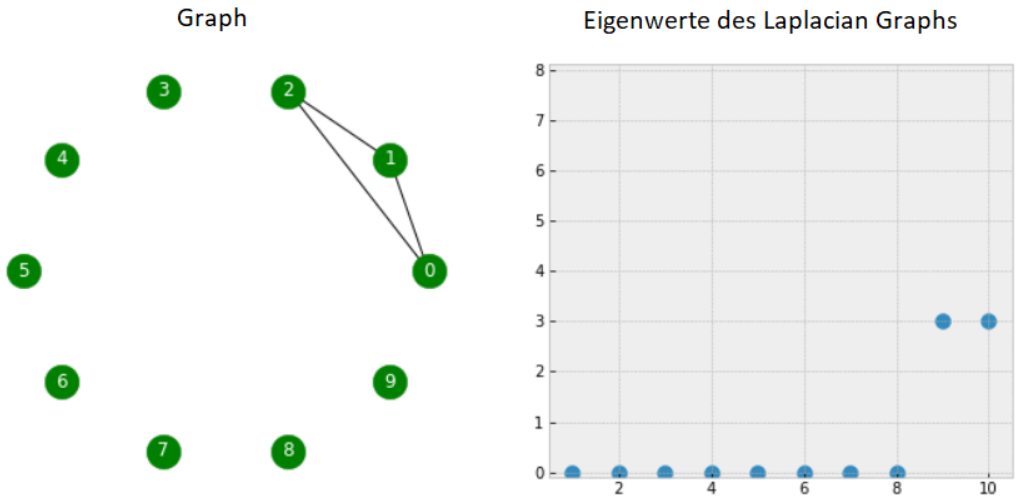


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

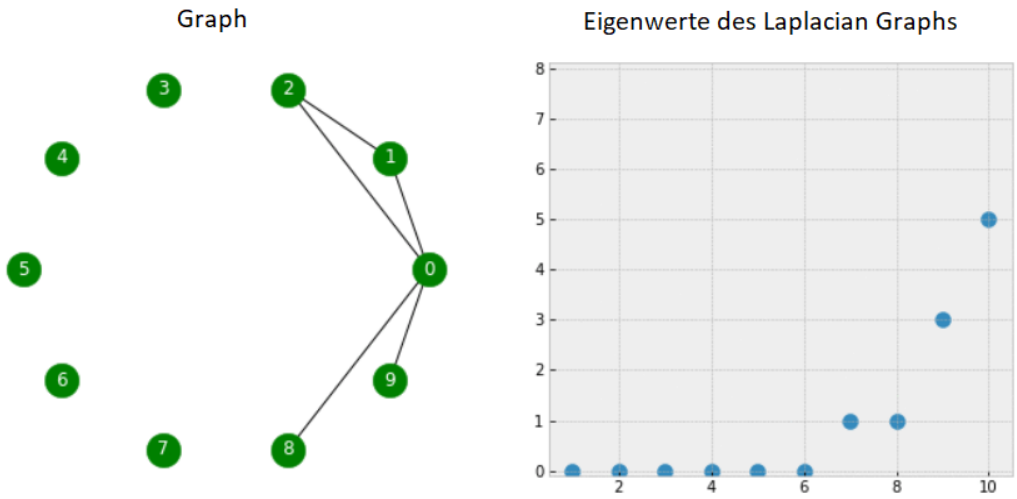


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

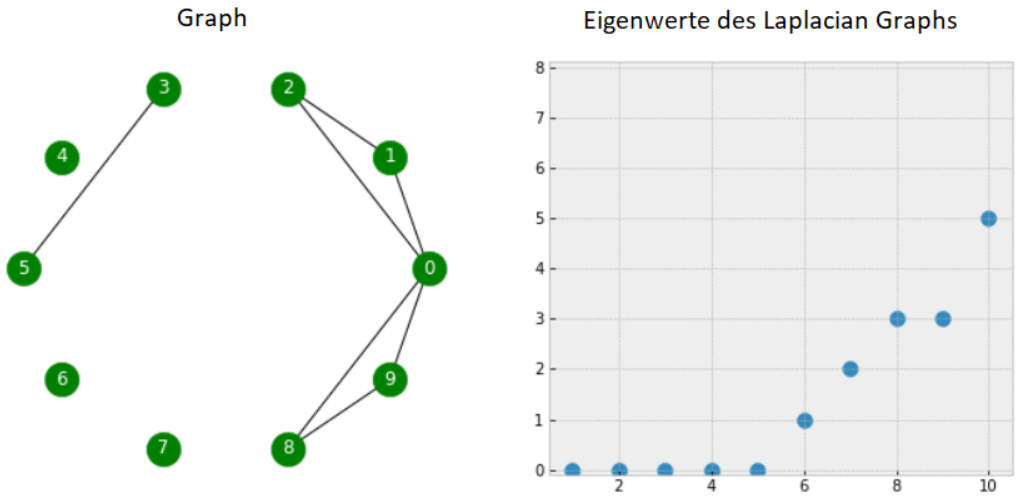


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

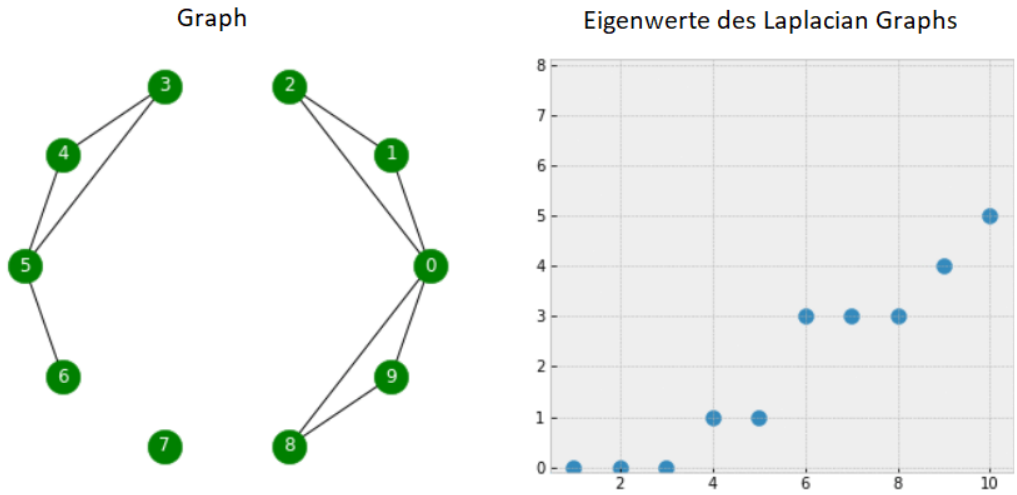


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

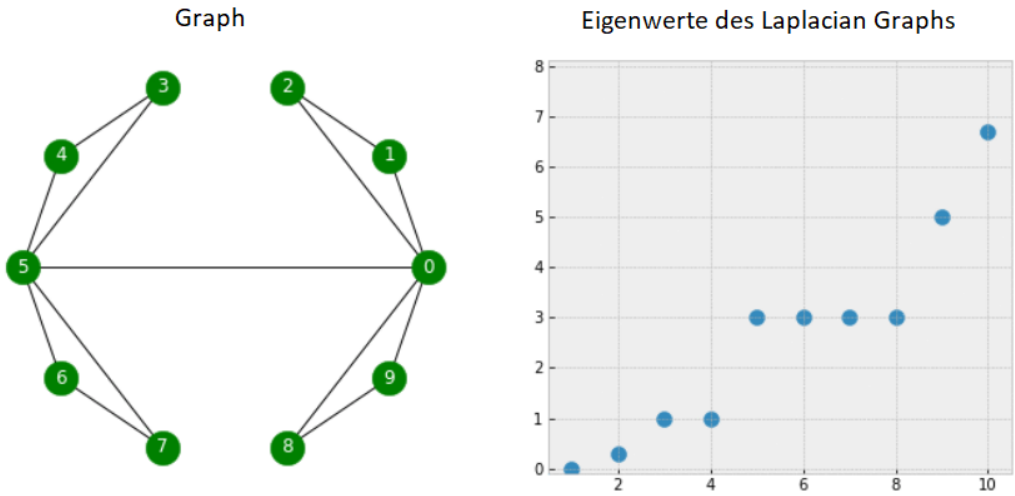


Abbildung: Zusammenhang zwischen den Eigenwerten von L und dem Graphen

Konkreter

- Der Graph ist vollständig getrennt, wenn alle zehn Eigenwerte 0 sind

Konkreter

- Der Graph ist vollständig getrennt, wenn alle zehn Eigenwerte 0 sind
- Fügen wir Kanten hinzu steigen einige unserer Eigenwerte

Konkreter

- Der Graph ist vollständig getrennt, wenn alle zehn Eigenwerte 0 sind
- Fügen wir Kanten hinzu steigen einige unserer Eigenwerte
- Die Anzahl der 0 Eigenwerte entspricht der Anzahl der verbundenen Komponenten in unserem Graphen

Konkreter

- Der Graph ist vollständig getrennt, wenn alle zehn Eigenwerte 0 sind
- Fügen wir Kanten hinzu steigen einige unserer Eigenwerte
- Die Anzahl der 0 Eigenwerte entspricht der Anzahl der verbundenen Komponenten in unserem Graphen
- Sind alle Eigenwerte bis auf einem ungleich Null haben wir einen zusammenhängenden Graphen

Konkreter

- Der erste Eigenwert ungleich Null wird als **Spektrumsabstand** bezeichnet

Konkreter

- Der erste Eigenwert ungleich Null wird als **Spektrumsabstand** bezeichnet
- Der Spektrumsabstand gibt uns eine Vorstellung von der Dichte des Graphens

Konkreter

- Der erste Eigenwert ungleich Null wird als **Spektrumsabstand** bezeichnet
- Der Spektrumsabstand gibt uns eine Vorstellung von der Dichte des Graphens
- Der zweite Eigenwert wird als **Fiedler-Wert** bezeichnet, und der entsprechende Vektor ist der **Fiedler-Vektor**

Konkreter

- Der erste Eigenwert ungleich Null wird als **Spektrumsabstand** bezeichnet
- Der Spektrumsabstand gibt uns eine Vorstellung von der Dichte des Graphens
- Der zweite Eigenwert wird als **Fiedler-Wert** bezeichnet, und der entsprechende Vektor ist der **Fiedler-Vektor**
- Der Fiedler-Wert nähert sich dem minimalen Graphenschnitt, der erforderlich ist, um den Graphen in zwei verbundene Komponenten zu trennen

Konkreter

- Der erste Eigenwert ungleich Null wird als **Spektrumsabstand** bezeichnet
- Der Spektrumsabstand gibt uns eine Vorstellung von der Dichte des Graphens
- Der zweite Eigenwert wird als **Fiedler-Wert** bezeichnet, und der entsprechende Vektor ist der **Fiedler-Vektor**
- Der Fiedler-Wert nähert sich dem minimalen Graphenschnitt, der erforderlich ist, um den Graphen in zwei verbundene Komponenten zu trennen
- Jeder Wert im Fiedler-Vektor gibt uns Aufschluss darüber, zu welcher Seite des Schnittes dieser Knoten gehört

Konkreter

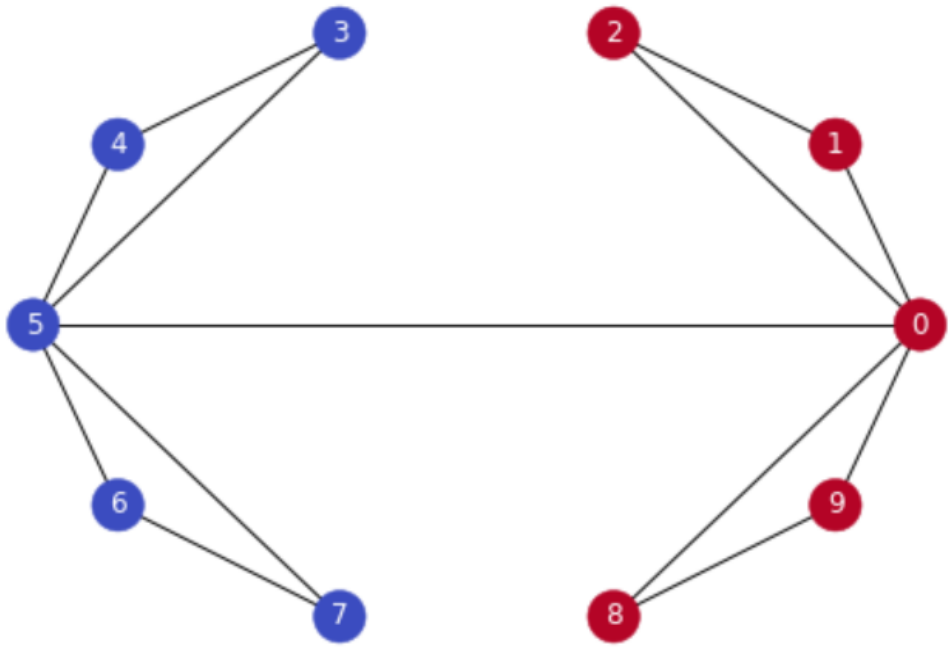


Abbildung: Die Färbung der Knoten richtet sich danach, ob ihr Eintrag im Fiedler-Vektor gößer oder kleiner Null ist

Konkreter

- Existieren zwei kleine Eigenwerte, haben wir zwei Cluster

Konkreter

- Existieren zwei kleine Eigenwerte, haben wir zwei Cluster
- Existieren drei kleine Eigenwerte, so haben wir drei Cluster und so weiter

Konkreter

- Existieren zwei kleine Eigenwerte, haben wir zwei Cluster
- Existieren drei kleine Eigenwerte, so haben wir drei Cluster und so weiter
- Im Allgemeinen wird oft nach der ersten großen Lücke zwischen den Eigenwerten gesucht, um die Anzahl der Cluster zu finden

Konkreter

- Existieren zwei kleine Eigenwerte, haben wir zwei Cluster
- Existieren drei kleine Eigenwerte, so haben wir drei Cluster und so weiter
- Im Allgemeinen wird oft nach der ersten großen Lücke zwischen den Eigenwerten gesucht, um die Anzahl der Cluster zu finden
- Die Vektoren, die den ersten $n-1$ positiven Eigenwerten zugeordnet sind, geben Aufschluss darüber, welche Schnitte im Graphen vorgenommen werden müssen, um jeden Knoten einer der n Cluster zuzuordnen

Konkreter

- Existieren zwei kleine Eigenwerte, haben wir zwei Cluster
- Existieren drei kleine Eigenwerte, so haben wir drei Cluster und so weiter
- Im Allgemeinen wird oft nach der ersten großen Lücke zwischen den Eigenwerten gesucht, um die Anzahl der Cluster zu finden
- Die Vektoren, die den ersten $n-1$ positiven Eigenwerten zugeordnet sind, geben Aufschluss darüber, welche Schnitte im Graphen vorgenommen werden müssen, um jeden Knoten einer der n Cluster zuzuordnen
- Diese $n-1$-Vektoren werden nun in eine neue Matrix geschrieben und mit Hilfe von k-Mean Clustering kann eine Zuordnung der Daten erfolgen

L als Blockmatrix

- Können die einzelnen Cluster einfach von einander unterschieden werden, dann ist L eine Blockmatrix

L als Blockmatrix

- Können die einzelnen Cluster einfach von einander unterschieden werden, dann ist L eine Blockmatrix
- Angenommen wir haben drei Hauptcluster $(C1, C2, C3)$

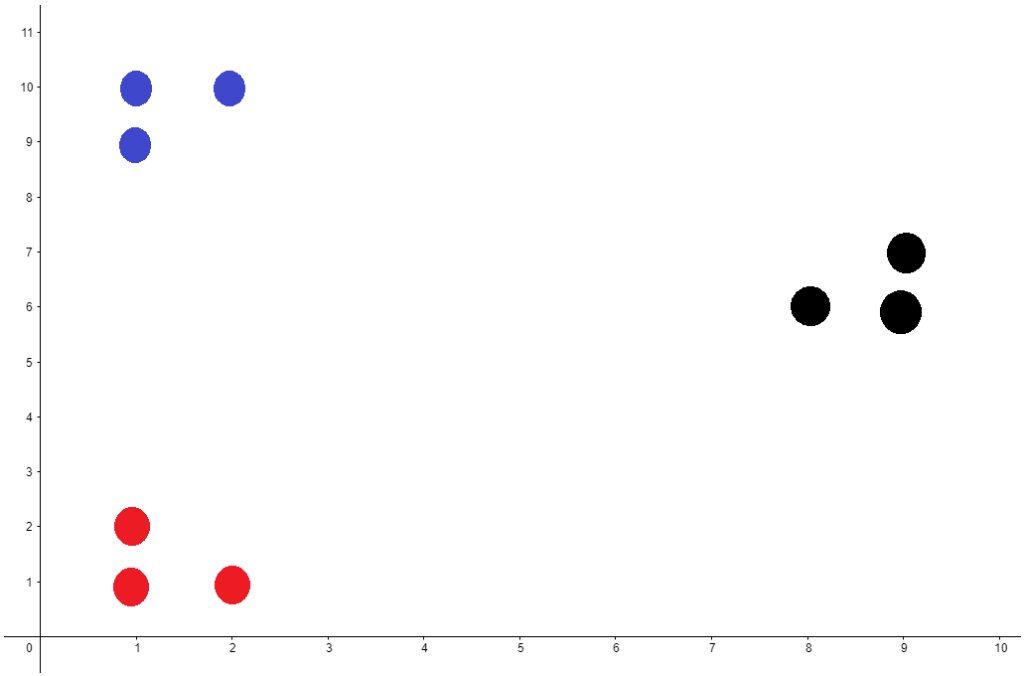


Abbildung: Datensatz mit drei Hauptclustern

L als Blockmatrix

$$\left(\begin{array}{cc|cc|cc} & & 0 & 0 & 0 & 0 \\ \multicolumn{2}{c|}{L_1} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{2}{c|}{\multirow{2}{*}{L_2}} & 0 & 0 \\ 0 & 0 & & & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{L_3}} \\ 0 & 0 & 0 & 0 & & \end{array}\right)$$

- Die einzelnen Blöcke entsprechen nun den einzelnen Clustern

Fragen

- Welche Art von Ähnlichkeitsgraphen soll benutzt werden?

Fragen

- Welche Art von Ähnlichkeitsgraphen soll benutzt werden?
- Wie viele Eigenvektoren sollen gewählt werden, die aus L extrahiert werden?

Fragen

- Welche Art von Ähnlichkeitsgraphen soll benutzt werden?
- Wie viele Eigenvektoren sollen gewählt werden, die aus L extrahiert werden?
- Wie viele Cluster werden benötigt?

Vielen Dank fürs Zuhören!

 T. Hastie, R. Tibshirani, J.H. Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, 2. Aufl.: Springer 2009.

 Naresh Kumar: Advantages and Disadvantages of Principal Component Analysis in Machine Learning. URL: http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of_4.html/ [27.06.2019].

 William Fleshman: Spectral Clustering Foundation and Application URL: https://towardsdatascience.com/spectral-clustering-aba2640c0d5b [27.06.2019].

 Prasoon Goyal: What are the disadvantages of a PCA? URL: https://www.quora.com/What-are-the-disadvantages-of-a-PCA [27.06.2019].