

Nichtlineare Projektion mit dem Sammon-Verfahren

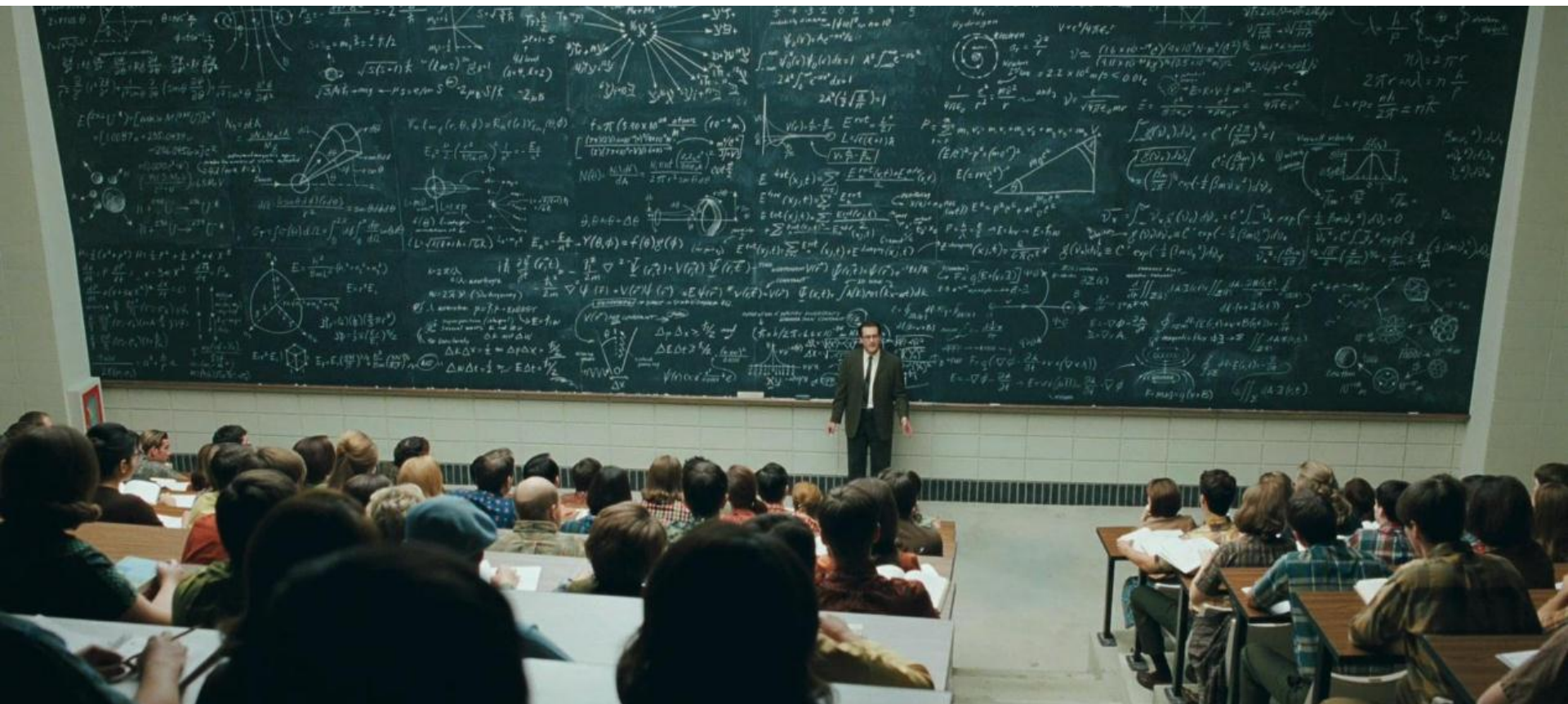
Datenvisualisierung in der Statistik

Tobias Siebig, Morten Terhart
16.03.2018, Mosbach

Gliederung

- » Zielsetzung / Nutzen
- » Methodik
- » Einordnung in den statistischen Kontext
- » Anwendung an Beispielen





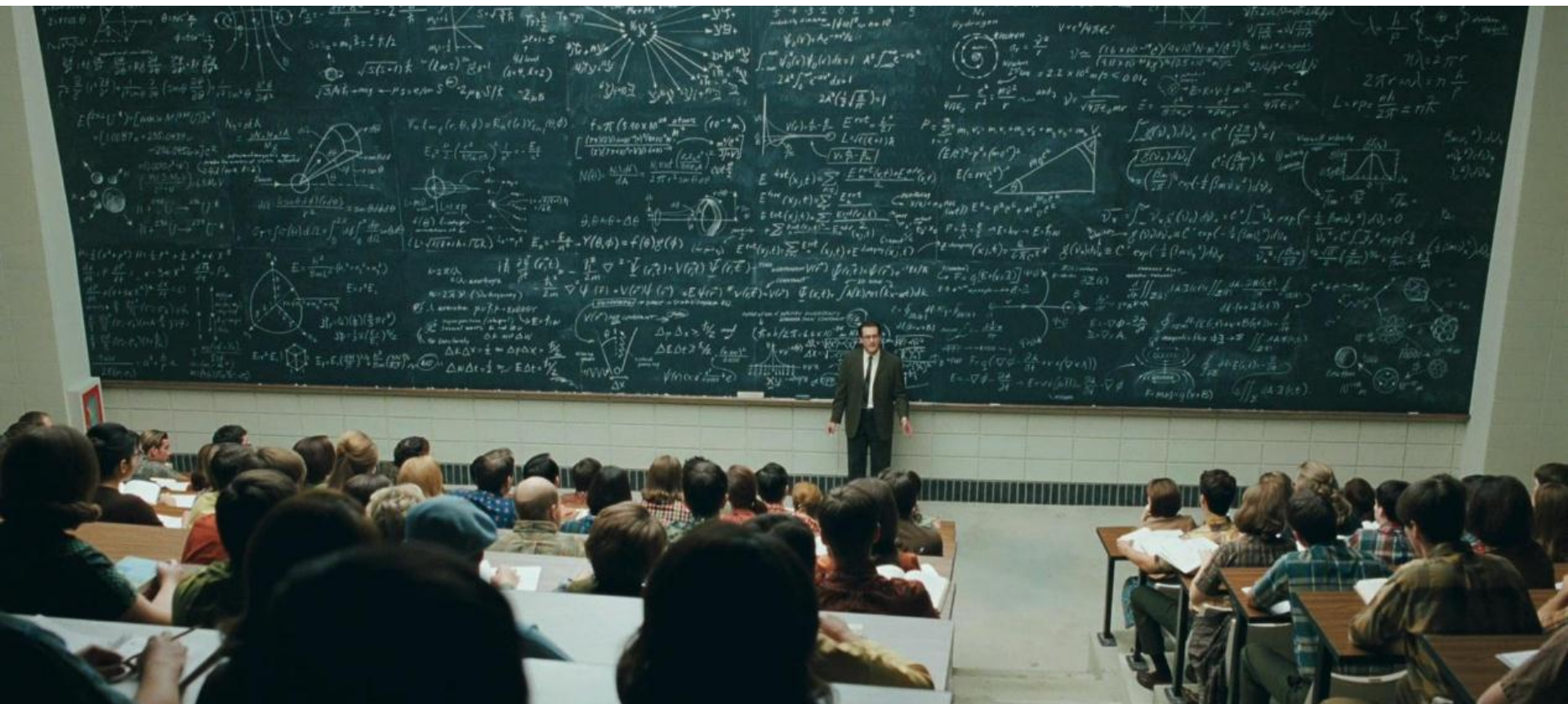
Zielsetzung / Nutzen



Zielsetzung / Nutzen

- » Skalierung von höher zu niedriger dimensional Datensätzen
 - Anschaulichere Darstellung
- » Nichtlineare Methode
- » Keine explizite Transformationsfunktion
- » Annäherung eines Datensatzes an die Struktur
 - Erhaltung der Strukturen





Methodik



Methodik

- » **Idee:** Abbildung eines Datensatzes X auf einen anderen Datensatz Y , sodass die Abstände zwischen den Punkten angenähert werden
- 1. Aufstellung der Distanzmatrizen D^x und D^y
- 2. Minimierung des Fehlers zwischen D^x und D^y mithilfe der *Sammon-Fehlerfunktion*

$$E_3 = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j}^N \frac{d_{ij}^* - d_{ij}}{d_{ij}^*}$$

- 3. Anwendung des Gradientenverfahrens zur Bestimmung des Minimums von E_3
 - Minimaler Fehler liefert nächste Annäherung an die Projektion
- 4. Wiederhole die Schritte 1 bis 3



Methodik - Beispiel

» **Initialisierung:** $Y = \{1, 2, 3, 4\}$

1. Berechne die Distanzmatrix

$$D^Y = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}.$$

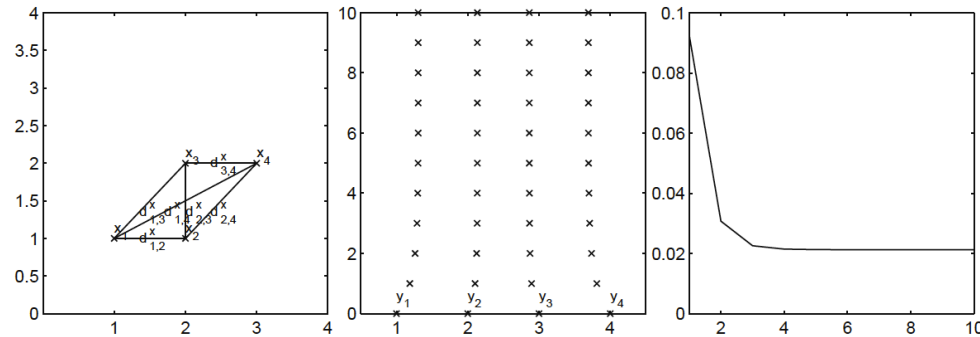
2. Bestimme den Wert der *Sammon-Fehlerfunktion*

$$E_3 = \frac{1}{3 \cdot 1 + 2 \cdot \sqrt{2} + \sqrt{5}} \cdot \left(2 \cdot \frac{(2 - \sqrt{2})^2}{\sqrt{2}} + \frac{(3 - \sqrt{5})^2}{\sqrt{5}} \right) \approx 0.0925.$$

3. Ermittle den minimalen Fehlers durch das Gradientenverfahren

$$\frac{\partial E_3}{\partial y_1} = \frac{2}{3 \cdot 1 + 2 \cdot \sqrt{2} + \sqrt{5}} \cdot \left(-\frac{2 - \sqrt{2}}{\sqrt{2}} - \frac{3 - \sqrt{5}}{\sqrt{5}} \right) \approx -0.1875.$$

» **Projektion:** Annäherung durch $Y \approx \{1.1875, 2.1027, 2.8973, 3.8125\}$



Quelle: Runkler, "Data Mining"





Einordnung in den statistischen Kontext



Einordnung in den statistischen Kontext

- » Nichtlineare Methode
- » Parallelen zur Hauptkomponentenanalyse
 - Dimensionale Reduktion
- » Projektion kann weitergehend analysiert werden, z.B. durch
 - Korrelation (linear oder multivariat)
 - Regression





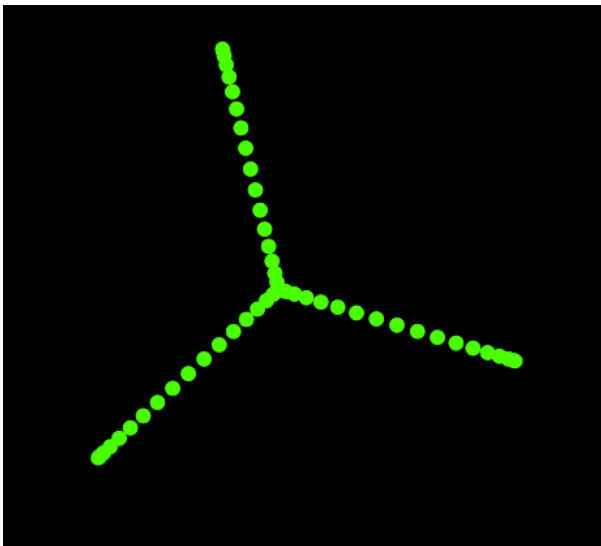
Anwendung an Beispielen



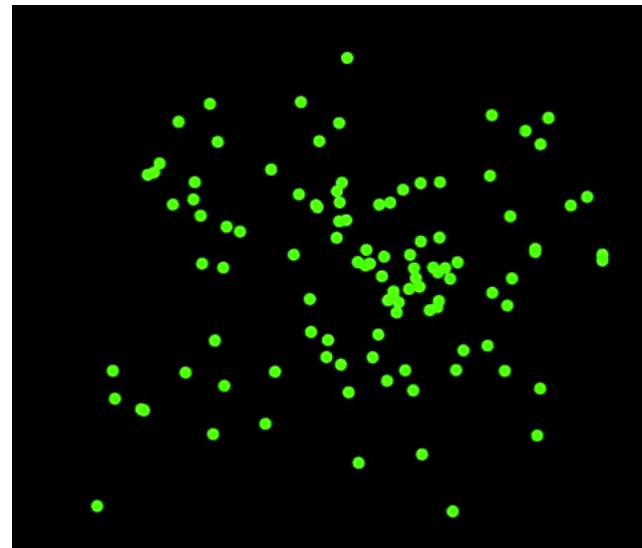
Projektion mit dem Sammon-Verfahren

Bouquet of Circles

- » 3 Ebenen in einem 6-dimensionalen Raum
- » Auf jeder Ebene ein Kreis, die sich in einem Punkt schneiden
- » Sammon-Verfahren klappt Kreise auf sichtbare Dimension um



Hauptachsentransformation

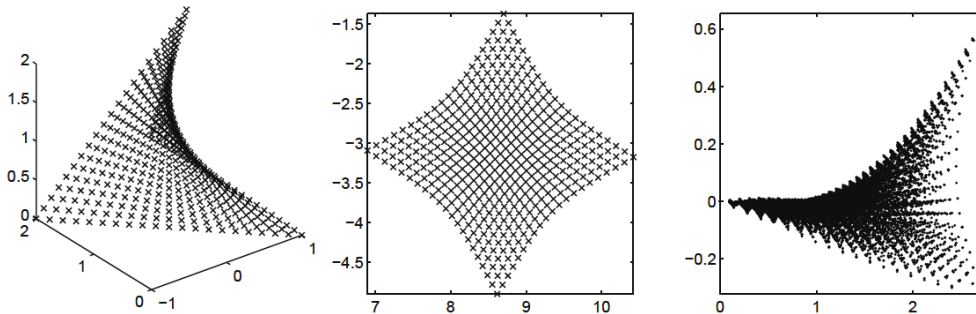


Sammon-Verfahren

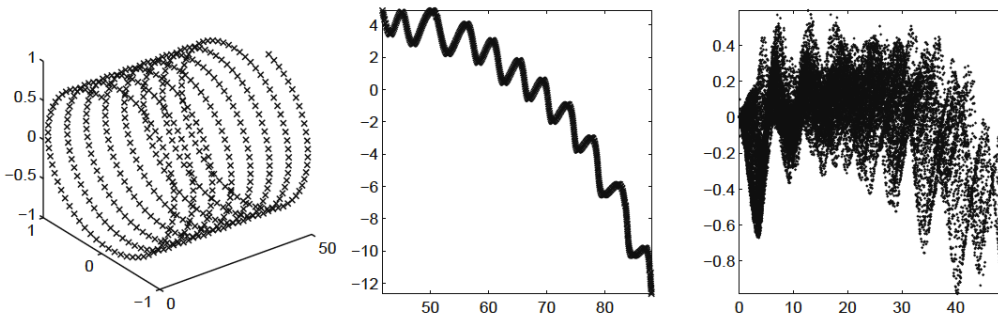
Weitergehende Informationen unter [Wikipedia \(Rose \(Topology\)\)](#)



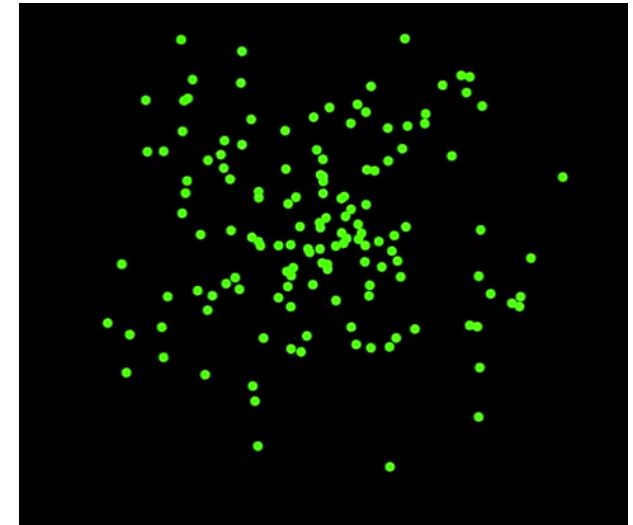
Projektion mit dem Sammon-Verfahren



Projektion eines gebogenen Quadrates



Projektion eines Helix-Datensatzes



Bouquet of Circles mit 4 Kreisen
im 8-dimensionalen Raum

Quelle: Runkler, „Data Mining“



Literatur

- » J.W. Sammon (1969). „*A nonlinear mapping for data structure analysis*“. IEEE Transactions on Computers. 18: 401, 402,403–409.
- » Thomas A. Runkler (2015). „*Data Mining – Modelle und Algorithmen intelligenter Datenanalyse*“. In Computational Intelligence, S. 47-51. 2. Auflage. Springer Vieweg. Berlin 2015.
- » Elzbieta Pękalska, Dick de Ridder, Robert P.W. Duin and Martin A. Kraaijveld (1999). „*A new method of generalizing Sammon mapping with application to algorithm speed-up*“. Proc. 5th Annual Conference of the Advanced School for Computing and Imaging (ASCI1999).
- » Scott Hotton and Jeff Hoshimi. „*HiSee – Projection Algorithms*“. Online: <http://hisee.sourceforge.net> (Abrufdatum: 08.03.2018).
- » Paul Henderson (2012). „*Sammon Mapping*“. Online: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0910/henderson.pdf (Abrufdatum: 08.03.2018).



Vielen Dank für eure Aufmerksamkeit.

Duale Hochschule Baden-Württemberg Mosbach
Lohrtalweg 10
74821 Mosbach

- » Tobias Siebig
Tob.Siebig.16@lehre.mosbach.dhbw.de

- » Morten Terhart
Mor.Terhart.16@lehre.mosbach.dhbw.de