

MUSTERERKENNUNG

Vorlesung im Sommersemester 2017

Prof. E.G. Schukat-Talamazzini

Stand: 6. März 2017

Teil V

Merkmalgewinnung

Aufgabe	Reihen	Spektrum	Dimensionen	PCA	LDA	SVD	f
---------	--------	----------	-------------	-----	-----	-----	---

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

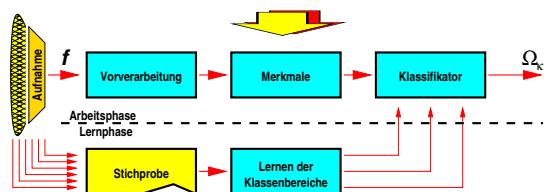
Matrixfaktorisierung

Mathematische Hilfsmittel

Aufgabe	Reihen	Spektrum	Dimensionen	PCA	LDA	SVD	f
---------	--------	----------	-------------	-----	-----	-----	---

Merkmalgewinnung

Transformation $\mathbb{R}^N \rightarrow \mathbb{R}^M$, $M \ll N$, M fest



Transformationen

Anwendungsbezogene
Merkmale

Spektralkoeffizienten

Vorhersagekoeffizienten

Unvollständige
Reihenentwicklung

Ziele

- reduzierte Dimension
- konstante Dimension
- scharfe Klassentrennung

Aufwand, Robustheit
alle Muster in einem \mathbb{R}^M

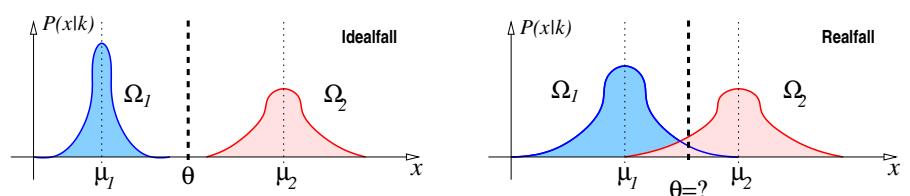
gleiche Klasse — geringe Distanz
(und umgekehrt)

Klassentrennung

Klassengebiete sind Zonen mit hoher Verteilungsdichte

Problem

Überlagern sich die Klassengebiete im gewählten Merkmalraum, so werden Muster im Überlappungsbereich häufig falsch klassifiziert.



Lösung

Verwende solche Merkmale x oder Merkmalkombinationen $\mathbf{x} = (x_1, \dots, x_D)$ zur Musterrepräsentation, die weitgehend disjunkte Klassengebiete gestalten.

Vorgehensweisen zur Merkmalgewinnung

Merkmalextraktion · Merkmaltransformation · Merkmalselektion

Keine Operation

Die ATW des Musters werden als Merkmale genutzt ($\mathbf{x} = \mathbf{f}$).

Heuristische Gewinnung (Extraktion)

Welche Merkmale $x_i(\mathbf{f})$ erscheinen subjektiv bedeutsam?

Analytische Gewinnung (Transformation)

Welche Merkmalkombination $\mathbf{x}(\mathbf{f})$ maximiert ein relevantes Gütekriterium?

(musterbezogene \leftrightarrow ergebnisbezogene Kriterien)

Kombinatorische Merkmalauswahl (Selektion)

Welche Teilmenge $\mathcal{Y} \subset \{x_1, \dots, x_D\}$ aller Merkmale erbringt die maximale Leistung?

Beispiele

Anwendungsabhängig („heuristisch“) definierte Merkmale

- das Muster $[f_n]$ bzw. $[f_{nm}]$ selbst als Merkmalvektor ($\in \mathbb{R}^N$)
- die Spektralkoeffizienten $|F_\nu|$ (hörbarer Bereich) einer Stimmprobe $[f_n]$
- das Grauerthistogramm $\{Q_\ell\}$ eines Zellschnittbildes ($\in \mathbb{R}^L$)
- die Winkelprojektionen \mathbf{f}^α — α in 45° -Schritten — eines gescannten Druckzeichens ($\in \mathbb{R}^{8N}$)
- der Kettencode der Mittelachse einer Unterschrift ($\in \{0, \dots, 7\}^*$)
- das Phasenbild $[\arg F_{\nu\mu}]$
das Kantenrichtungsbild $[\phi_{nm}]$
- mittlere, minimale, maximale Intensität des Abtastfeldes
die Streuung oder andere ausgewählte Momente μ_{pq}
- Konturlängen, Objektflächen, Volumina
-

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

Orthonormalsysteme

Vektorraumbasen mit paarweise senkrecht stehenden Basisvektoren der Länge 1

Definition

Eine Folge ϕ_1, \dots, ϕ_N linear unabhängiger Vektoren des \mathbb{R}^N mit der Eigenschaft

$$\phi_\nu^\top \phi_\mu = \begin{cases} 1 & \nu = \mu \\ 0 & \nu \neq \mu \end{cases}$$

heißt **Orthonormalsystem**.

$$\begin{array}{c|c|c} \Phi^\top & \Phi & E \end{array} \cdot \begin{array}{c|c|c} \Phi & \Phi^\top & ? \end{array} = \begin{array}{c|c|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

ist klar, aber gilt auch:

Lemma

Φ bezeichne die **Orthogonalmatrix** mit Basisvektor ϕ_ν in der ν -ten Spalte.
Dann gilt:

$$\Phi^\top \cdot \Phi = E = \Phi \cdot \Phi^\top$$

Beweis.

Die Aussage $\Phi^\top \Phi = E$ folgt unmittelbar aus der Definition einer Orthogonalmatrix.

Weil Φ^\top die Linksinverse von Φ ist, besitzt Φ den vollen Rang N .

Folglich besitzt Φ auch eine Rechtsinverse; diese muß aber auch gleich Φ^\top sein, weil man zeigen kann, daß sogar für beliebige quadratische Matrizen die Linksinverse und die Rechtsinverse eindeutig und gleich sind, falls sie existieren.

Aus $L \cdot Q = E = Q \cdot R$ folgt nämlich:

$$R - L = ER - LE = LQR - LQR = 0$$

□

Bemerkung

Achtung! Wenn man Matrizen Φ' betrachtet, deren Spalten eine **unvollständige Orthonormalbasis** bilden, so werden die Zusammenhänge etwas verwickelter:

$$\begin{array}{c|c|c} \Phi'^\top & \Phi' & E_{N'} \end{array} , \text{ aber: } \begin{array}{c|c|c} \Phi' & \Phi'^\top & (N \times N) \end{array} \neq \begin{array}{c|c|c} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{array} E_N$$

Vollständige Reihenentwicklung

Mit den richtigen Koeffizienten ist Muster f immer exakt reproduzierbar

Lemma

Es sei Φ ein Orthonormalsystem und $f \in \mathbb{R}^N$. Für die Summe

$$\hat{f} = \Phi \cdot c = \sum_{\nu=1}^N c_\nu \cdot \phi_\nu$$

mit den Entwicklungskoeffizienten

$$c_\nu = \phi_\nu^\top f \quad \text{bzw.} \quad c = \Phi^\top \cdot f$$

gilt die Aussage $\hat{f} = f$.

Beweis durch Einsetzen:

$$\hat{f} = \Phi c = \Phi \cdot (\Phi^\top f) = (\Phi \Phi^\top) \cdot f = Ef = f$$

Unvollständige Reihenentwicklung

Mit welchen Koeffizienten ergibt sich der kleinste Fehler zwischen f und \hat{f} ?

Satz

Für die unvollständige Darstellung

$$\hat{f} = \sum_{\nu=1}^{N'} c_\nu \cdot \phi_\nu, \quad N' < N$$

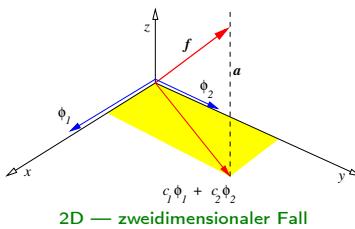
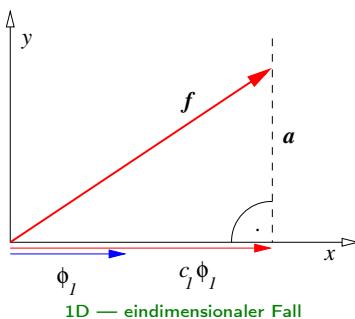
von f nach dem Orthonormalsystem Φ gilt: der mittlere quadratische Rekonstruktionsfehler

$$\varepsilon = \|f - \hat{f}\|^2 = (f - \hat{f})^\top (f - \hat{f})$$

wird minimal, wenn die Reihenkoeffizienten gemäß

$$c_\nu = \phi_\nu^\top f, \quad \nu = 1, \dots, N'$$

gewählt werden.

**Beweis.**

Das **Orthogonalitätsprinzip** besagt:
Der Abstand zwischen f und $c_1\phi_1$ ist minimal genau dann, wenn

$$\mathbf{a} = \mathbf{f} - c_1\phi_1$$

senkrecht auf ϕ_1 steht (siehe Abbildung 1D).

Also ist zu fordern:

$$\phi_1^\top (\mathbf{f} - c_1\phi_1) = 0$$

Allgemeiner (siehe Abbildung 2D) lautet für die auf $(N' \times N)$ reduzierte Orthogonalmatrix Φ' unsere Forderung

$$\begin{aligned}\Phi'^\top (\mathbf{f} - \Phi' \mathbf{c}) &= 0 \\ \Phi'^\top \mathbf{f} &= \Phi'^\top (\Phi' \mathbf{c}) \\ \Phi'^\top \mathbf{f} &= (\Phi'^\top \Phi') \mathbf{c} \\ \Phi'^\top \mathbf{f} &= \mathbf{E} \mathbf{c} = \mathbf{c}\end{aligned}$$

□

Walsh-Funktionen

Eulerformelartige Zerlegung in gerade und ungerade Anteile

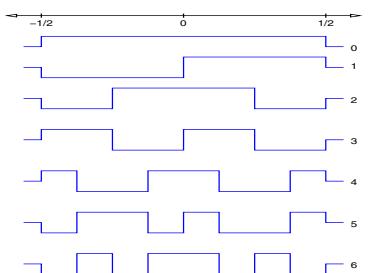
Definition

Die kontinuierlichen Walshfunktionen $wal_n(x)$ sind definiert durch:

$$wal_0(x) = \begin{cases} 1 & -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0 & \text{sonst} \end{cases}$$

$$wal_{2n}(x) = (-1)^{\lfloor \frac{n}{2} \rfloor} \cdot \left(wal_n\left(\frac{4x+1}{2}\right) + (-1)^n wal_n\left(\frac{4x-1}{2}\right) \right)$$

$$wal_{2n+1}(x) = (-1)^{\lfloor \frac{n}{2} \rfloor + 1} \cdot \left(wal_n\left(\frac{4x+1}{2}\right) + (-1)^{n+1} wal_n\left(\frac{4x-1}{2}\right) \right)$$

**Satz**

Die Walshfunktionen sind orthogonal. d.h. es gilt für alle n, m :

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} wal_n(x) \cdot wal_m(x) dx = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

DWT — Diskrete Walsh-Hadamard-Transformation

In Zweierpotenzordnung abgetastete Walshfunktionen

Definition

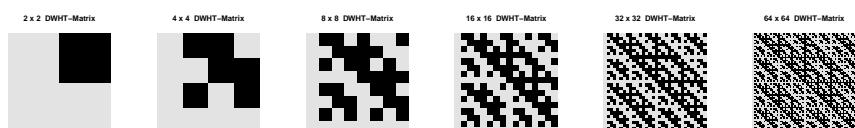
Die **Hadamardmatrizen** sind definiert durch:

$$\mathbf{H}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{H}_N = \mathbf{H}_2 \otimes \mathbf{H}_{N/2} = \begin{pmatrix} \mathbf{H}_{N/2} & \mathbf{H}_{N/2} \\ \mathbf{H}_{N/2} & -\mathbf{H}_{N/2} \end{pmatrix}$$

Für $N = 2^p$ heißt die Abbildung

$$\mathbf{c} = \text{WHT}(\mathbf{f}) = \mathbf{H}_N \mathbf{f} \quad \text{bzw.} \quad \mathbf{f} = \text{WHT}^{-1}(\mathbf{c}) = \frac{1}{N} \mathbf{H}_N \mathbf{c}$$

(inverse) **Walsh-Hadamard-Transformation** d. Ordnung N .

**Haar-Funktionen und DHT**

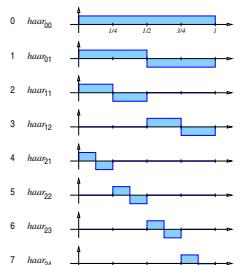
Eine ganz besonders „eckige“ Wavelet-Basis

Definition

Die **Haarfunktionen** $haar_{pq} : [0, 1] \rightarrow \mathbb{R}$ sind definiert durch:

$$\begin{aligned}haar_{00}(x) &= 1 / \sqrt{N} \\ haar_{pq}(x) &= \frac{1}{\sqrt{N}} \cdot \begin{cases} 2^{p/2} & q-1 \leq 2^p x < q-1/2 \\ -2^{p/2} & q-1/2 \leq 2^p x < q \\ 0 & \text{sonst} \end{cases}\end{aligned}$$

$k = 0, 1, \dots, N-1$ mit $k = 2^p + q - 1$ für gegebene Ordnung $N = 2^n \in \mathbb{N}$.

**Diskrete Haar-Transformation für $N = 8$**

$$\text{DHT}(\mathbf{f}) = \frac{1}{\sqrt{8}} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{pmatrix} \cdot \mathbf{f}$$

Legendre-Polynome

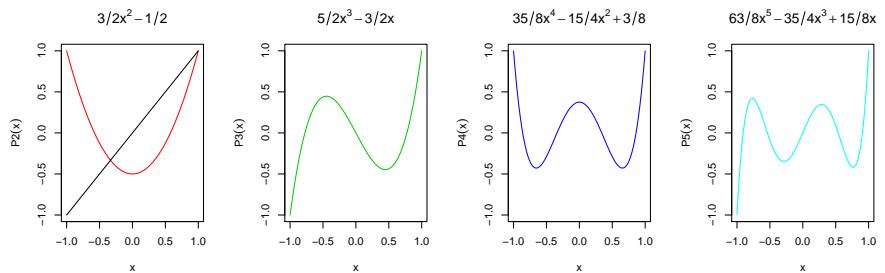
Definition

Die **Legendre-Polynome** sind rekursiv definiert durch:

$$Q_0(x) = 1$$

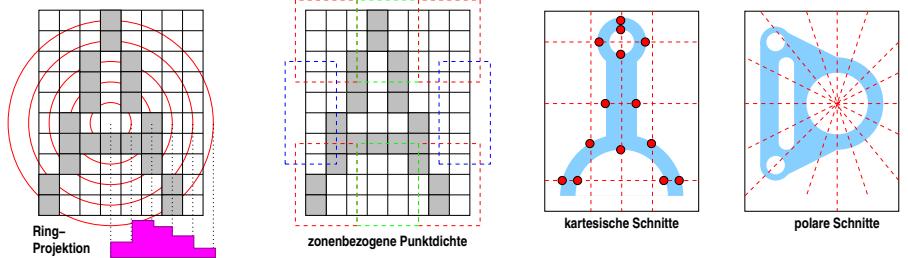
$$Q_1(x) = x$$

$$Q_p(x) = \frac{2p-1}{p}x \cdot Q_{p-1}(x) - \frac{p-1}{p} \cdot Q_{p-2}(x)$$



Heuristische Reihenentwicklungen

nicht Teil einer Orthonormalbasis · auf spezielle Anwendungen beschränkt



Projektionen

- \$x/y\$-Achse
- Winkelprojektion
- Ringprojektion

Schnittpunkte mit Testlinien

- kartesische Koordinaten \$(x, y)\$
- polare Koordinaten \$(r, \phi)\$

Zonenpunktdichten

$$c_D(f) = \sum_{(n,m) \in D} f_{nm}$$

Legendre-Momente und Legendre-Interpolation

Definition

Für die Funktion \$f : \mathbb{R}^2 \rightarrow \mathbb{R}\$ bezeichnet

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{+1} \int_{-1}^{+1} Q_p(x) Q_q(y) f(x, y) dx dy$$

das \$(p, q)\$-te **Legendre-Moment**. Für \$[f_{nm}]\$ bezeichnet

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{N \cdot M} \sum_{n=1}^N \sum_{m=1}^M Q_p(n\Delta x) Q_q(m\Delta y) f_{nm}$$

das \$(p, q)\$-te **diskrete Legendre-Moment**.

Satz

Für ein Muster \$[f_{nm}]\$ und seine Legendre-Momente \$\{\lambda_{pq}\}\$, \$p \leq n\$, \$q \leq m\$ gilt die Näherungsformel

$$f_{nm} \approx \hat{f}(n\Delta x, m\Delta y)$$

mit

$$\hat{f}(x, y) \stackrel{\text{def}}{=} \sum_{p=0}^N \sum_{q=0}^M \lambda_{pq} Q_p(x) Q_q(y)$$

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

Diskretes Fourierspektrum

Fakt

Die Koeffizienten

$$F_\nu = \sum_{n=0}^{N-1} f_n \cdot W_N^{n\nu} = \phi_\nu^\top \mathbf{f}$$

der diskreten Fourier-Transformation sind das Resultat einer komplexwertigen vollständigen Reihenentwicklung mit den orthogonalen Basisvektoren

$$\phi_\nu \stackrel{\text{def}}{=} \left(W_N^0, W_N^\nu, W_N^{2\nu}, \dots, W_N^{(N-1)\nu} \right)^\top.$$

Orthogonalität im Komplexwertigen:

$$\phi_\nu^\top \phi_\mu^* = \begin{cases} 1 & \nu = \mu \\ 0 & \nu \neq \mu \end{cases}, \quad \text{wobei } (a+bi)^* = (a-bi)$$

Lemma

Die Koeffizienten $c_\nu \stackrel{\text{def}}{=} |F_\nu|^2 = F_\nu \cdot F_\nu^*$ des **Betragssquadratspektrums** sind reellwertig und invariant gegenüber (zyklischen) Verschiebungen des Musters \mathbf{f} .

Beweis.

Die zyklische Verschiebung der Abtastwertfolge entspricht im Spektralbereich einer **Frequenzmodulation**.

$$\begin{aligned} [F_\nu] &= \text{DFT}\{[f_n]\} \\ [F_\nu^{(\tau)}] &= [F_\nu W_N^{-\nu\tau}] = \text{DFT}\{[f_{n-\tau}]\} = \text{DFT}\{[f_n^{(\tau)}]\} \\ |F_\nu^{(\tau)}|^2 &= F_\nu W_N^{-\nu\tau} \cdot F_\nu^* W_N^{+\nu\tau} = |F_\nu|^2 \cdot 1 \end{aligned}$$

Nach Betragsbildung heben sich die beiden Exponentialterme auf. □

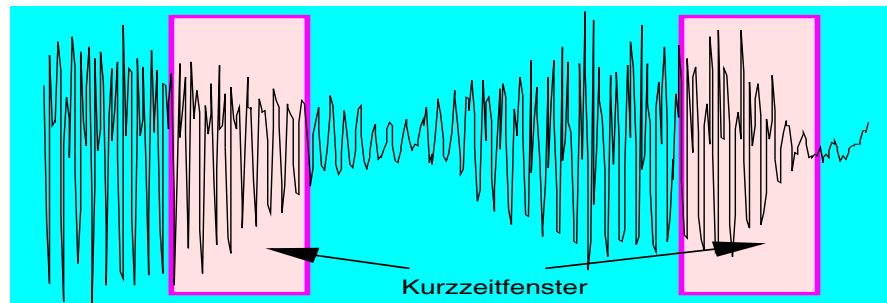
Bemerkung

Die Phaseninformation des komplexen Spektrums beschreibt allgemein die zeitliche Verschiebung aller Sinuskomponenten des Signals zueinander, speziell also auch die Gesamtverschiebung des Signals auf der Zeitachse.

Die Phaseninformation gilt als relativ unwichtig bei Schallsignalen, bei Bildsignalen jedoch als unverzichtbar!

Spektrum eindimensionaler Muster

Kurzzeitanalyse nicht stationärer Signale



Kurzzeitspektrum

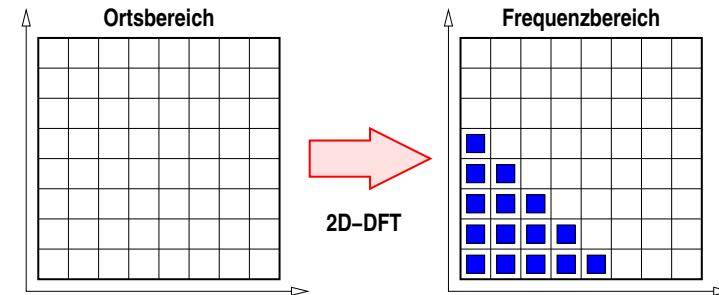
Erfassen zeitveränderlicher spektraler Signaleigenschaften:

$$[F_\nu^{(m)}] = \text{DFT}\{(f_m, \dots, f_{m+N-1})\}, \quad m = j \cdot r, \quad j \in \mathbb{Z}$$

N = Fensterbreite, m = Fensterposition, r = Fortschaltrate.

Spektrum zweidimensionaler Muster

Betrag und Phase ausgewählter Frequenzkomponenten

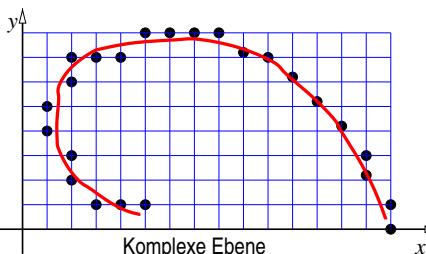


2D-Spektrum

Für kleine Frequenzindizes beschreibt $F_{\nu\mu}$ *langsame* Veränderungen (*große Strukturen*) des Musters.

Spektrum von Linien und Konturen

Fourier-Deskriptoren sind invariant gegenüber Translation und Konturstart



Fourier-Deskriptoren

Repräsentiere (Kontur-)Linie als Koordinatenfolge ($\mathbb{R}^2 \cong \mathbb{C}$)

$$[f_{nm}] \mapsto [z_k] \quad \text{mit} \quad z_k = x_k + iy_k \quad (k = 0, \dots, K-1)$$

und berechne das Betragsspektrum zum komplexwertigen Signal:

$$c_\kappa = |Z_\kappa|^2 \quad \text{mit} \quad [Z_\kappa] = \text{DFT}\{[z_k]\}$$

Lineares Vorhersagemodell

Klassisches Verfahren zur approximativen Signalkompression (Codec)

Definition

Der Ausdruck

$$\hat{f}_n \stackrel{\text{def}}{=} - \sum_{\nu=1}^p a_\nu \cdot f_{n-\nu}$$

heißt **lineares Vorhersagemodell** der Ordnung $p \in \mathbb{N}$; die Zahlen a_1, \dots, a_p sind die **Vorhersagekoeffizienten** des Modells.

Mit der Bezeichnung $a_0 = 1$ besitzt das **Fehlersignal** zwischen Original $[f_n]$ und Ideal $[\hat{f}_n]$ die Form

$$e_n \stackrel{\text{def}}{=} f_n - \hat{f}_n = \sum_{\nu=0}^p a_\nu \cdot f_{n-\nu}.$$

Berechnung der Vorhersagekoeffizienten

Welcher LPC-Vektor (a_0, \dots, a_p) ergibt den minimalen Fehler?

Satz (Markel 1972)

Der mittlere quadratische Vorhersagefehler

$$\varepsilon = \sum_{n=N_0}^{N_1} (f_n - \hat{f}_n)^2$$

der Prädiktionsformel p -ter Ordnung im Abtastintervall $[N_0, N_1]$ wird gerade durch diejenigen Prädiktionskoeffizienten a_0, \dots, a_p minimiert, welche Lösung des linearen Gleichungssystems

$$\sum_{\nu=1}^p a_\nu \cdot \psi_{\nu\mu} = -\psi_{0\mu}, \quad \mu = 1, \dots, p$$

sind, wobei die Abkürzung

$$\psi_{\nu\mu} \stackrel{\text{def}}{=} \sum_{n=N_0}^{N_1} f_{n-\nu} \cdot f_{n-\mu}, \quad \nu, \mu = 0, \dots, p$$

vereinbart sei.

Beweis.

Für den Vorhersagefehler gilt

$$\begin{aligned} \varepsilon &= \sum_{n=N_0}^{N_1} (f_n - \hat{f}_n)^2 = \sum_{n=N_0}^{N_1} \left(\sum_{\nu=0}^p a_\nu \cdot f_{n-\nu} \right)^2 \quad (a_0 = 1) \\ &= \sum_{n=N_0}^{N_1} \sum_{\nu=0}^p \sum_{\mu=0}^p a_\nu \cdot f_{n-\nu} \cdot f_{n-\mu} \cdot a_\mu = \sum_{\nu=0}^p \sum_{\mu=0}^p a_\nu \cdot \underbrace{\left(\sum_{n=N_0}^{N_1} f_{n-\nu} \cdot f_{n-\mu} \right)}_{\psi_{\nu\mu}} \cdot a_\mu \end{aligned}$$

Wir bilden nun die partiellen Ableitungen

$$\frac{\partial \varepsilon}{\partial a_\mu} = 2 \cdot \sum_{\nu=0}^p a_\nu \cdot \psi_{\nu\mu}$$

nach den Vorhersagekoeffizienten a_μ und setzen diese gleich Null. Wegen $a_0 = 1$ ergibt sich daraus das gesuchte LGS

$$\sum_{\nu=1}^p a_\nu \cdot \psi_{\nu\mu} = -\psi_{0\mu} \quad (\mu = 1, \dots, p)$$

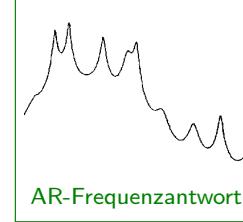
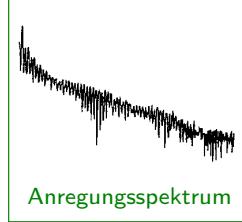
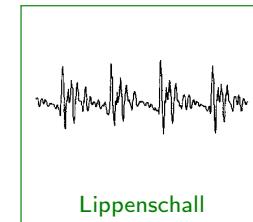
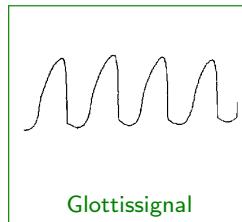
(mit symmetrischer Koeffizientenmatrix: *Autokorrelation!*) oder kompakter

$$\Psi^\top \mathbf{a} := \begin{pmatrix} \psi_{11} & \dots & \psi_{1p} \\ \vdots & \ddots & \vdots \\ \psi_{p1} & \dots & \psi_{pp} \end{pmatrix}^\top \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} -\psi_{01} \\ \vdots \\ -\psi_{0p} \end{pmatrix} =: -\mathbf{\psi}$$

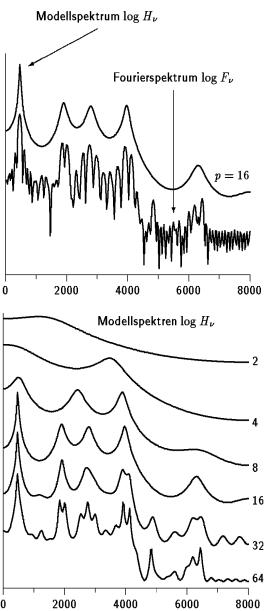
Source-Filter-Modell

Lineare Vorhersage $\hat{=}$ Autoregressives Erzeugungssystem

$$[e_n] = [f_n] * [a_n] \Rightarrow E(z) = F(z) \cdot A(z) \Rightarrow F(z) = E(z) \cdot \frac{1}{A(z)}$$



Modellspektrum der Linearen Vorhersage



Definition

Ist $[f_n]$ ein Sprachsignal mit den Vorhersagekoeffizienten $[a_k]$ der Ordnung p , so heißt

$$H(e^{i\omega}) = \frac{1}{A(e^{i\omega})} = \frac{1}{\sum_{k=0}^p a_k \cdot e^{-ik\omega}}$$

das **Modellspektrum** von $[f_n]$.

Konfigurationsparameter

Wie wählen wir die Modellordnung p ?
(Verschmierung \leftrightarrow Artefakte)

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

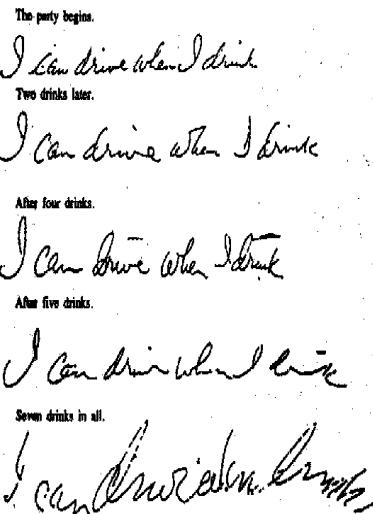
Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

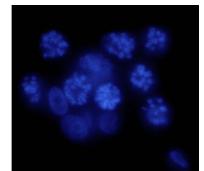
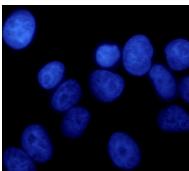
Dimension als Merkmal

Alkoholtest ohne Blutabnahme · Detektion von Tumorzellen



Fakt

Wenn er/sie noch sieben Gläser Vodka trinkt, füllt der Schriftzug eine kompakte Teilfläche der Papierebene aus ... !



- ⇒ Ausfüllungsgrad einer Punktmenge,
- ⇒ Filigranität einer Linie,
- ⇒ ... ?

Formfaktor

Maß für die „Filigranität“ zusammenhängender Flächen

Definition

Für eine einfach zusammenhängende Punktmenge $A \subset \mathbb{R}^2$ mit endlichem Flächeninhalt $\text{vol}_2(A)$ und endlichem Rand ∂A heißt der Quotient

$$c = \frac{(\text{Umfang})^2}{\text{Flächeninhalt}}$$

der **Formfaktor** von A .

Beispiele

$\frac{(2\pi r)^2}{\pi r^2} = 4\pi = 12.6$	$\frac{(4a)^2}{a^2} = 16$	$\frac{(2a(1+\lambda))^2}{\lambda \cdot a^2} = 4 \cdot \frac{(\lambda+1)^2}{\lambda}$	$\frac{(12a)^2}{5a^2} = 28.8$

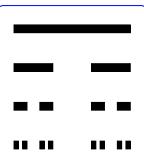
Geographische Objekte

Landmasse	Fläche [km²]	Küste [km]	Formfaktor
Japan	374,744	29,751	2361.9
Norwegen	307,442	25,148	2057.0
U.K.	241,590	12,429	639.4
Island	100,250	4,988	248.2
Australien	7,617,930	25,760	87.1
Spanien	499,542	4,964	49.3
Irland	68,890	1,448	30.4
Sri Lanka	64,740	1,340	27.7

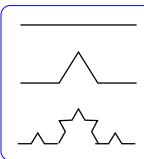
mehr Information

Selbstähnliche Mengen

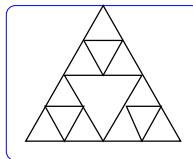
Kompliziertes Aussehen — einfaches induktives Bildungsgesetz



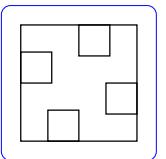
Cantormenge



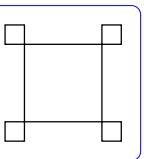
von Koch Kurve



Sierpinskidreieck



Cantorstaub



Falconermäander

Definition

Besteht die Punktmenge $A \subset \mathbb{R}^N$ aus genau m vielen, mit Faktor $0 < r < 1$ skalierten Kopien ihrer selbst, so heißt A **selbstähnlich**. Der Quotient

$$\dim_S A = \frac{\log m}{\log(\frac{1}{r})}$$

heißt (**Selbst-**)Ähnlichkeitsdimension von A .

Die Basis des Logarithmus' spielt keine Rolle.

Konstruktion selbstähnlicher Mengen als Grenzwert (?) einer unendlichen Copy+Paste-Sequenz

(Algorithmus) —————

- 1 Setze Startmenge E_0
- 2 Konstruiere E_{i+1} aus E_i :
 1. Skaliere E_i mit dem verkleinernden Faktor r , $0 < r < 1$.
 2. Erzeuge m Kopien der geschrumpften Menge.
 3. Wende spezifische Drehungen, Spiegelungen, Verschiebungen an.
- 3 Bilde den Grenzwert der Punkt mengenfolge

$$A \stackrel{\text{def}}{=} \lim_{i \rightarrow \infty} \bigcap_{j=i}^{\infty} E_j$$

A enthält genau jene Punkte x , die schließlich — d.h. von einem gewissen Index $i_0(x)$ ab — in allen E_i liegen.

Die Menge A gleicht i.a. weder der Vereinigung noch dem Durchschnitt der $\{E_i\}$.

Selbstähnlichkeitsdimension

Keine Kollision mit dem traditionellen Dimensionsbegriff

Das Potenzgesetz

$$m = \left(\frac{1}{r}\right)^{\dim_S A} = r^{-\dim_S A}$$

Linien, Quadrate, Kuben sind selbstähnlich und ihre s.ä. Dimension entspricht ihrer Standarddimension.

Beispiele (traditionell)

Linie	$\log m / \log m = 1$
Quadrat	$\log m^2 / \log m = 2$
Kubus	$\log m^3 / \log m = 3$

Beispiele (selbstähnlich)

Cantormenge	$\log 2 / \log 3 = 0.631$
von Kochkurve	$\log 4 / \log 3 = 1.262$
Sierpinskidreieck	$\log 3 / \log 2 = 1.585$
Cantorstaub	$\log 4 / \log 4 = 1$
Falconermenge	?

Problem

Wie finde ich die Konstruktionsparameter m , r einer selbstähnlichen Menge?!

Welche Rasterpunktmenge ist schon wirklich selbstähnlich?!

Grundprinzip der Dimensionsdefinition

Fakt

Für jedes n -dimensionale Volumenmaß $\mu_n : \mathfrak{P}(\mathbb{R}^n) \rightarrow \mathbb{R}$ mit der Eigenschaft

$$\mu_n \left(\bigcup_i A_i \right) = \sum_i \mu_n(A_i)$$

für disjunkte Mengensysteme (σ -Additivität) gilt die Skalierungseigenschaft

$$\mu_n(\lambda A) = \lambda^n \cdot \mu_n(A), \quad A \subset \mathbb{R}^n, \lambda \geq 0, \quad \lambda A = \{\lambda x \mid x \in A\}.$$

→ Dimensionsbestimmung aus dem Verlauf von $\lambda \mapsto \mu_n(\lambda A)$

Lemma

Für das s -dimensionale Hausdorff-Maß gilt die Skalierungseigenschaft

$$\mathfrak{h}_s(\lambda A) = \lambda^s \cdot \mathfrak{h}_s(A).$$

- $\mathfrak{h}_0(A) =$ Anzahl der Punkte in A
- $\mathfrak{h}_1(A) =$ Länge einer glatten Kurve
- $\mathfrak{h}_2(A) = 1/4 \cdot \pi$ mal Fläche von A
- $\mathfrak{h}_3(A) = 4/3 \cdot \pi$ mal Volumen von A

Hausdorff-Dimension

Denn es kann nur eine geben ...

Satz

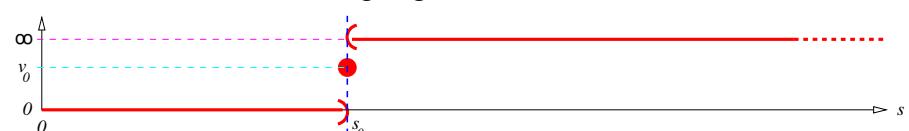
Es gibt für jede Punktmenge A genau ein $s_0 > 0$ mit der Eigenschaft

$$\mathfrak{h}_s(A) = \begin{cases} \infty & s < s_0 \\ v_0 & s = s_0 \\ 0 & s > s_0 \end{cases}$$

Dieses $s_0 \in \mathbb{R}$ heißt die **Hausdorff-Dimension** von A ($\dim_H A = s_0$); der Wert v_0 heißt **Hausdorff-Volumen** ($\text{vol}_H A = v_0$).

Bemerkung

Für alle selbstähnlichen Mengen gilt $\dim_H A = \dim_S A$.



Hausdorff-Maß

Gesamtvolumen immer kleinerer Reisetaschen, die gerade noch in den Kofferraum passen

Durchmesser von $U \subseteq \mathbb{R}^n$

$$|U| \stackrel{\text{def}}{=} \sup \{|x - y| \mid x, y \in U\}$$

δ -Überdeckung von A

$$A \subseteq \bigcup_{i=1}^{\infty} U_i \quad \text{und} \quad \forall i : 0 < |U_i| \leq \delta$$



Standard-Volumenelement

δ -körniges s -Volumen von A

$$\mathfrak{h}_s^\delta(A) \stackrel{\text{def}}{=} \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s \mid \{U_i\} \text{ ist } \delta\text{-Überdeckg.} \right\}$$



„Auslitern“ des Kofferraums

s -dimensionales Hausdorffmaß von A

$$\mathfrak{h}_s(A) \stackrel{\text{def}}{=} \lim_{\delta \rightarrow 0} \mathfrak{h}_s^\delta(A)$$

Kolmogorov-Entropien

Alternative Verfahren: Auslitern durch Auszählen

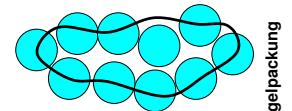
(a) $N_\delta^{(a)}(A) =$ minimale Anzahl von Mengen mit Durchmesser $\leq \delta$, die A überdecken

(b) $N_\delta^{(e)}(A) =$ maximale Anzahl disjunkter δ -Hyperkugeln mit Mittelpunkt in A

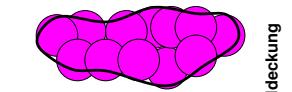
(c) $N_\delta^{(d)}(A) =$ minimale Anzahl von δ -Hyperkugeln, die A überdecken

(d) $N_\delta^{(c)}(A) =$ minimale Anzahl von δ -Hyperwürfeln, die A überdecken

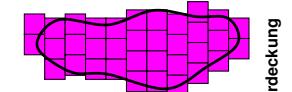
(e) $N_\delta^{(b)}(A) =$ Anzahl der δ -Netz Hyperwürfel, die A schneiden



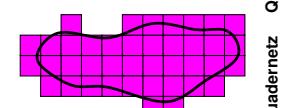
Kugelpackung



Kugeldeckung



Quaderdeckung



Quadernetz

Box-Dimension(en)

Alle Kolmogorov-Dimensionen sind äquivalent und $\geq \dim_H$

Definition

Für eine Punktmenge $A \subseteq \mathbb{R}^n$ und $\ell \in \{a, b, c, d, e\}$ definiere

$$\uparrow \dim_B^{(\ell)} A = \liminf_{\delta \rightarrow 0} \frac{\log N_\delta^{(\ell)}(A)}{-\log \delta}$$

$$\downarrow \dim_B^{(\ell)} A = \limsup_{\delta \rightarrow 0} \frac{\log N_\delta^{(\ell)}(A)}{-\log \delta}$$

sowie

$$\dim_B^{(\ell)} A \stackrel{\text{def}}{=} \uparrow \dim_B^{(\ell)} A, \quad \text{falls } \uparrow \dim_B^{(\ell)} A = \downarrow \dim_B^{(\ell)} A.$$

Lemma

Die Unter- und Ober-Boxdimensionen $\uparrow \dim_B^{(\ell)}$ bzw. $\downarrow \dim_B^{(\ell)}$ stimmen für alle $\ell \in \{a, b, c, d, e\}$ überein.

Lemma

Für jede Punktmenge $A \subseteq \mathbb{R}^n$ gilt der folgende Zusammenhang zwischen Hausdorff- und Boxdimensionen:

$$\dim_H A \leq \uparrow \dim_B A \leq \downarrow \dim_B A$$

Minkowski-Dimension

Definition

Für eine Punktmenge $A \subseteq \mathbb{R}^n$ und $\delta > 0$ heißt

$$A_\delta \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid |x - y| \leq \delta \text{ für ein } y \in A\}$$

die **δ -parallele Umgebung** von A .

Verhält sich das n -dimensionale Volumen $\text{vol}_n(A_\delta)$ asymptotisch gemäß

$$\text{vol}_n(A_\delta) \sim v \cdot \delta^{n-s},$$

so heißt s die **Minkowski-Dimension** und v der **Minkowski-Gehalt** von A .

Lemma

Für eine Punktmenge $A \subseteq \mathbb{R}^n$ gelten die Aussagen

$$\uparrow \dim_B A = n - \liminf_{\delta \rightarrow 0} \frac{\log \text{vol}_n(A_\delta)}{\log \delta}$$

$$\downarrow \dim_B A = n - \limsup_{\delta \rightarrow 0} \frac{\log \text{vol}_n(A_\delta)}{\log \delta}$$

Etappendimension

Fraktale Dimension für Kurven (Konturlinien)

Definition

Sei A das Bild der stetigen Bijektion $f : [a, b] \rightarrow \mathbb{R}^n$ und $M_\delta(A)$ die maximale Länge $M+1$ von Punktfolgen $x_0, \dots, x_M \in A$ mit Abstand $\|x_m - x_{m-1}\| = \delta$. Dann heißt

$$\dim_E A \stackrel{\text{def}}{=} \lim_{\delta \rightarrow 0} \frac{\log M_\delta(A)}{-\log \delta}$$

die **Etappendimension** der Kurve A bzw. der Abbildung f .

Etappenzahl M_δ einer Kurve x_1, \dots, x_T

Berechne für ausgewählte $\delta > 0$ die Haltepunkte

$$t_i = \begin{cases} 1 & i = 0 \\ \min \left\{ t > t_{i-1} \mid \|x_t - x_{t_{i-1}}\| > \delta \right\} & \text{falls ex.} \\ T & \text{sonst} \end{cases}$$



und daraus die Schrittzahl $M_\delta = \min \{i \mid t_i = T\}$

Die britische Küstenlinie

Etappendimension der britischen Westküste: $\dim_E \approx 1.25$



$$\delta = 200 \text{ km}$$



$$\delta = 100 \text{ km}$$

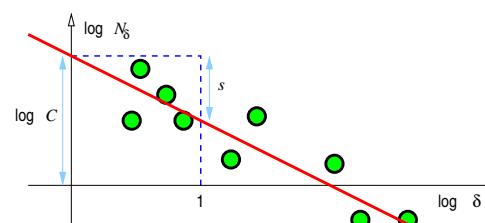


$$\delta = 50 \text{ km}$$

aus: Wikipedia (engl.)

Logarithmiertes Potenzgesetz

„Lächerlich grenzwertferne“ Extrapolation mittels Ausgleichsgerade



$$N_\delta \approx C \cdot \delta^{-s}$$

$$\log N_\delta \approx \log C - s \cdot \log \delta$$

(Algorithmus)

1 Berechne N_δ_i für ausgewählte Werte $\delta_i > 0$.

2 Plaziere die Punkte $(\log \delta_i, \log N_\delta_i)$.

3 Bestimme die Steigung der Ausgleichsgeraden.

Löse LGS $\begin{cases} m_{20} \cdot a + m_{10} \cdot b = m_{11} \\ m_{10} \cdot a + m_{00} \cdot b = m_{01} \end{cases}$ mit Momenten $m_{pq} \stackrel{\text{def}}{=} \sum_{i=1}^I x_i^p y_i^q$.

(Summlösung)

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

Ausliteration von Rasterbildern

Fläche der δ -parallelen Hülle eines Binärbildes f

Für $\delta = 1, 2, \dots$ berechne das dilatierte Bild

$$f^{(\delta)} = f \oplus s_\delta$$

sowie seine Fläche

$$\text{vol}_2(f^{(\delta)}) = \sum_{n,m} f_{nm}^{(\delta)}$$

Fläche der δ -Rasterung von f

$$U_{nm}^{(\delta)} \stackrel{\text{def}}{=} [n\delta, (n+1)\delta] \times [m\delta, (m+1)\delta]$$

ist die Originalrasterpunktmenge des Bildpunktes (n, m) im δ -vergrößerten Bild $f^{(\delta)}$. Setze nun

$$f_{nm}^{(\delta)} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{falls ex. } (\nu, \mu) \in U_{nm}^{(\delta)} \text{ mit } f_{\nu\mu} = 1 \\ 0 & \text{sonst} \end{cases}$$

und

$$\text{vol}_2(f^{(\delta)}) = \sum_{n,m} f_{nm}^{(\delta)}$$

Analytische Methoden zur Merkmalsgewinnung

Systematische Berechnung der $n < N$ besten Merkmale x_1, \dots, x_n

$$f \mapsto \Phi^\top f, \quad \Phi = (\phi_1, \dots, \phi_n)$$

Optimierung eines Gütekriteriums

- Maximale Streuung der Merkmale für Muster $f \in \Omega$
- Kompaktheit der Merkmalsgebiete von Mustern einer Klasse
- Trennung der Merkmalsgebiete von Mustern verschiedener Klassen
- Kompatibilität mit einem statistischen Verteilungsmodell
- Korrelation mit der wahren Klassenzugehörigkeit
- Minimale Fehlklassifikationsrate

... bezogen auf eine Lernstichprobe von (Beispiel-)Mustern

Etikettierte Stichprobe

Definition

Es sei Ω ein Problemkreis mit Mustern aus K Klassen. Eine Teilmenge $\omega \subseteq \Omega$ heißt **Stichprobe** aus Ω . Ein System

$$\{\omega_\kappa \mid \kappa = 1, \dots, K\}$$

von Mengen $\omega_\kappa \subseteq \Omega$ von Mustern der Klasse κ heißt **klassifizierte** oder **etikettierte Stichprobe** aus Ω .

Wir verwenden dieselben Bezeichnungen auch für die Ergebnisse

$$\omega^T \quad \text{bzw.} \quad \{\omega_\kappa^T \mid \kappa = 1, \dots, K\}$$

der Anwendung einer Transformation $\mathfrak{T} : \Omega \rightarrow \mathbb{R}^n$ von Mustern in den Merkmalraum \mathbb{R}^n .

Vier Gütekriterien (1)

Kompaktheit und Separiertheit der transformierten Daten

K1 — maximaler wechselseitiger Abstand \rightsquigarrow PCA

aller Merkmalvektoren einer Stichprobe $\omega \subseteq \mathbb{R}^n$ von allen anderen

$$s_1 = \frac{1}{|\omega|^2} \underbrace{\sum_{x \in \omega} \sum_{y \in \omega} (x - y)^\top (x - y)}_{d(\omega, \omega)}$$

K2 — maximale Außerklassenabstände

Mittlerer quadratischer Abstand aller Merkmalvektoren einer Klasse Ω_κ von denen einer anderen Klasse Ω_λ

$$s_2 = \frac{2}{K(K-1)} \underbrace{\sum_{\kappa=1}^K \sum_{\lambda < \kappa} \frac{1}{|\omega_\kappa| \cdot |\omega_\lambda|} \sum_{x \in \omega_\kappa} \sum_{y \in \omega_\lambda} (x - y)^\top (x - y)}_{d(\omega_\kappa, \omega_\lambda)}$$

Vier Gütekriterien (2)

Kompaktheit und Separiertheit der transformierten Daten

K3 — minimale Innerklassenabstände

Mittlerer quadratischer Abstand aller Merkmalvektoren einer Stichprobe $\{\omega_\kappa\}$ von Merkmalvektoren der *eigenen* Klasse

$$s_3 = \frac{1}{K} \sum_{\kappa=1}^K \underbrace{\frac{1}{|\omega_\kappa|^2} \sum_{x \in \omega_\kappa} \sum_{y \in \omega_\kappa} (x - y)^\top (x - y)}_{d(\omega_\kappa, \omega_\kappa)}$$

K4 — kombinierte Optimierung von K2 und K3 \rightsquigarrow LDA

$$s_4 = s_2 + \theta \cdot s_3 \quad (\theta = \text{Lagrange-Multiplikator})$$

Stichprobenstatistiken

Empirische Mittelwertvektoren und Kovarianzmatrizen

Definition

Für eine (endliche) Stichprobe $\omega = \{x^{(1)}, \dots, x^{(T)}\}$ bezeichne $|\omega| = T$ die **Mächtigkeit** (den Umfang) der Stichprobe,

$$\mu = \mu(\omega) = \frac{1}{T} \cdot \sum_{x \in \omega} x$$

den (empirischen) **Mittelwertvektor** und

$$S = S(\omega) = \frac{1}{T} \cdot \sum_{x \in \omega} (x - \mu)(x - \mu)^\top = \underbrace{\frac{1}{T} \cdot \sum_{x \in \omega} xx^\top - \mu\mu^\top}_{R(\omega)}$$

die (empirische) **Kovarianzmatrix** (bzw. **Momentenmatrix** $R(\omega)$).

Bemerkung

Für die Mittelwertvektoren, Momenten- und Kovarianzmatrizen der klassenbezogenen Komponenten ω_κ einer etikettierten Stichprobe verwenden wir auch die Kurzbezeichnungen μ_κ , R_κ und S_κ .

Momente und Statistiken abgeleiteter Stichproben

$$\begin{aligned} m_0 &= \sum_x 1 \quad \text{und} \quad \mathbf{m}_1 = \sum_x \mathbf{x} \quad \text{und} \quad \mathbf{M}_2 = \sum_x \mathbf{x} \mathbf{x}^\top \\ \boldsymbol{\mu} &= \frac{\mathbf{m}_1}{m_0} \quad \text{und} \quad \mathbf{R} = \frac{\mathbf{M}_2}{m_0} \quad \text{und} \quad \mathbf{S} = \mathbf{R} - \boldsymbol{\mu} \boldsymbol{\mu}^\top \end{aligned}$$

Probe	Statistiken			Momente		
	ω'	m'_0	\mathbf{m}'_1	\mathbf{M}'_2	μ'	\mathbf{R}'
$\alpha \cdot \omega$	m_0	$\alpha \mathbf{m}_1$	$\alpha^2 \mathbf{M}_2$	$\alpha \boldsymbol{\mu}$	$\alpha^2 \mathbf{R}$	$\alpha^2 \mathbf{S}$
$-\omega$	m_0	$-\mathbf{m}_1$	\mathbf{M}_2	$-\boldsymbol{\mu}$	\mathbf{R}	\mathbf{S}
$\omega - \mu$	m_0	$\mathbf{0}$	$\mathbf{M}_2 - m_0 \boldsymbol{\mu} \boldsymbol{\mu}^\top$	$\mathbf{0}$	\mathbf{S}	\mathbf{S}
$\mathbf{A} \cdot \omega$	m_0	$\mathbf{A} \mathbf{m}_1$	$\mathbf{A} \mathbf{M}_2 \mathbf{A}^\top$	$\mathbf{A} \boldsymbol{\mu}$	$\mathbf{A} \mathbf{R} \mathbf{A}^\top$	$\mathbf{A} \mathbf{S} \mathbf{A}^\top$
$\omega \cup \omega$	$2m_0$	$2\mathbf{m}_1$	$2\mathbf{M}_2$	$\boldsymbol{\mu}$	\mathbf{R}	\mathbf{S}
$\omega \cup -\omega$	$2m_0$	$\mathbf{0}$	$2\mathbf{M}_2$	$\mathbf{0}$	\mathbf{R}	\mathbf{R}

Zwei Rechenregeln für symmetrische Matrizen

(Die Spur einer quadratischen Matrix ist die Summe der Diagonalelemente)

Lemma

- Ist die Matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ symmetrisch, so gilt für alle Vektoren $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$:

$$\underbrace{\mathbf{u}^\top}_{\mathbf{u}^\top} \cdot \underbrace{\mathbf{A}}_{\mathbf{A}} \cdot \underbrace{\mathbf{v}}_{\mathbf{v}} = \text{spur} \left(\underbrace{\mathbf{A}}_{\mathbf{A}} \cdot \underbrace{\mathbf{u} \mathbf{v}^\top}_{\mathbf{u} \mathbf{v}^\top} \right)$$

- Ist die Matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ symmetrisch, und ist $\Phi = (\phi_1, \dots, \phi_n) \in \mathbb{R}^{N \times n}$, so gilt:

$$\text{spur} \left(\underbrace{\Phi}_{\Phi} \cdot \underbrace{\Phi^\top}_{\Phi^\top} \cdot \underbrace{\mathbf{A}}_{\mathbf{A}} \right) = \sum_{\nu=1}^n \underbrace{\phi_\nu^\top}_{\phi_\nu^\top} \cdot \underbrace{\mathbf{A}}_{\mathbf{A}} \cdot \underbrace{\phi_\nu}_{\phi_\nu}$$

Maximierung des Varianzkriteriums (K1)

Lemma

Ist $\Phi = (\phi_1, \dots, \phi_n) \in \mathbb{R}^{N \times n}$ eine orthonormale Transformationsmatrix ($\Phi^\top \Phi = \mathbf{E}$) und sind ω' und $\{\omega_\kappa'\}$ die (un)klassifizierten Merkmalstichproben zu ω und $\{\omega_\kappa\}$ hinsichtlich der Transformation

$\mathbf{c} = \Phi^\top \mathbf{f}$, so besitzen das Kompaktheitskriterium K1 — bezogen auf die transformierten Muster — die Gestalt

$$s_1 = 2 \cdot \sum_{\nu=1}^n \phi_\nu^\top \mathbf{Q}^{(1)} \phi_\nu$$

mit der symmetrischen, positiv-definiten **Kernmatrix** $\mathbf{Q}^{(1)} = \mathbf{S}(\omega)$.

Bemerkung

Auch für die Kriterien K2, K3 findet sich eine solche Darstellung als **Summe quadratischer Formen**, und zwar mit den Kernmatrizen

$$\mathbf{Q}^{(2)} \stackrel{\text{def}}{=} \frac{1}{K} \sum_{\kappa=1}^K \mathbf{R}(\omega_\kappa) - \frac{1}{K(K-1)} \sum_{\kappa \neq \lambda} \boldsymbol{\mu}(\omega_\kappa) \boldsymbol{\mu}(\omega_\lambda)^\top \quad \text{bzw.} \quad \mathbf{Q}^{(3)} \stackrel{\text{def}}{=} \frac{1}{K} \sum_{\kappa=1}^K \mathbf{S}(\omega_\kappa)$$

Beweis.

Wir reformulieren s_1 für transformierte orthonormal transformierte Merkmalvektoren $\mathbf{x}' = \Phi^\top \mathbf{x}$:

$$\begin{aligned} s_1 &= \frac{1}{|\omega'|^2} \sum_{\mathbf{x}' \in \omega'} \sum_{\mathbf{y}' \in \omega'} (\mathbf{x}' - \mathbf{y}')^\top (\mathbf{x}' - \mathbf{y}') = \frac{1}{|\omega|^2} \sum_{\mathbf{x} \in \omega} \sum_{\mathbf{y} \in \omega} (\Phi^\top \mathbf{x} - \Phi^\top \mathbf{y})^\top (\Phi^\top \mathbf{x} - \Phi^\top \mathbf{y}) \\ &= \frac{1}{|\omega|^2} \sum_{\mathbf{x} \in \omega} \sum_{\mathbf{y} \in \omega} (\mathbf{x} - \mathbf{y})^\top \Phi \Phi^\top (\mathbf{x} - \mathbf{y}) \\ &= \frac{1}{|\omega|^2} \sum_{\mathbf{x} \in \omega} \sum_{\mathbf{y} \in \omega} \text{spur} \left(\Phi \Phi^\top (\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^\top \right) \end{aligned}$$

denn die Matrix $\Phi \Phi^\top$ ist offensichtlich symmetrisch.

$$\begin{aligned} s_1 &= \text{spur} \left(\Phi \Phi^\top \cdot \frac{1}{|\omega|^2} \sum_{\mathbf{x} \in \omega} \sum_{\mathbf{y} \in \omega} (\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^\top \right) \\ &= \text{spur} \left(\Phi \Phi^\top \cdot \left\{ \frac{1}{|\omega|^2} \sum_{\mathbf{x}, \mathbf{y}} \mathbf{x} \mathbf{x}^\top - \frac{2}{|\omega|^2} \sum_{\mathbf{x}, \mathbf{y}} \mathbf{x} \mathbf{y}^\top + \frac{1}{|\omega|^2} \sum_{\mathbf{x}, \mathbf{y}} \mathbf{y} \mathbf{y}^\top \right\} \right) \\ &= \text{spur} \left(\Phi \Phi^\top \cdot \left\{ \frac{1}{|\omega|} \sum_{\mathbf{x}} \mathbf{x} \mathbf{x}^\top - 2 \frac{1}{|\omega|} \sum_{\mathbf{x}} \mathbf{x} \cdot \frac{1}{|\omega|} \sum_{\mathbf{y}} \mathbf{y}^\top + \frac{1}{|\omega|} \sum_{\mathbf{y}} \mathbf{y} \mathbf{y}^\top \right\} \right) \\ &= \text{spur} \left(\Phi \Phi^\top \cdot \left\{ \mathbf{R}(\omega) - 2\boldsymbol{\mu}(\omega) \boldsymbol{\mu}(\omega)^\top + \mathbf{R}(\omega) \right\} \right) \\ &= 2 \cdot \text{spur} \left(\Phi \Phi^\top \cdot \mathbf{S}(\omega) \right) = 2 \cdot \text{spur} \left(\Phi \Phi^\top \cdot \mathbf{Q}^{(1)} \right) = 2 \cdot \sum_{\nu=1}^n \phi_\nu^\top \mathbf{Q}^{(1)} \phi_\nu \end{aligned}$$

Karhunen-Loéve-Transformation (KLT)

Definition

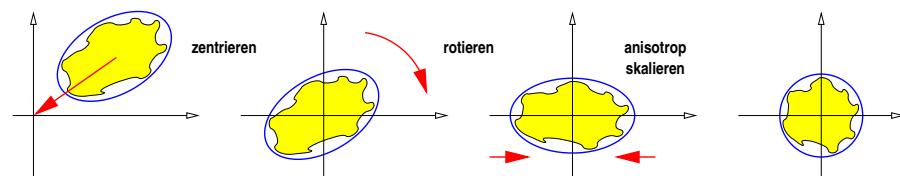
Sind μ und \mathbf{S} Mittelwertvektor und Kovarianzmatrix der Muster einer Stichprobe $\omega \subseteq \mathbb{R}^N$, und ist $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ die Matrix der zu den nach absteigender Größe sortierten Eigenwerten $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_N)$ gehörenden Eigenvektoren von \mathbf{S} , so heißt die Abbildung

$$\mathcal{T}_{\text{KL}} : \mathbf{f} \mapsto \mathbf{U}^\top \mathbf{f}$$

die **Karhunen-Loéve-Transformation** von ω . Wir bezeichnen ferner

$$\mathbf{f} \mapsto \mathbf{U}^\top(\mathbf{f} - \mu) \quad \text{und} \quad \mathbf{f} \mapsto (\mathbf{U}\mathbf{D}^{-1/2})^\top(\mathbf{f} - \mu)$$

als **zentrierte** bzw. **normierte** Karhunen-Loéve-Transformation.



Beweis.

- Die Kovarianzmatrix \mathbf{S} einer Datenprobe oder Verteilung wird nach linearer Transformation mit der Abbildungsmatrix Φ bekanntlich zu $\mathbf{S}' = \Phi \mathbf{S} \Phi^\top$. Für eine KLT gilt:

$$\mathbf{S}' = (\mathbf{U}^\top) \cdot (\mathbf{U} \mathbf{D} \mathbf{U}^\top) \cdot (\mathbf{U}^\top)^\top = \mathbf{U}^\top \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^\top \mathbf{U} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{E} = \mathbf{D}$$

Für eine normierte KLT gilt sogar:

$$\mathbf{S}' = (\mathbf{U}\mathbf{D}^{-1/2})^\top \cdot (\mathbf{U}\mathbf{D}\mathbf{U}^\top) \cdot (\mathbf{U}\mathbf{D}^{-1/2}) = \mathbf{D}^{-1/2} \cdot \mathbf{U}^\top \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^\top \mathbf{U} \cdot \mathbf{D}^{-1/2} = \mathbf{D}^{-1/2} \cdot \mathbf{D} \cdot \mathbf{D}^{-1/2} = \mathbf{E}$$

- Wir entwickeln die Muster \mathbf{f} der Probe nach der ONB $\{\phi_1, \dots, \phi_N\}$ und verwenden die verzerrungsminimalen Koeffizienten $c_i = \phi_i^\top \mathbf{f}$ dafür; der Fehler ist:

$$\begin{aligned} \varepsilon(\omega) &= \sum_{\mathbf{f} \in \omega} \|\mathbf{f} - \tilde{\mathbf{f}}\|^2 = \sum_{\mathbf{f} \in \omega} \left\| \sum_{i=n+1}^N c_i \phi_i \right\|^2 = \sum_{\mathbf{f} \in \omega} \left(\sum_{i=n+1}^N c_i \phi_i \right)^\top \left(\sum_{j=n+1}^N c_j \phi_j \right) \\ &= \sum_{\mathbf{f} \in \omega} \sum_{i=n+1}^N \sum_{j=n+1}^N c_i c_j \cdot \phi_i^\top \phi_j = \sum_{\mathbf{f} \in \omega} \sum_{i=n+1}^N c_i^2 = \sum_{\mathbf{f} \in \omega} \sum_{i=n+1}^N (\phi_i^\top \mathbf{f})(\phi_i^\top \mathbf{f}) \\ &= \sum_{\mathbf{f} \in \omega} \sum_{i=n+1}^N \phi_i^\top \mathbf{f} \mathbf{f}^\top \phi_i = \sum_{i=n+1}^N \phi_i^\top \left(\sum_{\mathbf{f} \in \omega} \mathbf{f} \mathbf{f}^\top \right) \phi_i = |\omega| \sum_{i=n+1}^N \phi_i^\top R(\omega) \phi_i \end{aligned}$$

Wegen der Mittelwertfreiheit von ω gilt $R(\omega) = \mathbf{S}(\omega)$, und wir haben $K1 + \varepsilon(\omega) = C$ hergeleitet.

- Maximierung der Varianz bzw. Minimierung des Rekonstruktionsfehlers führt also auf dieselbe Entwicklung, und zwar — dem Satz über Summen quadratischer Formen — unter Verwendung der Eigenvektoren zu den n größten Eigenwerten von $\mathbf{S}(\omega)$.

Hauptachsentransformation

PCA — Principal Component Analysis

Lemma

- Die oben definierte Karhunen-Loéve-Transformation bildet Mustervektoren auf dekorrelierte Merkmalvektoren ab, d.h., die Kovarianzmatrix wird zur Diagonalmatrix (oder zur Einheitsmatrix bei der normierten KLT).
- Die **unvollständige Reihenentwicklung**

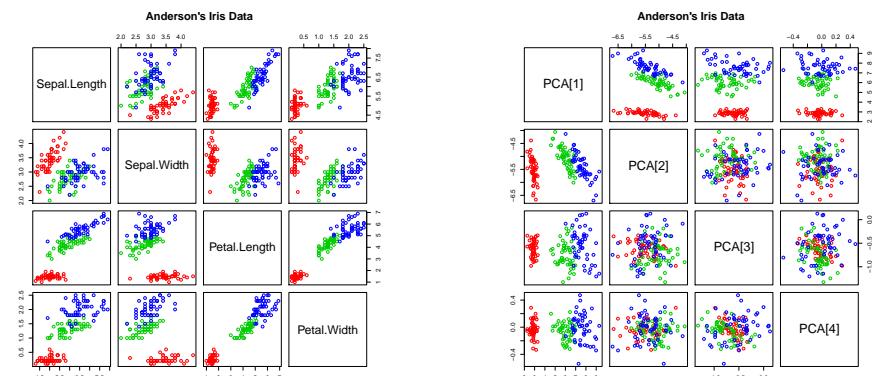
$$\mathcal{T}_{\text{PCA}}^{(n)} : \mathbf{f} \mapsto (\mathbf{u}_1, \dots, \mathbf{u}_n)^\top \mathbf{f}$$

erzielt unter allen orthonormalen Reihenentwicklungen dieser Ordnung den niedrigsten mittleren quadratischen Rekonstruktionsfehler für die Muster einer (mittelwertfreien) Stichprobe ...

- ... und die Gesamtvarianz der resultierenden Merkmalvektoren ist maximal (Varianzkriterium K1).

Beispiel: der IRIS-Datensatz & eine glückliche PCA

4 Originalmerkmale · 3 Klassen · 150 Objekte



Kovarianzmatrix \mathbf{S}

$$\begin{pmatrix} 0.685 & -0.042 & 1.274 & 0.516 \\ -0.042 & 0.189 & -0.329 & -0.121 \\ 1.274 & -0.329 & 3.116 & 1.295 \\ 0.516 & -0.121 & 1.295 & 0.581 \end{pmatrix}$$

Eigenwerte

$$\begin{pmatrix} \lambda_1 = 4.2282 \\ \lambda_2 = 0.2426 \\ \lambda_3 = 0.0782 \\ \lambda_4 = 0.0238 \end{pmatrix} \begin{pmatrix} 0.361 & -0.656 & -0.582 & 0.315 \\ -0.084 & -0.730 & 0.597 & -0.319 \\ 0.856 & 0.173 & 0.076 & -0.479 \\ 0.358 & 0.075 & 0.545 & 0.753 \end{pmatrix}$$

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

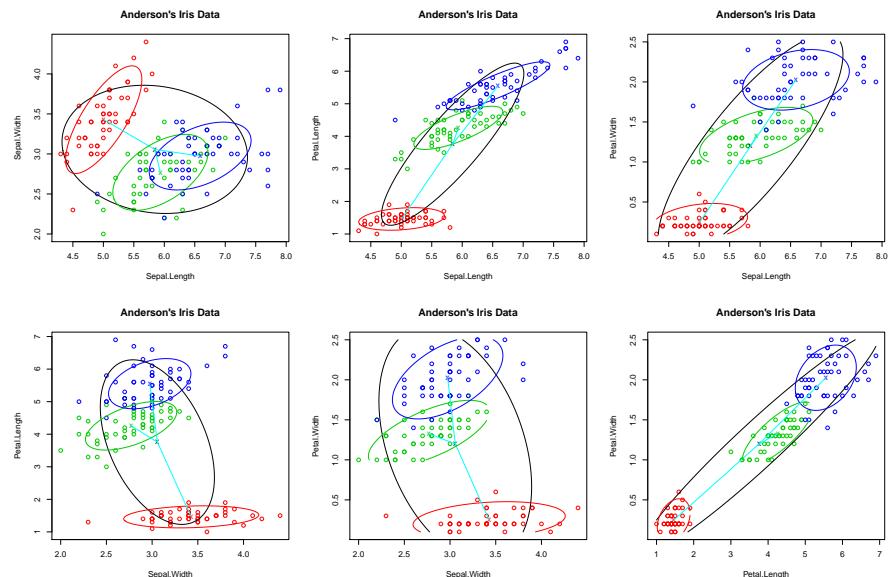
Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

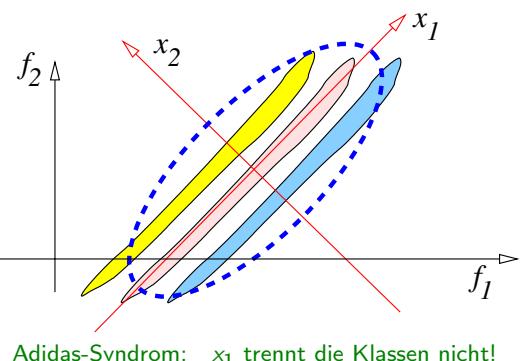
Kovarianzzerlegung

Gesamtstreuung & klassenweise Streuung für alle IRIS-Koordinatenpaare



Varianz \neq Klassentrennschärfe

Die Hauptachsen der PCA repräsentieren nicht notwendig die beste aller Datenansichten!



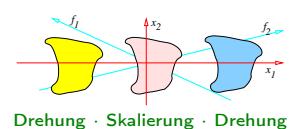
Problem

Die PCA „weiß nichts“ von den Klassenzugehörigkeiten der Datenobjekte.

Lösung

Maximiere **Fisher-Diskriminanz** ($K4$)

1. Ballungsgebiete kompakt halten
2. Klassenzentren auf spreizen



Drehung · Skalierung · Drehung

Kovarianzzerlegung

Gesamtstreuung = Innerklassenstreuung + Zwischenklassenstreuung

Definition

Ist $\{\omega_\kappa\}$ eine klassifizierte Stichprobe, so heißt

$$\mathbf{S}_W = \sum_{\kappa=1}^K \frac{|\omega_\kappa|}{|\omega|} \cdot \mathbf{S}_\kappa = \frac{1}{|\omega|} \sum_{\kappa=1}^K \sum_{x \in \omega_\kappa} (\mathbf{x} - \boldsymbol{\mu}_\kappa)(\mathbf{x} - \boldsymbol{\mu}_\kappa)^\top$$

die (empirische) **Intraklassenstreuungsmatrix** ('within-class scatter') und

$$\mathbf{S}_B = \sum_{\kappa=1}^K \frac{|\omega_\kappa|}{|\omega|} \cdot (\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})(\boldsymbol{\mu}_\kappa - \boldsymbol{\mu})^\top$$

die (empirische) **Interklassenstreuungsmatrix** ('between-class scatter'); dabei bezeichne $\boldsymbol{\mu}_\kappa$ den κ -ten klassenbezogenen und $\boldsymbol{\mu}$ den globalen Mittelwertvektor der Stichprobe.

Lemma

Für die Gesamtstreuungsmatrix $\mathbf{S} = \mathbf{S}(\omega)$ der Stichprobe $\{\omega_\kappa\}$ gilt die Aussage

$$\mathbf{S} = \mathbf{S}_W + \mathbf{S}_B$$

Kitano-Transformation

Analytische Reihenentwicklung zur optimalen Klassentrennung

1 Drehung & Anisotrope Skalierung

Transformation $\mathbf{x} \mapsto \tilde{\mathbf{x}} = \Phi_1^\top \mathbf{x}$ mit

$$\Phi_1 \stackrel{\text{def}}{=} \mathbf{U} \mathbf{D}^{-1/2} \quad \text{und} \quad \tilde{\mathbf{S}}_W \mathbf{U} \stackrel{!}{=} \mathbf{U} \mathbf{D}$$

$\tilde{\mathbf{S}}_W$ ist die Innerklassenstreuung der Originaldaten.

2 Drehung & Achsenauswahl

Transformation $\tilde{\mathbf{x}} \mapsto \hat{\mathbf{x}} = \Phi_2^\top \tilde{\mathbf{x}}$ mit

$$\Phi_2 \stackrel{\text{def}}{=} \mathbf{V} \quad \text{und} \quad \tilde{\mathbf{S}}_B \mathbf{V} \stackrel{!}{=} \mathbf{V} \mathbf{C}$$

$\tilde{\mathbf{S}}_B$ ist die Zwischenklassenstreuung der Φ_1 -transformierten Daten.

(Algorithmus)

Kitano-Transformation

Sehr erfolgversprechend — wenn alle Klassengebiete „ähnlich“ sind

Lemma

1. Die erste Transformation Φ_1 von Kitano macht die Innerklassenstreuung der Daten zur Einheitsmatrix.
2. Sind die Kovarianzmatrizen aller Klassen ähnlich ($\mathbf{S}_\kappa \propto \mathbf{S}_0$), so macht die Kitano-Transformation alle Klassengebiete **sphärisch**.
3. Die Gesamttransformation $\Phi \stackrel{\text{def}}{=} \Phi_1 \Phi_2$ von Kitano und die Diagonalmatrix \mathbf{C} des obigen Algorithmus lösen die (nicht notwendig symmetrische!) Eigenwertaufgabe

$$(\mathbf{S}_W^{-1} \mathbf{S}_B) \cdot \mathbf{U} \stackrel{!}{=} \mathbf{U} \mathbf{D} .$$

(zumthmogIA)

Beweis.

1. Für die Innerklassenstreuung der transformierten Stichprobe gilt:

$$\begin{aligned} \tilde{\mathbf{S}}_W &= \Phi_1^\top \mathbf{S}_W \Phi_1 = (\mathbf{U} \mathbf{D}^{-1/2})^\top \cdot (\mathbf{U} \mathbf{D} \mathbf{U}^\top) \cdot (\mathbf{U} \mathbf{D}^{-1/2}) \\ &= \mathbf{D}^{-1/2} \cdot \mathbf{U}^\top \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^\top \mathbf{U} \cdot \mathbf{D}^{-1/2} = \mathbf{D}^{-1/2} \mathbf{D} \mathbf{D}^{-1/2} = \mathbf{E} \end{aligned}$$

2. Offenbar gilt dann auch $\mathbf{S}_\kappa \propto \mathbf{S}_0 \propto \mathbf{S}_W$, und die obige Herleitung gilt auch für alle \mathbf{S}_κ . Die zweite Komponente Φ_2 ist eine Drehung und lässt die Sphären intakt.

3. Für die fragliche Kernmatrix $\mathbf{S}_W^{-1} \mathbf{S}_B$ und Kitanos Φ gilt:

$$\begin{aligned} \mathbf{S}_W^{-1} \mathbf{S}_B \Phi &= (\mathbf{U} \mathbf{D}^{-1/2} \mathbf{U}^\top) \mathbf{S}_B \underbrace{(\mathbf{U} \mathbf{D}^{-1/2})}_{\Phi_1} \underbrace{\mathbf{V}}_{\Phi_2} \\ &= (\mathbf{U} \mathbf{D}^{-1/2}) (\mathbf{U} \mathbf{D}^{-1/2})^\top \mathbf{S}_B (\mathbf{U} \mathbf{D}^{-1/2}) \mathbf{V} \\ &= (\mathbf{U} \mathbf{D}^{-1/2}) \tilde{\mathbf{S}}_B \mathbf{V} = \underbrace{(\mathbf{U} \mathbf{D}^{-1/2})}_{\Phi_1} \underbrace{\mathbf{V}}_{\Phi_2} \mathbf{C} \\ &\quad \underbrace{\mathbf{C}}_{\Phi} \end{aligned}$$

Folglich ist Φ tatsächlich Lösung der EWA zur Matrix $\mathbf{S}_W^{-1} \mathbf{S}_B$.

□

Bemerkungen

1. Der Rechenaufwand für $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ ist unter Umstand viel geringer als für

$$\Phi_1 \circ \Phi_2 : \mathbb{R}^N \rightarrow \mathbb{R}^N \rightarrow \mathbb{R}^n$$

2. Leider gibt es keine schnellen Algorithmen zur *sukzessiven* Berechnung der Eigenvektoren **nicht-symmetrischer** Matrizen wie $\mathbf{S}_W^{-1} \mathbf{S}_B$.

3. Die Interklassenstreuung \mathbf{S}_B hat einen Rang $\leq K - 1$. Dasselbe gilt auch für $\mathbf{S}_W^{-1} \mathbf{S}_B$, also gibt es *maximal* ($K-1$) Eigenwerte ungleich Null.

4. Nach Wilks (1962) führt auch das Optimierungsproblem $\text{spur}(\Phi^\top \mathbf{S} \Phi) \rightarrow \max$ mit der Nebenbedingung

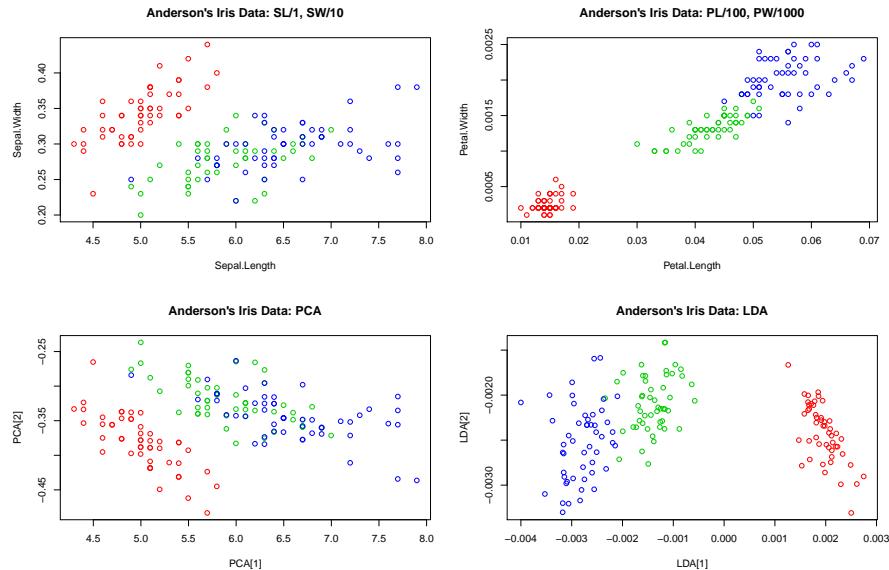
$$\text{spur}(\Phi^\top \mathbf{S}_W \Phi) = \text{konstant}$$

auf dasselbe Orthonormalsystem.

5. Das Wilkssche Optimierungsproblem entspricht der Maximierung des Kompaktheitskriteriums K4, sofern alle klassenbezogenen Stichproben ω_κ denselben Umfang besitzen.

Beispiel: der IRIS-Datensatz — PCA vs. LDA

4 Originalmerkmale · 3 Klassen · 150 Objekte

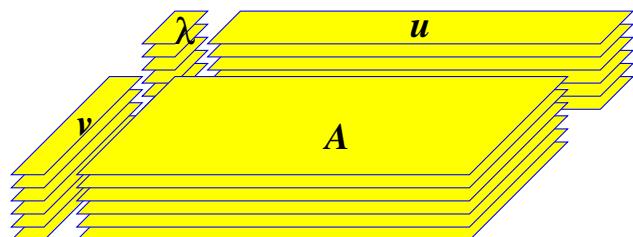


Willkürliche umskalierter IRIS-Datensatz

Reihenentwicklung für Matrizen

Geht das auch für *rechteckige* Matrizen („Bilder“) ?

$$\mathbf{A} \approx \mathbf{A}^{(r)} \stackrel{\text{def}}{=} \sum_{i=1}^r c_i \cdot \mathbf{v}_i \mathbf{u}_i^\top$$



Lemma

Sei $\mathbf{A} \in \mathbb{R}^{M \times N}$ eine reellwertige Rechteckmatrix. Die (symmetrischen quadratischen) Matrizen $\mathbf{A}^\top \mathbf{A}$ und $\mathbf{A} \mathbf{A}^\top$ besitzen genau dieselben Eigenwerte $\{\lambda_i\}$.

Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

Beweis.

Es sei $\mathbf{u} \in \mathbb{R}^N$ ein Eigenvektor von $\mathbf{A}^\top \mathbf{A}$, d.h. es gelte

$$\mathbf{A}^\top \mathbf{A} \mathbf{u} = \lambda \mathbf{u}$$

Betrachte den Vektor $\mathbf{v} = \mathbf{A}\mathbf{u} \in \mathbb{R}^M$:

$$\mathbf{A}\mathbf{A}^\top \mathbf{v} = \mathbf{A}\mathbf{A}^\top \mathbf{A}\mathbf{u} = \mathbf{A}\lambda\mathbf{u} = \lambda\mathbf{A}\mathbf{u} = \lambda\mathbf{v}$$

Also ist \mathbf{v} ein Eigenvektor von $\mathbf{A}\mathbf{A}^\top$.

Es sei umgekehrt $\mathbf{v} \in \mathbb{R}^M$ ein Eigenvektor von $\mathbf{A}\mathbf{A}^\top$, d.h. es gelte

$$\mathbf{A}\mathbf{A}^\top \mathbf{v} = \lambda \mathbf{v}$$

Betrachte den Vektor $\mathbf{u} = \mathbf{A}^\top \mathbf{v} \in \mathbb{R}^N$:

$$\mathbf{A}^\top \mathbf{A} \mathbf{u} = \mathbf{A}^\top \mathbf{A} \mathbf{A}^\top \mathbf{v} = \mathbf{A}^\top \lambda \mathbf{v} = \lambda \mathbf{A}^\top \mathbf{v} = \lambda \mathbf{u}$$

Also ist \mathbf{u} ein Eigenvektor von $\mathbf{A}^\top \mathbf{A}$. □

Beweis.

Wir nehmen der Einfachheit halber an, daß alle Singulärwerte verschieden sind. Dann gibt es zu jedem (gemeinsamen)

Eigenwert $\lambda = d^2$ von $\mathbf{A}^\top \mathbf{A}$ und $\mathbf{A} \mathbf{A}^\top$ je genau einen Eigenvektor \mathbf{u} und \mathbf{v} , wenn man von der Skalierung einmal absieht. Da wir bereits oben gezeigt haben, daß auch $\mathbf{A}\mathbf{u}$ ein Eigenvektor von $\mathbf{A}\mathbf{A}^\top$ zum Eigenwert λ ist, gilt jedenfalls

$$C \cdot \mathbf{v} = \mathbf{A}\mathbf{u}$$

für eine geeignete Konstante $C \in \mathbb{R}$.

Weil \mathbf{u} normiert ist gilt

$$\begin{aligned} \|\mathbf{A}\mathbf{u}\|^2 &= (\mathbf{A}\mathbf{u})^\top (\mathbf{A}\mathbf{u}) = \mathbf{u}^\top \mathbf{A}^\top \mathbf{A}\mathbf{u} \\ &= \mathbf{u}^\top (\mathbf{A}^\top \mathbf{A}\mathbf{u}) = \mathbf{u}^\top d^2 \mathbf{u} \\ &= d^2 \mathbf{u}^\top \mathbf{u} = d^2 \end{aligned}$$

d.h. es ist $\|\mathbf{A}\mathbf{u}\| = d$.

Weil auch \mathbf{v} normiert ist, folgt unmittelbar

$$|C| = |C| \|\mathbf{v}\| = \|C\mathbf{v}\| = \|\mathbf{A}\mathbf{u}\| = d$$

□

Singulärwerte und Singulärvektoren

Rechte und linke Singulärvektoren verhalten sich dual

Definition

Sei $\mathbf{A} \in \mathbb{R}^{M \times N}$ eine reellwertige Rechteckmatrix. Die Vektoren $\mathbf{u}_i \in \mathbb{R}^N$ mit

$$\mathbf{A}^\top \mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

heißen **Rechtssingulärvektoren** von \mathbf{A} , die Vektoren $\mathbf{v}_i \in \mathbb{R}^M$ mit

$$\mathbf{A} \mathbf{A}^\top \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

heißen **Linkssingulärvektoren** von \mathbf{A} .

Die Wurzeln $d_i = \sqrt{\lambda_i}$ der gemeinsamen Eigenwerte von $\mathbf{A}^\top \mathbf{A}$ und $\mathbf{A} \mathbf{A}^\top$ heißen **Singulärwerte** von \mathbf{A} .

Lemma

Für zwei zusammengehörende normierte Rechts- und Linkssingulärvektoren \mathbf{u}, \mathbf{v} mit Singulärwert d gelten die Aussagen

$$\mathbf{A}\mathbf{u} = d\mathbf{v} \quad \text{und} \quad \mathbf{A}^\top \mathbf{v} = d\mathbf{u}.$$

Singulärwertzerlegung

Satz

Besitzt die Matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ den Rang r , und sind $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_r)$ und $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_r)$ die Matrizen, welche die zu den nichtverschwindenden Singulärwerten d_1, \dots, d_r gehörenden orthonormierten Rechts- und Linkssingulärvektoren als Spalten enthalten, so gilt für \mathbf{A} die Darstellungsformel

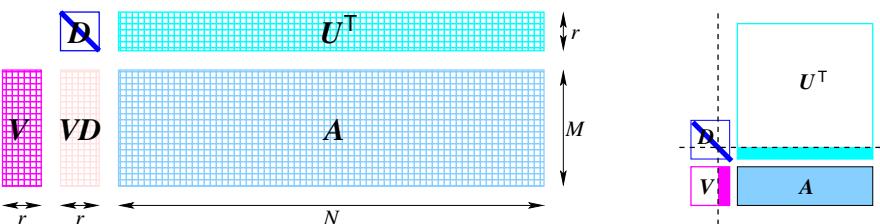
$$\mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{U}^\top = \sum_{\rho=1}^r d_\rho \mathbf{v}_\rho \mathbf{u}_\rho^\top$$

Dabei bezeichne $\mathbf{D} = \text{diag}(d_1, \dots, d_r)$.

Matrixrang und Singulärkomponenten

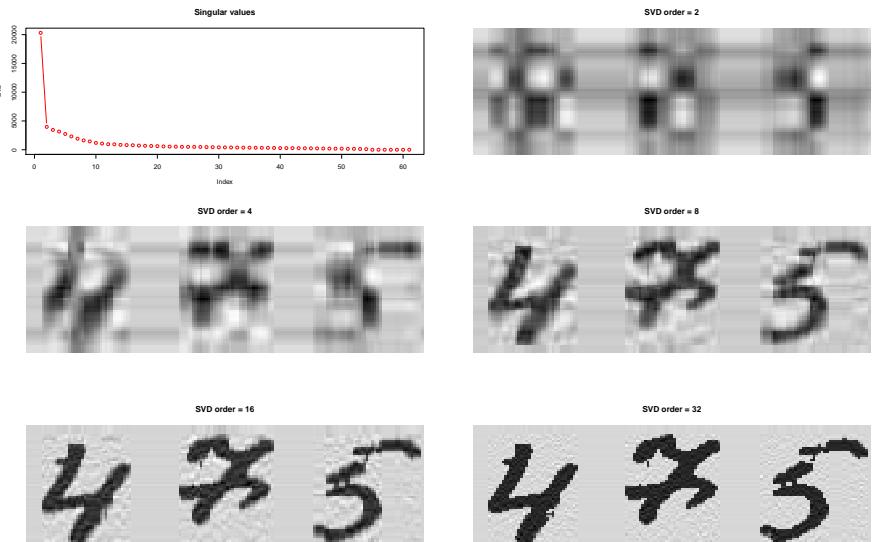
Satz

Besitzt die Matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ den Rang r , so existieren genau r positive sowie $N - r$ bzw. $M - r$ verschwindende Singulärwerte, zusammen mit ihren orthonormierten Rechts- und Linkssingulärvektoren.



Beispiel

Grauwertbild (151×61) mit Ziffern



Aufgabenstellung

Unvollständige Reihenentwicklungen

Spektrale Merkmale

Fraktale Merkmale

Analytische Merkmale

Lineare Diskriminanzanalyse

Matrixfaktorisierung

Mathematische Hilfsmittel

Diskrete Zufallsvariable

Definition

Eine Zufallsvariable \mathbb{X} heißt **diskret**, falls der Wertebereich von \mathbb{X} abzählbar ist.

Bemerkung

In diesem Falle gilt dann

$$\mathcal{V} = \bigcup_{n \in \mathbb{N}} \{\varepsilon \mid \mathbb{X}(\varepsilon) = x_n\} = \bigcup_{n \in \mathbb{N}} \{\mathbb{X} = x_n\}$$

für den abzählbaren Wertebereich $\{x_1, x_2, x_3, \dots\}$ sowie auch $P(\mathcal{V}) = 1$. Für die Wahrscheinlichkeit $P(\mathbb{X} = x_n)$ schreiben wir auch kürzer p_n .

Definition

Ist \mathbb{X} eine diskrete Zufallsvariable mit dem Wertebereich $\{x_n \mid n \in \mathbb{N}\}$, dann heißt

$$p_{\mathbb{X}} : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ x & \mapsto \begin{cases} P(\mathbb{X} = x_n) & (\exists n) x = x_n \\ 0 & \text{sonst} \end{cases} \end{cases}$$

diskrete Wahrscheinlichkeitsdichtefunktion von \mathbb{X} . Die Werte x_n heißen **Massenpunkte** von \mathbb{X} .

Zufallsvariable eines Wahrscheinlichkeitsraumes und ihre kumulative Verteilungsfunktion

Definition

Sei $(\mathcal{U}, \mathfrak{E}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung

$$\mathbb{X} : \mathcal{U} \rightarrow \mathbb{R}$$

heißt **Zufallsvariable** genau dann wenn gilt:

$$A_r = \{\varepsilon \in \mathcal{U} \mid \mathbb{X}(\varepsilon) \leq r\} \in \mathfrak{E} \quad \text{für alle } r \in \mathbb{R}$$

Definition

Die Abbildung

$$F_{\mathbb{X}} : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ r & \mapsto P(A_r) \end{cases}, \quad A_r = \{\varepsilon \in \mathcal{U} \mid \mathbb{X}(\varepsilon) \leq r\}$$

heißt **kumulative Verteilungsfunktion** der Zufallsvariablen \mathbb{X} . Für $P(A_r)$ schreiben wir üblicherweise

$$P(\mathbb{X} \leq r).$$

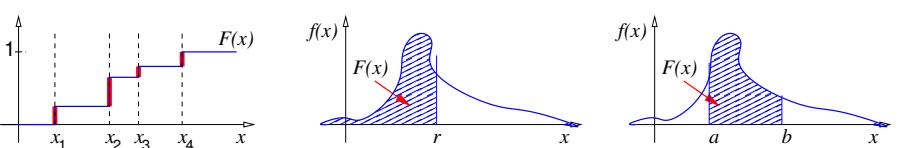
Stetige Zufallsvariable

Definition

Eine Zufallsvariable \mathbb{X} heißt **kontinuierlich**, falls der Wertebereich von \mathbb{X} nicht abzählbar ist.

Bemerkung

1. Für eine kontinuierliche Zufallsvariable \mathbb{X} sei die Wahrscheinlichkeit dafür, daß sie einen bestimmten Wert $x \in \mathbb{R}$ annimmt, gleich 0.
2. Wir betrachten daher stattdessen die Wahrscheinlichkeit dafür, daß \mathbb{X} in einem endlichen Intervall $[a, b]$ oder einem unendlichen Anfangsstück $(-\infty, r]$ von \mathbb{R} liegt.



kontinuierliche (stetige) Zufallsvariable

Wahrscheinlichkeitsverteilungsdichtefunktion

Die Dichte ist die Ableitung der Verteilung

Definition

Ist \mathbb{X} eine (kontinuierliche) Zufallsvariable mit der Eigenschaft

$$F_{\mathbb{X}}(r) = P(\mathbb{X} \leq r) = \int_{-\infty}^r f_{\mathbb{X}}(\xi) d\xi \quad \text{für alle } r \in \mathbb{R}$$

so heißt die Funktion $f_{\mathbb{X}} : \mathbb{R} \rightarrow \mathbb{R}$ die **Wahrscheinlichkeitsdichte** der **stetigen** Zufallsvariable \mathbb{X} .

Lemma

Für eine stetige Zufallsvariable \mathbb{X} mit der Dichtefunktion $f_{\mathbb{X}}$ und der Verteilungsfunktion $F_{\mathbb{X}}$ gilt für alle $a, b \in \mathbb{R}$ mit $a < b$ die Aussage

$$P(a < \mathbb{X} \leq b) = F_{\mathbb{X}}(b) - F_{\mathbb{X}}(a) = \int_a^b f_{\mathbb{X}}(\xi) d\xi.$$

(Zentrale) Momente einer Verteilung

Mittel und (quadratische) Standardabweichung (\sim Varianz)

Definition

Für eine stetige Zufallsvariable \mathbb{X} mit der Dichtefunktion $f_{\mathbb{X}}$ bezeichnen wir

$$\mu_{\mathbb{X}} = \mathcal{E}[\mathbb{X}] = \int_{-\infty}^{+\infty} x \cdot f_{\mathbb{X}}(x) dx$$

als den **Erwartungswert** der Zufallsvariablen \mathbb{X} selbst,

$$\text{Var}[\mathbb{X}] = \sigma_{\mathbb{X}}^2 = \mathcal{E}[(\mathbb{X} - \mu_{\mathbb{X}})^2] = \int_{-\infty}^{+\infty} (x - \mu_{\mathbb{X}})^2 \cdot f_{\mathbb{X}}(x) dx$$

als die **Varianz** (Streuung, Dispersion), $\sigma_{\mathbb{X}}$ als die **Standardabweichung** und

$$\mathcal{E}[\mathbb{X}^N] \quad \text{bzw.} \quad \mathcal{E}[(\mathbb{X} - \mu_{\mathbb{X}})^N]$$

als das (zentrale) **N-te Moment** von \mathbb{X} .

Erwartungswerte

Zufallsvariable · Funktion einer Zufallsvariablen

Definition

Für eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ ist mit \mathbb{X} auch $g(\mathbb{X})$ eine Zufallsvariable. Die Summe bzw. das Integral

$$\begin{aligned} \mathcal{E}[g(\mathbb{X})] &= \sum_n g(x_n) \cdot P(\mathbb{X} = x_n) \\ \mathcal{E}[g(\mathbb{X})] &= \int_{-\infty}^{+\infty} g(x) \cdot f_{\mathbb{X}}(x) dx \end{aligned}$$

heißt — im Falle der Konvergenz — der **Erwartungswert** der Zufallsgröße $g(\mathbb{X})$.

Eigenschaften der Verteilungsmomente

Lemma

Für die Erwartungswerte von Funktionen diskreter oder stetiger Zufallsvariablen gelten die Aussagen:

1. Für alle $a \in \mathbb{R}$ ist $\mathcal{E}[a] = a$.
2. Homogenität: $\mathcal{E}[a \cdot g(\mathbb{X})] = a \cdot \mathcal{E}[g(\mathbb{X})]$
3. Additivität: $\mathcal{E}[g_1(\mathbb{X}) + g_2(\mathbb{X})] = \mathcal{E}[g_1(\mathbb{X})] + \mathcal{E}[g_2(\mathbb{X})]$
4. Monotonie: $g_1(x) \leq g_2(x) \Rightarrow \mathcal{E}[g_1(\mathbb{X})] \leq \mathcal{E}[g_2(\mathbb{X})]$
5. Falls $\mathcal{E}[\mathbb{X}^2]$ existiert, so gilt $\text{Var}[\mathbb{X}] = \mathcal{E}[\mathbb{X}^2] - \mathcal{E}[\mathbb{X}]^2$

Ungleichungen für Wahrscheinlichkeiten

Verteilungsunabhängige Abschätzungen

Satz (Tschebyscheff-Ungleichung)

Ist die Abbildung $g : \mathbb{R} \rightarrow \mathbb{R}$ nichtnegativ, so gilt für jedes $\lambda > 0$:

$$P(g(\mathbb{X}) \geq \lambda) \leq \frac{\mathcal{E}[g(\mathbb{X})]}{\lambda}$$

Im Falle endlicher Varianz von \mathbb{X} gilt die bekannte Form:

$$P(|\mathbb{X} - \mu_{\mathbb{X}}| \geq c \cdot \sigma_{\mathbb{X}}) \leq \frac{1}{c^2}$$

Satz (Jensen-Ungleichung)

Für eine Zufallsvariable \mathbb{X} und eine konvexe Abbildung $g : \mathbb{R} \rightarrow \mathbb{R}$ gilt:

$$\mathcal{E}[g(\mathbb{X})] \geq g(\mathcal{E}[\mathbb{X}])$$

Folgerung

Weil bekanntlich $g(x) = x^2$ konvex ($g'' \geq 0$) ist, folgt

$$\text{Var}[\mathbb{X}] = \mathcal{E}[\mathbb{X}^2] - \mathcal{E}[\mathbb{X}]^2 \geq 0.$$

Zufallsvektoren

Multivariate Wahrscheinlichkeitsverteilungen

Definition

Ist \mathfrak{X} ein Vektor von Zufallsvariablen $\mathbb{X}_1, \dots, \mathbb{X}_D$, so heißt

$$F_{\mathfrak{X}}(\mathbf{x}) = F_{\mathfrak{X}}(x_1, \dots, x_D) = P(\mathbb{X}_1 \leq x_1, \dots, \mathbb{X}_D \leq x_D)$$

die **multivariate Verteilungsfunktion** von \mathfrak{X} .

Sie heißt **stetig**, falls eine Funktion $f_{\mathfrak{X}} : \mathbb{R}^D \rightarrow \mathbb{R}$ existiert mit der Eigenschaft

$$F_{\mathfrak{X}}(x_1, \dots, x_D) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_D} f_{\mathfrak{X}}(x_1, \dots, x_D) dx_1 \dots dx_D$$

$f_{\mathfrak{X}}$ heißt dann **multivariate Verteilungsdichte** von \mathfrak{X} .

Momente multivariater Verteilungen

Erwartungswertvektor und Kovarianzmatrix

Definition

Für einen stetigen Vektor \mathfrak{X} von Zufallsvariablen $\mathbb{X}_1, \dots, \mathbb{X}_D$ mit der multivariaten Verteilungsdichte $f_{\mathfrak{X}}$ definieren wir

$$\boldsymbol{\mu}_{\mathfrak{X}} = \mathcal{E}[\mathfrak{X}] = \int_{\mathbb{R}^D} \mathbf{x} \cdot f_{\mathfrak{X}}(\mathbf{x}) d\mathbf{x}$$

als den **Erwartungswertvektor** und

$$\mathbf{S}_{\mathfrak{X}} = \begin{pmatrix} \sigma_{\mathbb{X}_1 \mathbb{X}_1} & \dots & \sigma_{\mathbb{X}_1 \mathbb{X}_D} \\ \vdots & \ddots & \vdots \\ \sigma_{\mathbb{X}_D \mathbb{X}_1} & \dots & \sigma_{\mathbb{X}_D \mathbb{X}_D} \end{pmatrix}$$

als die **Kovarianzmatrix** von \mathfrak{X} . Dabei bezeichne für alle $j = 1, \dots, D$:

$$\begin{aligned} \sigma_{\mathbb{X}_i \mathbb{X}_j} &= \text{Cov}[\mathbb{X}_i, \mathbb{X}_j] = \mathcal{E}[(\mathbb{X}_i - \mu_{\mathbb{X}_i})(\mathbb{X}_j - \mu_{\mathbb{X}_j})] \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (\mathbb{X}_i - \mu_{\mathbb{X}_i})(\mathbb{X}_j - \mu_{\mathbb{X}_j}) \cdot f_{\mathfrak{X}, \mathbb{X}_j}(\mathbb{X}_i, \mathbb{X}_j) d\mathbb{X}_i d\mathbb{X}_j \end{aligned}$$

Statistische Unabhängigkeit

Statistische Unkorreliertheit

Definition

Die Zufallsvariablen $\mathbb{X}_1, \dots, \mathbb{X}_D$ heißen **unabhängig**, wenn gilt

$$f_{\mathfrak{X}}(x_1, \dots, x_D) = \prod_{d=1}^D f_{\mathbb{X}_d}(x_d) = f_{\mathbb{X}_1}(x_1) \cdot \dots \cdot f_{\mathbb{X}_D}(x_D)$$

Die Zufallsvariablen heißen **unkorreliert**, wenn gilt

$$\mathcal{E}\left[\prod_{d=1}^D \mathbb{X}_d\right] = \prod_{d=1}^D \mathcal{E}[\mathbb{X}_d]$$

Unabhängigkeit & Unkorreliertheit

Einige wichtige Eigenschaften

1. Aus der Unabhängigkeit folgt die Unkorreliertheit.
2. Aus der Unkorreliertheit folgt i.a. **nicht** die Unabhängigkeit.
3. Zwei Zufallsvariablen \mathbb{X}, \mathbb{Y} sind unkorreliert, wenn $E[\mathbb{X}\mathbb{Y}] = \mu_{\mathbb{X}}\mu_{\mathbb{Y}}$ gilt, also gdw. gilt $\text{Cov}[\mathbb{X}, \mathbb{Y}] = 0$.
4. Sind die Zufallsvariablen $\mathbb{X}_1, \dots, \mathbb{X}_D$ unabhängig, so sind sie auch paarweise unabhängig und natürlich auch paarweise unkorreliert.
Folglich gilt für ihre Kovarianzmatrix

$$\mathbf{S}_{\mathbb{X}} = \text{diag}(\sigma_{\mathbb{X}_1}^2, \dots, \sigma_{\mathbb{X}_D}^2) = \begin{pmatrix} \sigma_{\mathbb{X}_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\mathbb{X}_2}^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\mathbb{X}_D}^2 \end{pmatrix}$$

mit $\sigma_{\mathbb{X}_d}^2 = \sigma_{\mathbb{X}_d \mathbb{X}_d}$, $d = 1, \dots, D$.

Eigenzerlegung symmetrischer Matrizen

Eigenvektoren $\hat{=}$ Orthogonalbasis

Lemma

Für reellwertige symmetrische Matrizen ($\mathbf{A}^\top = \mathbf{A} \in \mathbb{R}^{n \times n}$) gilt:

1. Eigenvektoren zu **verschiedenen** Eigenwerten sind orthogonal:
- $$\left\{ \begin{array}{l} \mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \\ \mathbf{A}\mathbf{u}_j = \lambda_j \mathbf{u}_j \end{array} \right\}, \quad \lambda_i \neq \lambda_j \Rightarrow \mathbf{u}_i^\top \mathbf{u}_j = 0$$
2. Die Eigenvektoren **mehrerer** Eigenwerte sind orthogonalisierbar (Gram-Schmidt).
 3. Die normierten Eigenvektoren $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ von \mathbf{A} bilden ein Orthonormalsystem:
- $$\mathbf{U}^\top \mathbf{U} = \mathbf{E} = \mathbf{U} \mathbf{U}^\top$$

Matrixschreibweise für die spezielle EWA

$$\underbrace{\mathbf{A}}_{A} \cdot \underbrace{\mathbf{U}}_{U} = \underbrace{\mathbf{U}}_{U} \cdot \underbrace{\mathbf{D}}_{D}, \quad \mathbf{D} \stackrel{\text{def}}{=} \text{diag}(\lambda_1, \dots, \lambda_n)$$

Spezielle Eigenwertaufgabe

Definition

Sind $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ zwei quadratische Matrizen, so heißt

$$\mathbf{Ax} = \lambda \mathbf{x} \quad \text{bzw.} \quad \mathbf{Ax} = \lambda \mathbf{Bx} \quad (\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n)$$

spezielle (allgemeine) **Eigenwertaufgabe**. Ist obige Gleichung erfüllt, so heißt \mathbf{x} **Eigenvektor** zum **Eigenwert** λ .

Lemma

Für die spezielle Eigenwertaufgabe gelten folgende Aussagen:

1. Skalierungseigenschaft: $\mathbf{Ax} = \lambda \mathbf{x} \Rightarrow \mathbf{A}(\alpha \mathbf{x}) = \lambda(\alpha \mathbf{x})$
2. Die spezielle EWA ist äquivalent zum homogenen LGS $(\mathbf{A} - \lambda \mathbf{E}) \cdot \mathbf{x} = \mathbf{0}$.
3. Die EWA besitzt nichttriviale Lösungen $\mathbf{x} \neq \mathbf{0}$ genau zu den Lösungen des **charakteristischen Polynoms** $P(\lambda) = \det(\mathbf{A} - \lambda \mathbf{E}) = 0$.
4. \mathbf{A} besitzt genau n (nicht notwendig verschiedene) Eigenwerte $\lambda_1, \dots, \lambda_n$.
5. Ist \mathbf{A} symmetrisch, so sind alle $\lambda_1, \dots, \lambda_n$ reell.
6. Sind alle λ_i verschieden, so sind $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ linear unabhängig.

Eigenzerlegung symmetrischer Matrizen

Diagonalisierung und dyadische Reihenentwicklung

Lemma

Sind $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ bzw. $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ die (normierten) Eigenvektoren und Eigenwerte der reellen symmetrischen Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, so gilt:

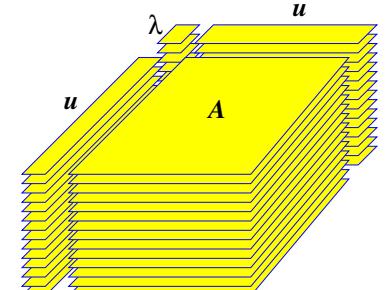
1. **Diagonalisierung**

$$\underbrace{\mathbf{A}}_{D} = \underbrace{\mathbf{U}}_{U^\top} \cdot \underbrace{\mathbf{D}}_{A} \cdot \underbrace{\mathbf{U}}_{U}$$
2. **Zerlegung**

$$\underbrace{\mathbf{A}}_{A} = \underbrace{\mathbf{U}}_{U} \cdot \underbrace{\mathbf{D}}_{D} \cdot \underbrace{\mathbf{U}^\top}_{U^\top}$$
3. **Invertierung**

$$\underbrace{\mathbf{A}^{-1}}_{A^{-1}} = \underbrace{\mathbf{U}}_{U} \cdot \underbrace{\mathbf{D}^{-1}}_{D^{-1}} \cdot \underbrace{\mathbf{U}^\top}_{U^\top}$$
4. **Dyadische Reihe**

$$\mathbf{A} = \sum_{\nu=1}^n \lambda_\nu \cdot \mathbf{u}_\nu \cdot \mathbf{u}_\nu^\top$$



Quadratische Formen

Positiv-(semi)definite Matrizen

Definition

Ist \mathbf{A} reell und symmetrisch, so heißt die Abbildung

$$Q : \mathbf{x} \mapsto \mathbf{x}^\top \mathbf{A} \mathbf{x}$$

quadratische Form. Dieses $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt **positiv-definit**, wenn für alle \mathbf{x} gilt:

$$\mathbf{x} \neq \mathbf{0} \Rightarrow Q(\mathbf{x}) > 0$$

analog: negativ-definit ($Q(\mathbf{x}) < 0$), positiv-semidefinit ($Q(\mathbf{x}) \geq 0$), negativ-semidefinit ($Q(\mathbf{x}) \leq 0$).

Lemma

Die quadratische Form $Q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$ ist $\begin{cases} \text{positiv-definit} \\ \text{negativ-definit} \\ \text{positiv-semidefinit} \\ \text{negativ-semidefinit} \end{cases}$ genau dann,

wenn \mathbf{A} ausschließlich $\begin{cases} \text{positive} \\ \text{negative} \\ \text{nichtnegative} \\ \text{nichtpositive} \end{cases}$ Eigenwerte besitzt.

Bemerkung

Eine positiv-semidefinit symmetrische Matrix ist invertierbar, gdw. sie sogar positiv-definit ist.

Beweis.

Für die quadratische Form gilt die Darstellung

$$\begin{aligned} Q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} &= \mathbf{x}^\top \cdot (\mathbf{U} \mathbf{D} \mathbf{U}^\top) \cdot \mathbf{x} \\ &= (\mathbf{U}^\top \mathbf{x})^\top \cdot \mathbf{D} \cdot (\mathbf{U}^\top \mathbf{x}) \\ &= \tilde{\mathbf{x}}^\top \mathbf{D} \tilde{\mathbf{x}} \quad \text{mit } \tilde{\mathbf{x}} = \mathbf{U}^\top \mathbf{x} \\ &= \sum_{i=1}^n \sum_{j=1}^n \tilde{x}_i D_{ij} \tilde{x}_j = \sum_{i=1}^n \tilde{x}_i^2 \lambda_i \end{aligned}$$

Ferner gilt wegen der Normierungsvoraussetzung $\|\mathbf{x}\| = 1$ auch für die transformierten Vektorkoordinaten

$$\sum_{i=1}^n \tilde{x}_i^2 = \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}} = (\mathbf{U}^\top \mathbf{x})^\top (\mathbf{U}^\top \mathbf{x}) = \mathbf{x}^\top \mathbf{U} \mathbf{U}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{x} = 1.$$

Also ist $\sum_i \tilde{x}_i^2 \lambda_i$ eine Konvexitätskombination und nimmt maximal den Wert

$$\lambda_k = \max \{ \lambda_i \mid i = 1, \dots, n \}$$

an; dazu muß $\tilde{\mathbf{x}} = \mathbf{U}^\top \mathbf{x}$ der k -te Einheitsvektor sein, also folgt

$$\mathbf{u}_k^\top \mathbf{x} \stackrel{!}{=} \begin{cases} 1 & i=k \\ 0 & i \neq k \end{cases} \Rightarrow \mathbf{x} = \mathbf{u}_k$$

□

Quadratische Formen

Maximierung und Minimierung quadratischer Formen und ihrer Summen

Satz

Die quadratische Form $Q(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$ nimmt — unter der Nebenbedingung $\|\mathbf{x}\| = 1$ — ihren maximalen (minimalen) Wert an, wenn \mathbf{x} der Eigenvektor \mathbf{u}_k zum größten (kleinsten) Eigenwert λ_k ist. Es gilt dann

$$Q(\mathbf{u}_k) = \mathbf{u}_k^\top \mathbf{A} \mathbf{u}_k = \lambda_k.$$

Lemma

Die Summe

$$s = \sum_{\mu=1}^m Q(\phi_\mu) = \sum_{\mu=1}^m \phi_\mu^\top \mathbf{A} \phi_\mu$$

quadratischer Formen nimmt — unter der Voraussetzung normierter, paarweise orthogonaler Vektoren ϕ_μ — ihr Maximum (Minimum) an, wenn für die ϕ_μ die zu den m größten (kleinsten) Eigenwerten $\lambda_1, \dots, \lambda_m$ von \mathbf{A} gehörenden Eigenvektoren $\mathbf{u}_1, \dots, \mathbf{u}_m$ eingesetzt werden. Es gilt dann

$$s = \sum_{\mu=1}^m Q(\mathbf{u}_\mu) = \sum_{\mu=1}^m \lambda_\mu.$$

Von Mises Iteration

zur Berechnung des eigenwertgrößten Eigenvektors einer Matrix

Bemerkung

Für einen normierten Eigenvektor \mathbf{u} einer symmetrischen Matrix \mathbf{A} mit Eigenwert λ gelten die Aussagen

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u}, \quad \|\mathbf{u}\| = 1$$

und damit die „iterationsverdächtige“ Beziehung

$$\mathbf{u} = \frac{\mathbf{A}\mathbf{u}}{\|\mathbf{A}\mathbf{u}\|}$$

Lemma

Ist \mathbf{A} eine symmetrische, positiv-definite ($N \times N$)-Matrix, so strebt im Iterationsverfahren

$$\lambda^{(i+1)} = \|\mathbf{A}\mathbf{u}^{(i)}\| \quad \text{nebst} \quad \mathbf{u}^{(i+1)} = \frac{\mathbf{A}\mathbf{u}^{(i)}}{\|\mathbf{A}\mathbf{u}^{(i)}\|}$$

die Folge der $\lambda^{(i)}$ gegen den größten Eigenwert λ von \mathbf{A} und die Folge der $\mathbf{u}^{(i)}$ gegen den/einen zu λ gehörenden Eigenvektor, sofern der Startvektor $\mathbf{u}^{(0)} \in \mathbb{R}^N$ „allgemein“ bezüglich des Orthonormalsystems der Eigenvektoren war.

Von Mises Deflation

zur Berechnung der $M < N$ eigenwertgrößten Eigenvektoren einer Matrix

Bemerkung

Wegen der dyadischen Summenformel

$$\mathbf{A} = \sum_{n=1}^N \lambda_n \mathbf{u}_n \mathbf{u}_n^\top$$

für quadratische Matrizen mit Eigenvektorbasis lässt sich die von Mises-Iteration zu einem Verfahren zur sukzessiven Bestimmung *mehrerer* Eigenvektoren zu absteigend geordneten Eigenwerten ausbauen.

(Algorithmus)

- 0 Setze $\mathbf{A}^{(1)}$ gleich \mathbf{A} . Wiederhole für $n = 1, 2, \dots, M$:
- 1 Berechne λ_n und \mathbf{u}_n mit

$$\mathbf{A}^{(n)} \mathbf{u}_n = \lambda_n \mathbf{u}_n, \quad \|\mathbf{u}_n\| = 1$$

nach dem von Mises-Verfahren.

- 2 Deflatiere die aktuelle Matrix um den neuen Eigenvektor

$$\mathbf{A}^{(n+1)} = \mathbf{A}^{(n)} - \lambda_n \mathbf{u}_n \mathbf{u}_n^\top.$$

(summeinA)

Zusammenfassung (5)

1. Die Merkmalgewinnung dient der **trennscharfen** Repräsentation von Mustern in einem Raum konstanter, möglichst **niedriger Dimension**.
2. Unvollständige **Reihenentwicklungen** wie Walsh, Haar, Legendre gehören zu den datenunabhängigen Heuristiken.
3. **Spektralkoeffizienten** sind durch komplexwertige Reihenentwicklung definiert.
4. Die **linearen Vorhersagekoeffizienten** (LPC) stellen ein geglättetes Spektrum dar.
5. Die **Filigranität** von Kurven oder Flächen kann durch näherungsweise Berechnung **fraktaler Dimensionen** charakterisiert werden.
6. **Analytische** Reihenentwicklungen passen sich an die statistischen Mustereigenschaften an; sie basieren auf dem mathematischen Konzept der **Eigenräume**.
7. Die **Hauptachsentransformation** (PCA) rotiert die Datenkoordinaten in die Richtungen **maximaler Streuung**.
8. Die **Diskriminanzanalyse** (LDA) orientiert sich an den **Klassenzugehörigkeiten** der Muster und besitzt meistens die besseren Trennungseigenschaften.
9. Die **Singulärwertzerlegung** (SVD) erlaubt eine unvollständige Reihenentwicklung für **Rechteckmatrizen**.