
2. Merkmale

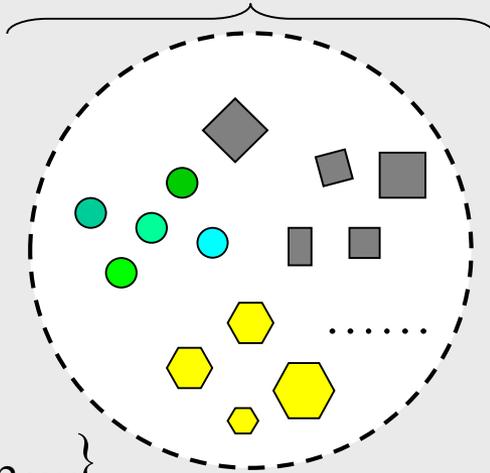
2. Merkmale

- 2.1. Merkmalstypen
- 2.2. Sichtung des Merkmalsraumes
- 2.3. Transformation der Merkmale
- 2.4. Abstandsmessung im Merkmalsraum
- 2.5. Normalisierung
- 2.6. Auswahl und Konstruktion von Merkmalen
 - 2.6.1. Deskriptive Merkmale
 - 2.6.2. Modellparameter als Merkmale
 - 2.6.3. Konstruktion invarianter Merkmale
- 2.7. Reduktion der Dimension des Merkmalsraumes
 - 2.7.1. Hauptkomponentenanalyse (PCA)
 - 2.7.2. Nichtlineare Hauptkomponentenanalyse (*Kernelized PCA*)
 - 2.7.3. Independent Component Analysis (ICA)
 - 2.7.4. Multiple Discriminant Analysis
 - 2.7.5. Dimensionsreduktion durch Merkmalsauswahl

2. Merkmale Wiederholung

Domäne \subseteq Welt

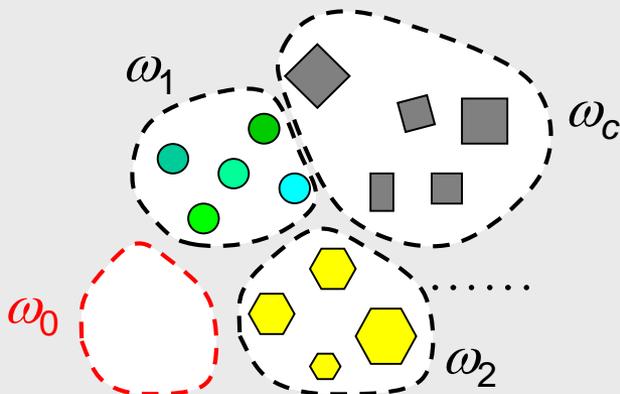
Objekte



$$\Omega = \{o_1, o_2, \dots\}$$

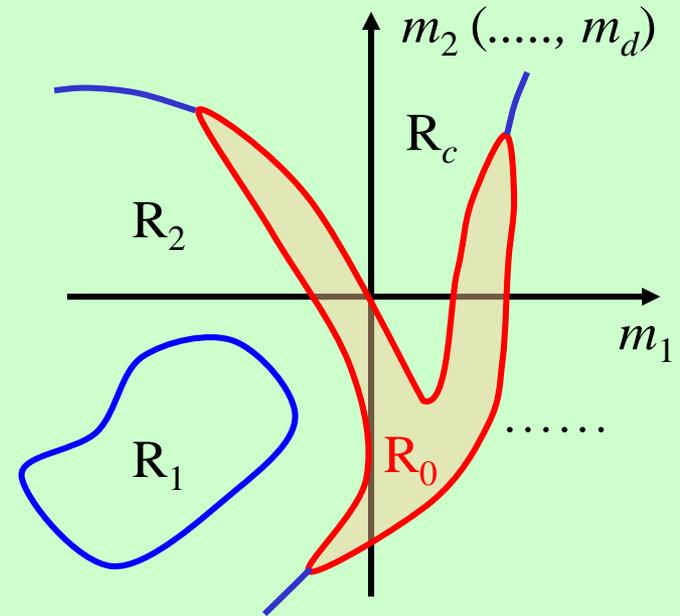
$$\omega_i = [o_j]_{\sim}$$

$$\Omega / \sim = \{\omega_1, \dots, \omega_c\} = \{[o_j]_{\sim} \mid o_j \in \Omega\}$$



Beobachten,
messen,
beschreiben:
Objekt α \mathbf{m}

Merkmalsraum \mathbb{M}



Partition von \mathbb{M} :

$$R_i \cap R_j = \emptyset \text{ für } i \neq j$$

$$\bigcup_{i=0}^c R_i = \mathbb{M}$$

Merkmalsvektor:

$$\mathbf{m} := (m_1, \dots, m_d)^T \in \mathbb{M}$$

Klassifikation:

$$\mathbf{m} \in R_j \Rightarrow \hat{\omega} := \omega_j$$

2. Merkmale

Merkmale (Synonyme: Deskriptoren, Kenngrößen)

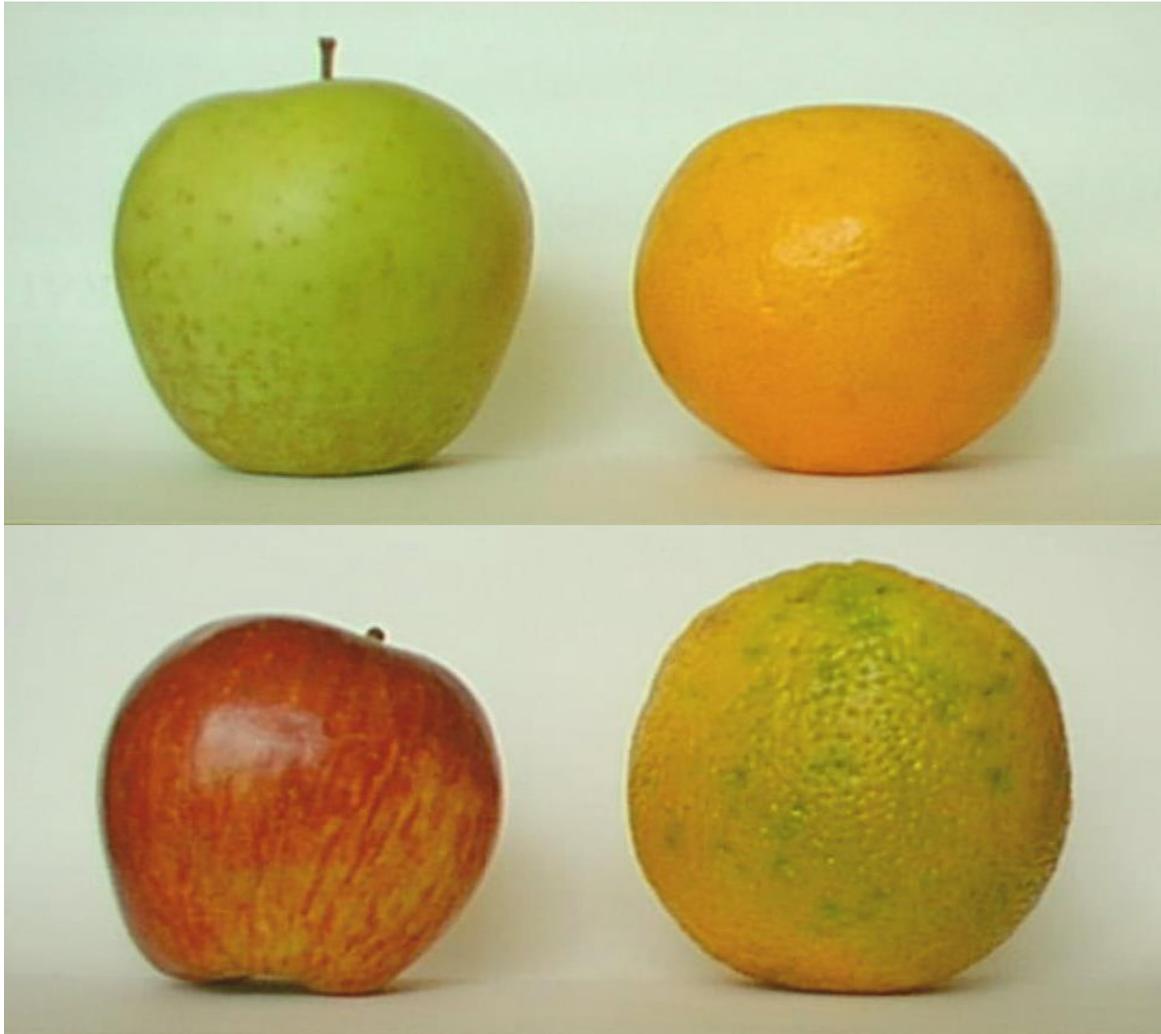
Beobachtbare bzw. physikalisch messbare Eigenschaften von Objekten, die im Kontext der Mustererkennung geeignet sein sollen, eine möglichst fehlerfreie Klassifikation zu bewerkstelligen.

Gewünschte Eigenschaften von Merkmalen:

- beobachtbar bzw. physikalisch messbar
- mit geringem Aufwand zu gewinnen
- interpretierbar, anschauliche Bedeutung (falls möglich)
- invariant bezüglich irrelevanter Transformationen und Variationen
- hohe Relevanz bezüglich der Aufgabenstellung
 - geringe Variation innerhalb der Klassen
 - große Variation zwischen den Klassen } hohe Diskriminanz
- robust (unempfindlich gegenüber Störungen)
- möglichst geringe Abhängigkeit zw. unterschiedlichen Merkmalen
- möglichst geringe Zahl von Merkmalen (Merkmalsvektor niedriger Dimension)

2. Merkmale

Beispiel: Klassifikation von Äpfeln und Orangen



2.1. Merkmalstypen

	Skala				
	qualitativ		quantitativ (metrisch)		
	Nominal-	Ordinal-	Intervall-	Verhältnis-	Absolut-
Empirische Relationen	~ Äquivalenz	~ Äquivalenz ϕ Ordnung	~ Äquivalenz ϕ Ordnung \oplus Emp. Addition	~ Äquivalenz ϕ Ordnung \oplus Emp. Addition \otimes Emp. Multipl.	~ Äquivalenz ϕ Ordnung \oplus Emp. Addition \otimes Emp. Multipl.
Zulässige Transformationen	$m' = f(m)$ $f(.)$ bijektiv	$m' = f(m)$ $f(.)$ streng monoton	$m' = am + b$ mit $a > 0$	$m' = am$ mit $a > 0$	$m' = m$
Beispiele zugehörige Merkmale	Telefonnum., Kfz-Kennz., Typen, PLZ, Geschlecht	Güteklassen, Härtegrad, Windstärke	Temp. in C°, F°, Kalenderzeit, geographische Höhe	Masse, Länge, el. Strom,	Quantenzahlen, Teilchenanzahl, Fehlerzahl
Werte von m	Zahlen, Namen, Symbole	i.d.R. natürliche Zahlen	i.d.R. reelle Zahlen	i.d.R. reelle Zahlen > 0	i.d.R. natürliche Zahlen
Aussagekraft	gering	→	→	→	hoch

2.2. Sichtung des Merkmalsraumes

Merkmale sichten:

Visualisierung der Stichproben im Merkmalsraum, um im Entwurfsprozess einen Eindruck von der Leistungsfähigkeit der Merkmale zu gewinnen.

Probleme:

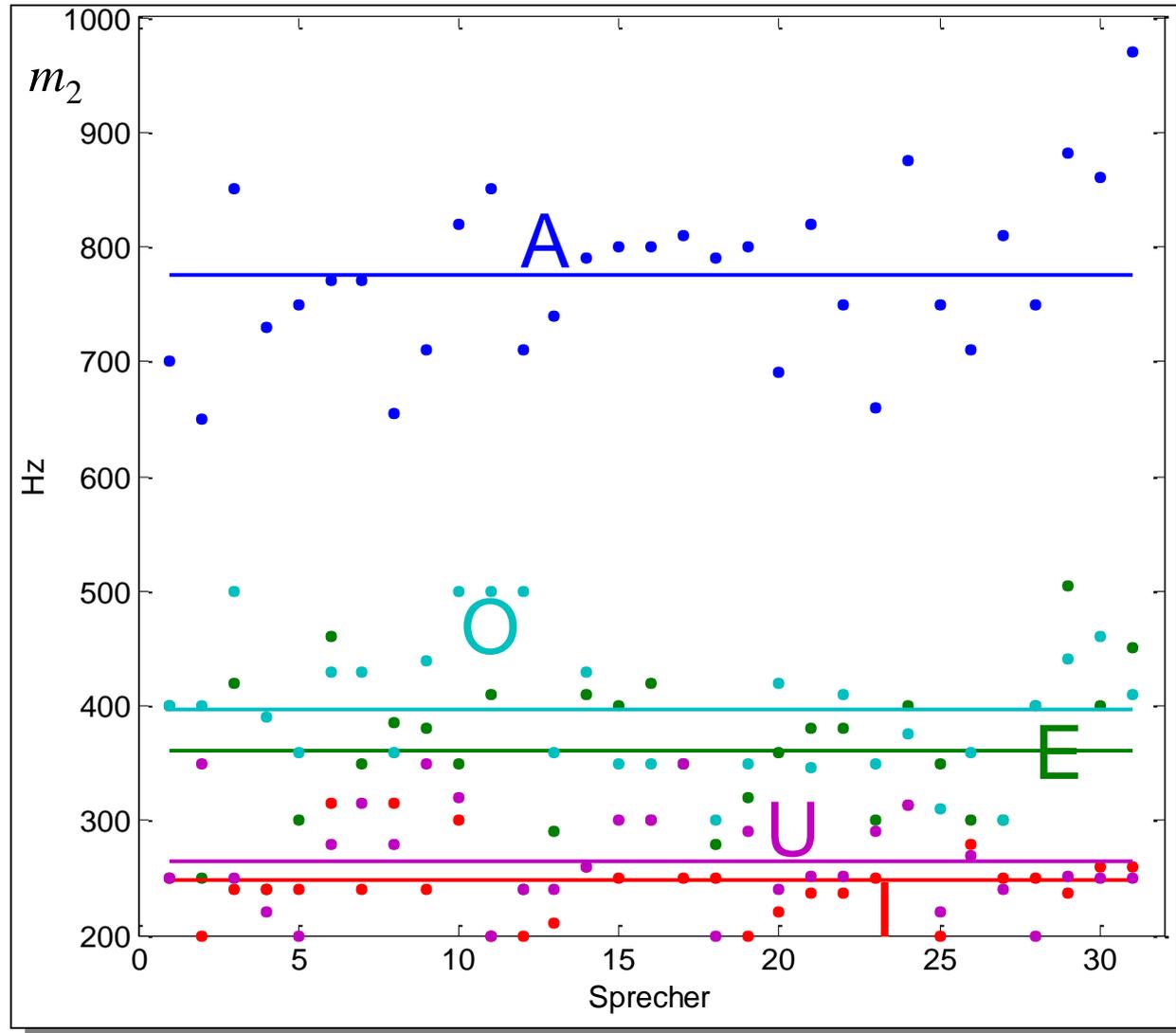
- menschliche Vorstellungskraft reicht nur für 3 Dimensionen
- nur 2 Dimensionen sind graphisch darstellbar

Methoden:

- Betrachtung 2-dimensionaler Projektionen des Merkmalsraumes
- Betrachtung von 2D-Schnitten (Scheiben) im Merkmalsraum

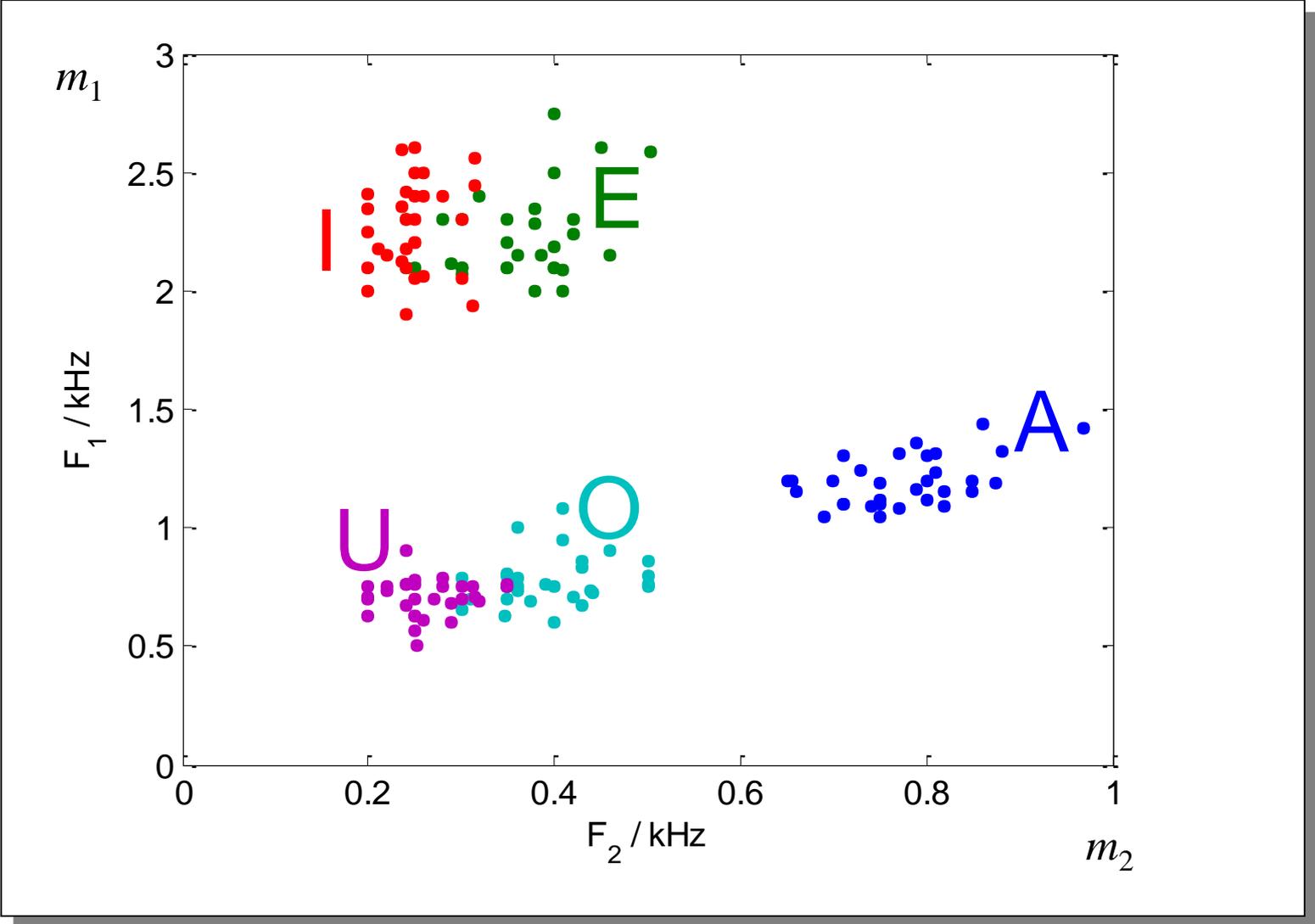
2.2. Sichtung des Merkmalsraumes

Beispiel: Klassifikation von Vokalen anhand des zweiten Formanten



2.2. Sichtung des Merkmalsraumes

Beispiel: Klassifikation von Vokalen anhand der ersten beiden Formanten



2.2. Sichtung des Merkmalsraumes

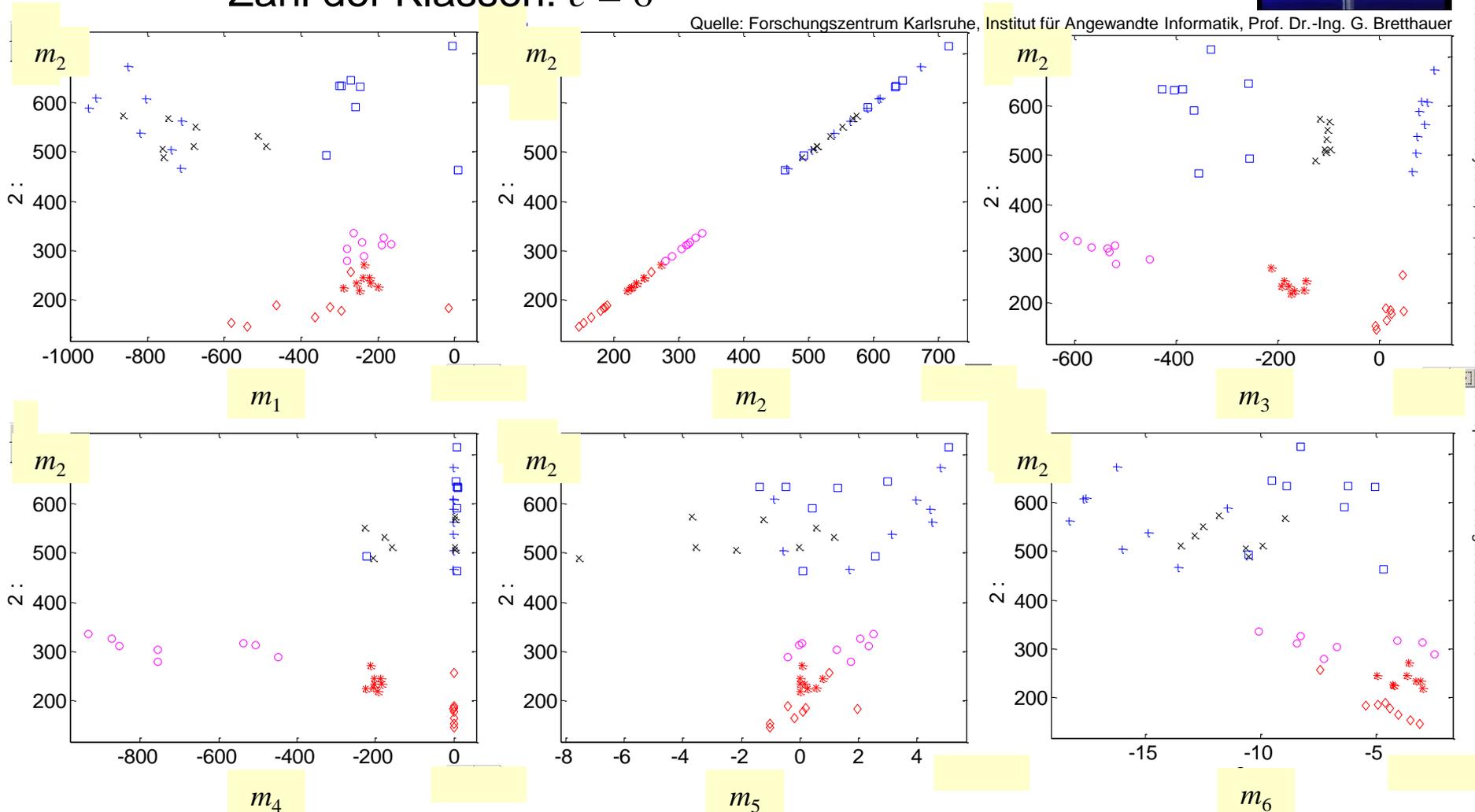
Beispiel: Stichproben für „die Hand“, 2D-Projektionen

Dimension des Merkmalsraumes: $d = 6$

Zahl der Klassen: $c = 6$



Quelle: Forschungszentrum Karlsruhe, Institut für Angewandte Informatik, Prof. Dr.-Ing. G. Bretthauer



2.2. Sichtung des Merkmalsraumes



Beispiel: IRIS Datensatz

- Der Iris Datensatz wurde 1936 von R. A. Fisher veröffentlicht.
- Aufgabe: Klassifikation der Spezies:
 - Iris Versicolor
 - Iris Virginica
 - Iris Setosa
- Datensatz umfasst 150 Beispiele, 50 Bsp./Klasse
- Merkmale:
 - sepal length (Kelchblattlänge)
 - sepal width (Kelchblattbreite)
 - petal length (Blütenblattlänge)
 - petal width (Blütenblattbreite)

2.2. Sichtung des Merkmalsraumes

TABLE I
THE IRIS DATA: FISHER [2]

Iris setosa				Iris versicolor				Iris virginica			
Sepal Leng.	Sepal Width	Petal Leng.	Petal Width	Sepal Leng.	Sepal Width	Petal Leng.	Petal Width	Sepal Leng.	Sepal Width	Petal Leng.	Petal Width
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
4.6	3.2	1.4	0.2	6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
5.0	3.3	1.4	0.2	5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8

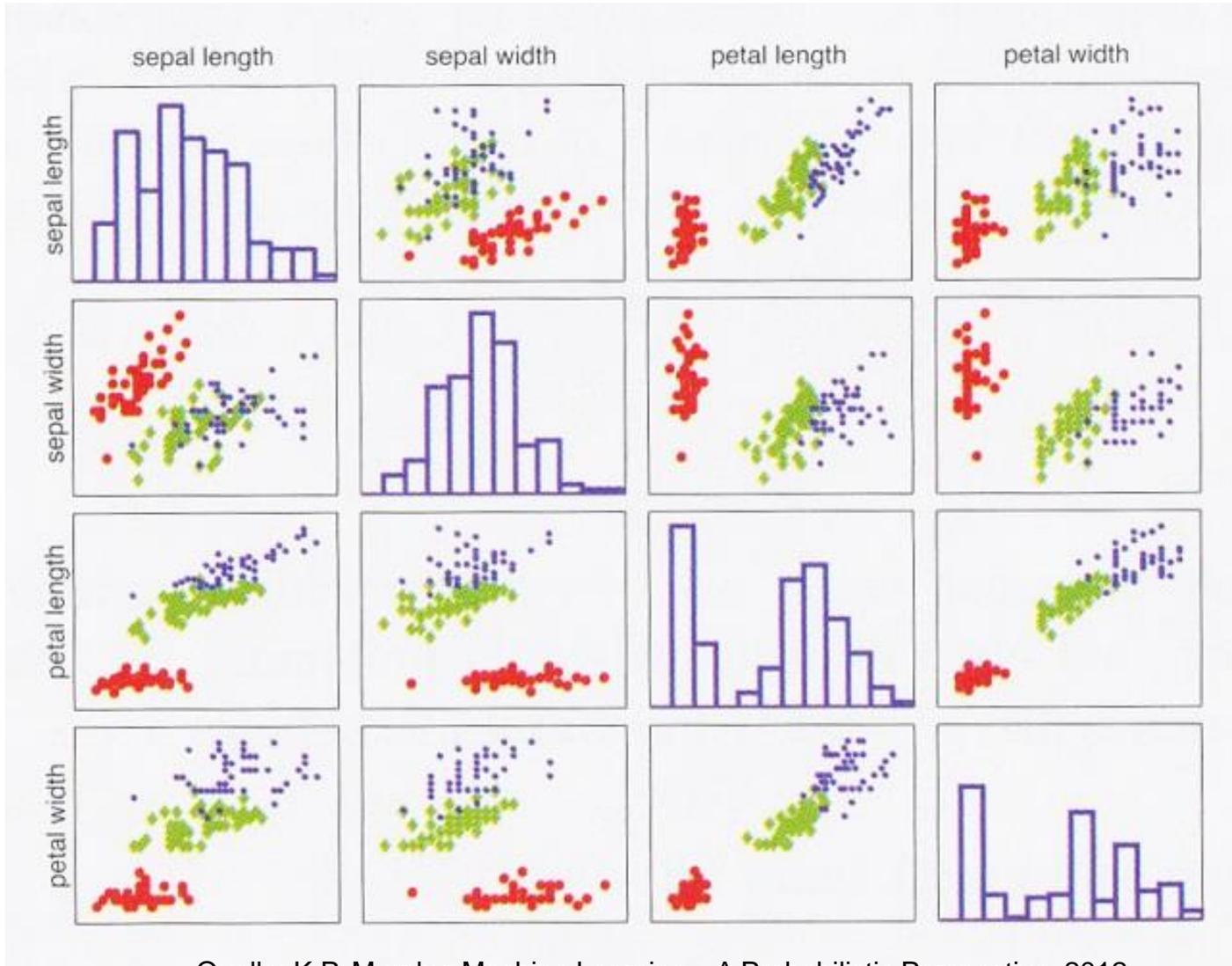
Merkmale:



- m_1 = sepal length (Kelchblattlänge)
- m_2 = sepal width (Kelchblattbreite)
- m_3 = petal length (Blütenblattlänge)
- m_4 = petal width (Blütenblattbreite)

2.2. Sichtung des Merkmalsraumes

Bsp.: IRIS Datensatz

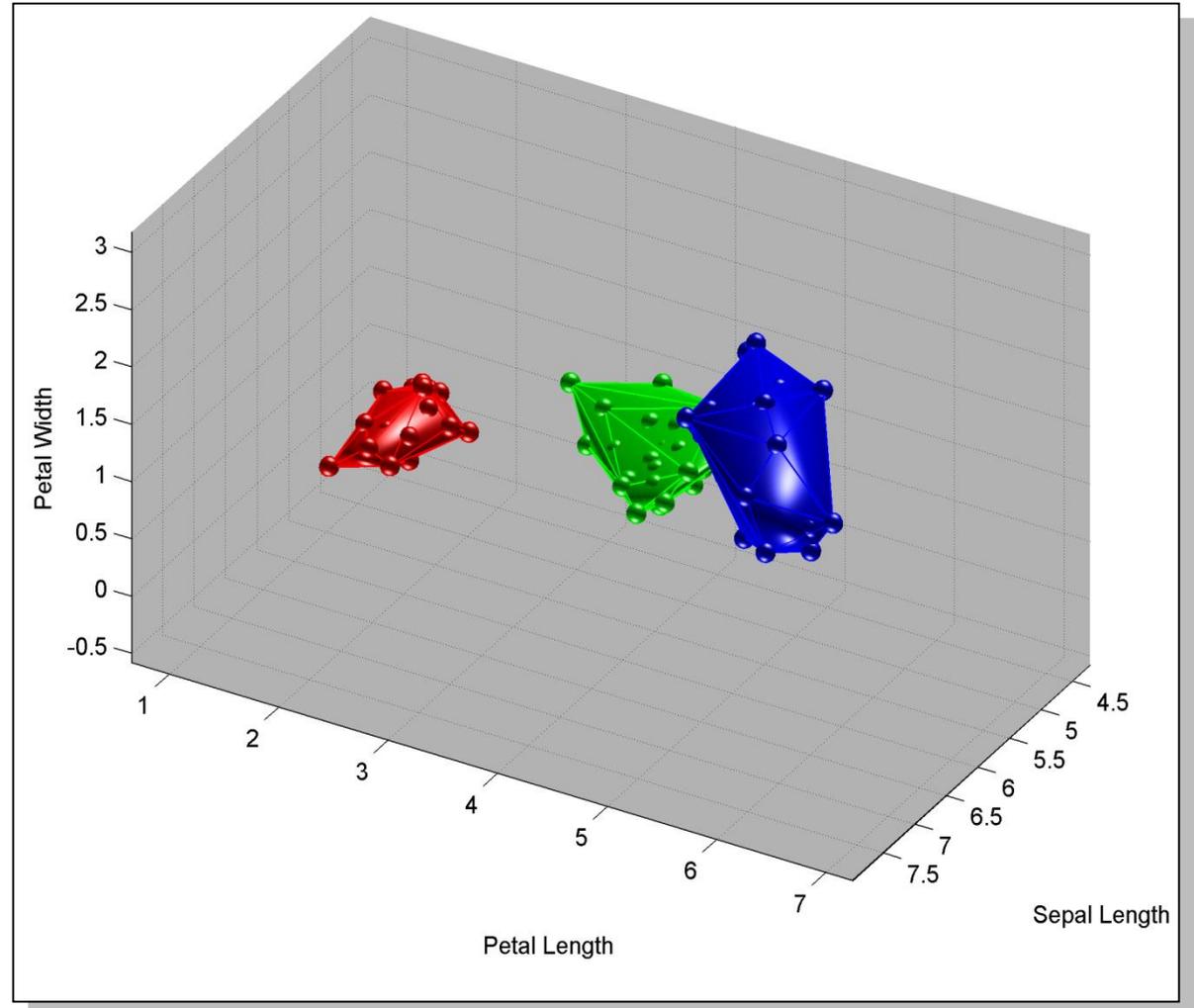


Quelle: K.P. Murphy; Machine Learning – A Probabilistic Perspective, 2012

2.2. Sichtung des Merkmalsraumes

Bsp.: IRIS Datensatz

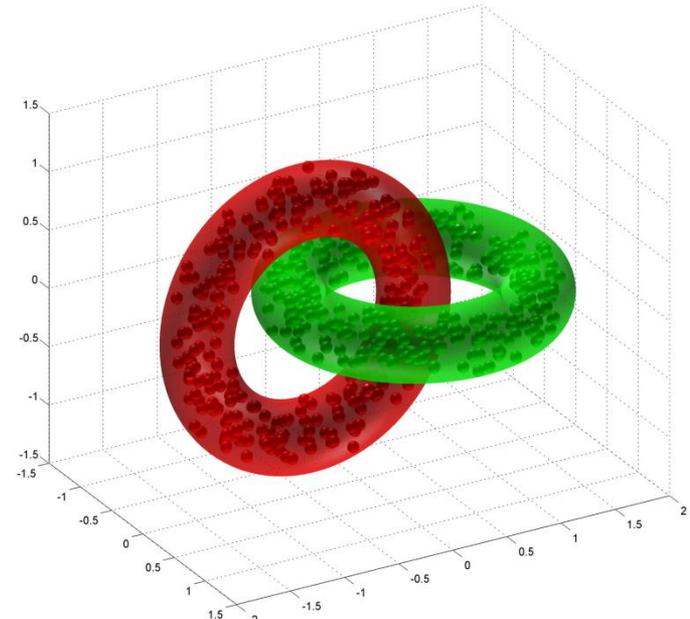
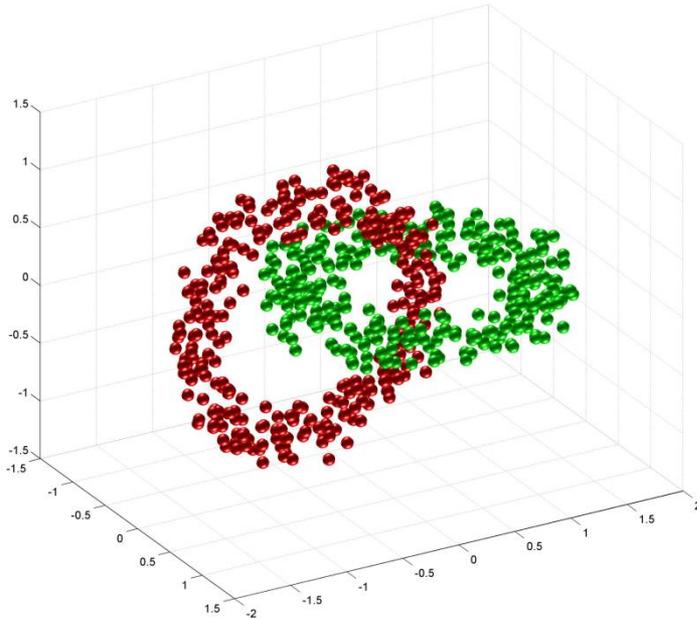
- Iris Setosa
- Iris Virginica
- Iris Versicolor



2.2. Sichtung des Merkmalsraumes

Probleme bei der 2D-Projektion:

„Verschlungene“ Klassenstrukturen schwer erkennbar



Probleme bei 2D-Schnitten:

Meist werden keine Stichprobenpunkte auf dem Schnitt zu liegen kommen.

Abhilfe: Projektion einer Scheibe der Dicke 2ε

2.2. Sichtung des Merkmalsraumes

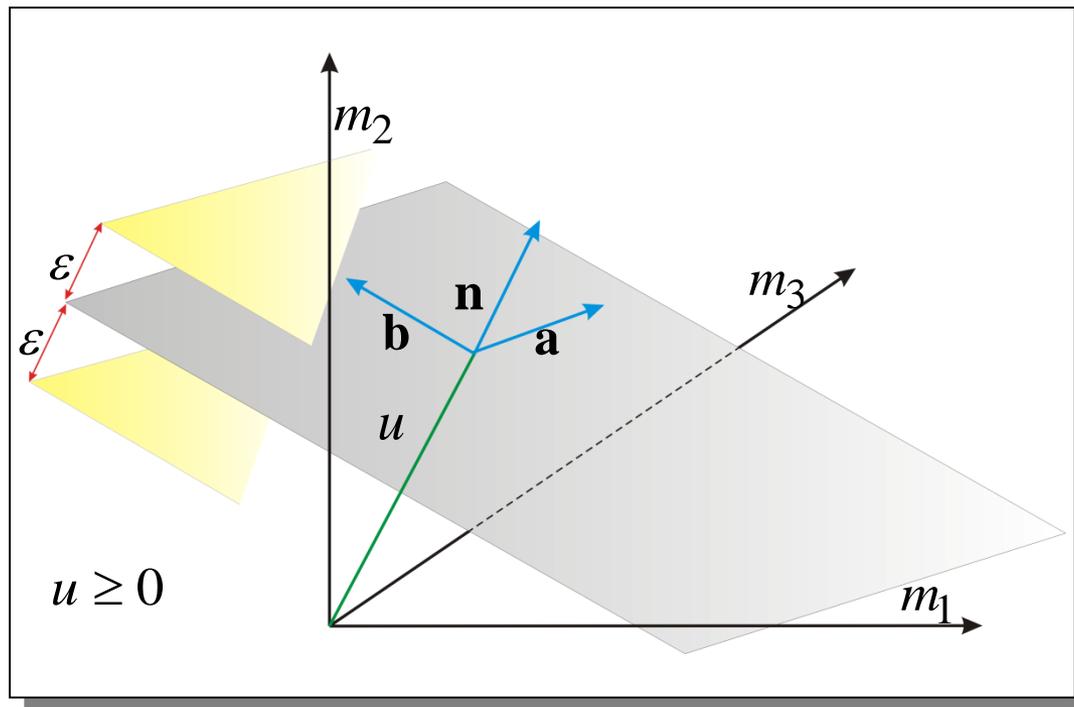
2D-Projektion einer Scheibe der Dicke 2ε

Bsp: 3D \rightarrow 2D Sei $\|\mathbf{a}\|=1$ $\|\mathbf{b}\|=1$ $\mathbf{a}^T \mathbf{b} = 0$ $\mathbf{n} = \mathbf{a} \times \mathbf{b}$

2D-Darstellung durch Projektion auf Ebene $\mathbf{n}^T \mathbf{m} - u = 0$ innerhalb einer Scheibe der Dicke 2ε

\rightarrow 2D-Darstellung der Menge:

$$\left\{ \begin{pmatrix} \mathbf{a}^T \mathbf{m}_i \\ \mathbf{b}^T \mathbf{m}_i \end{pmatrix} \right\} \text{ f\"ur alle } \mathbf{m}_i \text{ mit } \left| \mathbf{n}^T \mathbf{m}_i - u \right| \leq \varepsilon$$



2.2. Sichtung des Merkmalsraumes

2D-Projektion einer Scheibe der Dicke 2ε

Verallgemeinerung für $\mathbf{m}, \mathbf{a}, \mathbf{b} \in \mathbb{R}^d$

Sei $\|\mathbf{a}\|=1$ $\|\mathbf{b}\|=1$ $\mathbf{a} \perp \mathbf{b}$

$$\mathbf{n}_i^T \mathbf{n}_j = \delta_i^j \quad i, j \in \{1, \dots, d-2\} \quad \text{span}(\mathbf{n}_1, \dots, \mathbf{n}_{d-2}) \perp \text{span}(\mathbf{a}, \mathbf{b})$$

$$\text{2D-Ebene im } \mathbb{R}^d : \begin{bmatrix} \mathbf{n}_1^T \mathbf{m} - u_1 = 0 \\ \vdots \\ \mathbf{n}_{d-2}^T \mathbf{m} - u_{d-2} = 0 \end{bmatrix}$$

$\mathbf{n}_j^T \mathbf{m} - u_j = v_j$: Abstand von \mathbf{m} zur 2D-Ebene in Richtung \mathbf{n}_j

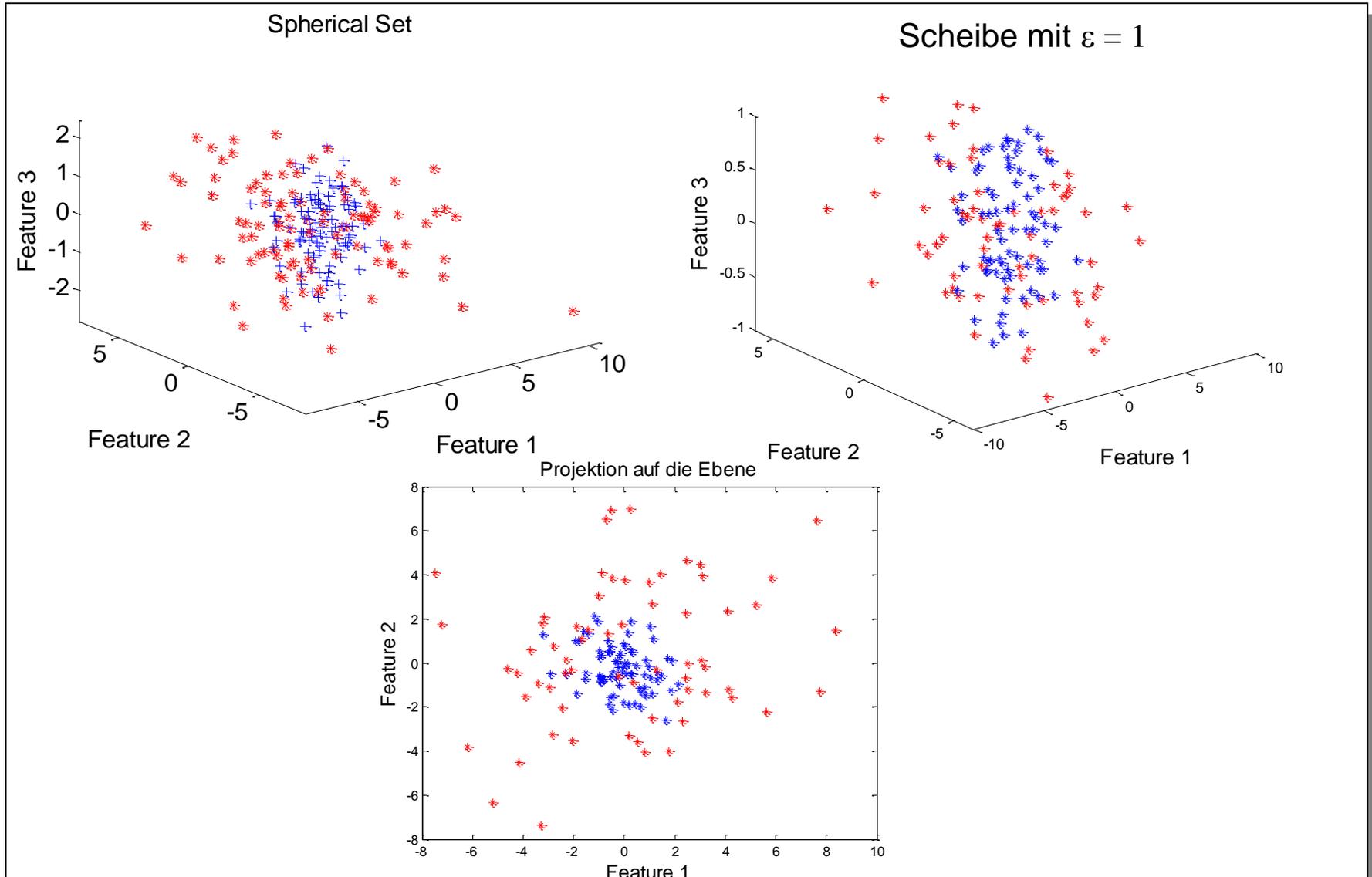
$v = \sqrt{\sum_{j=1}^{d-2} (\mathbf{n}_j^T \mathbf{m} - u_j)^2}$: Abstand von \mathbf{m} zur 2D-Ebene

Lösung: 2D-Darstellung der Menge:

$$\left\{ \begin{pmatrix} \mathbf{a}^T \mathbf{m}_i \\ \mathbf{b}^T \mathbf{m}_i \end{pmatrix} \right\} \quad \text{für alle } \mathbf{m}_i \text{ mit } \sqrt{\sum_{j=1}^{d-2} (\mathbf{n}_j^T \mathbf{m}_i - u_j)^2} \leq \varepsilon$$

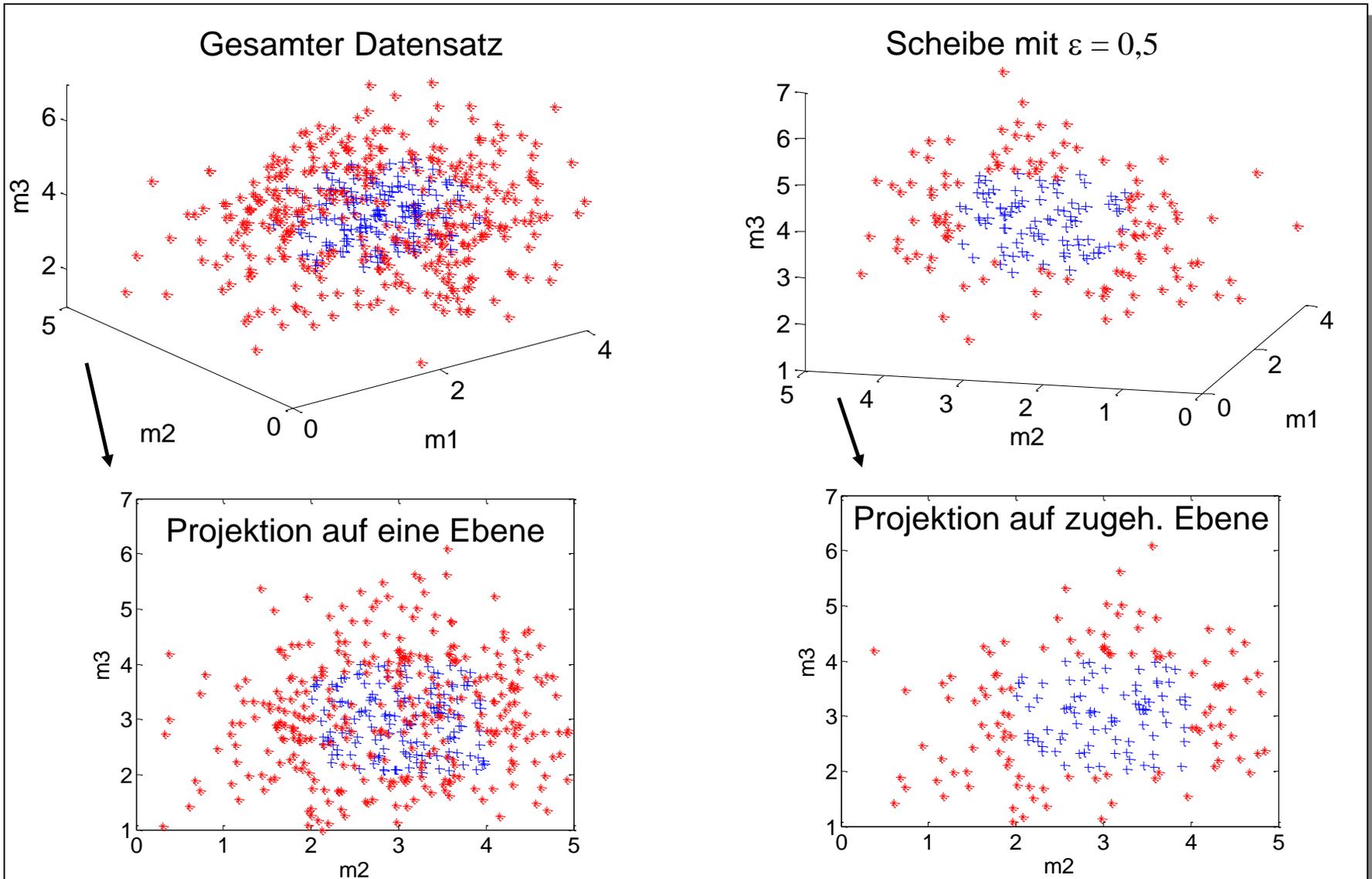
2.2. Sichtung des Merkmalsraumes

Beispiel: Zwei kugelförmig streuende Klassen durchdringen sich.



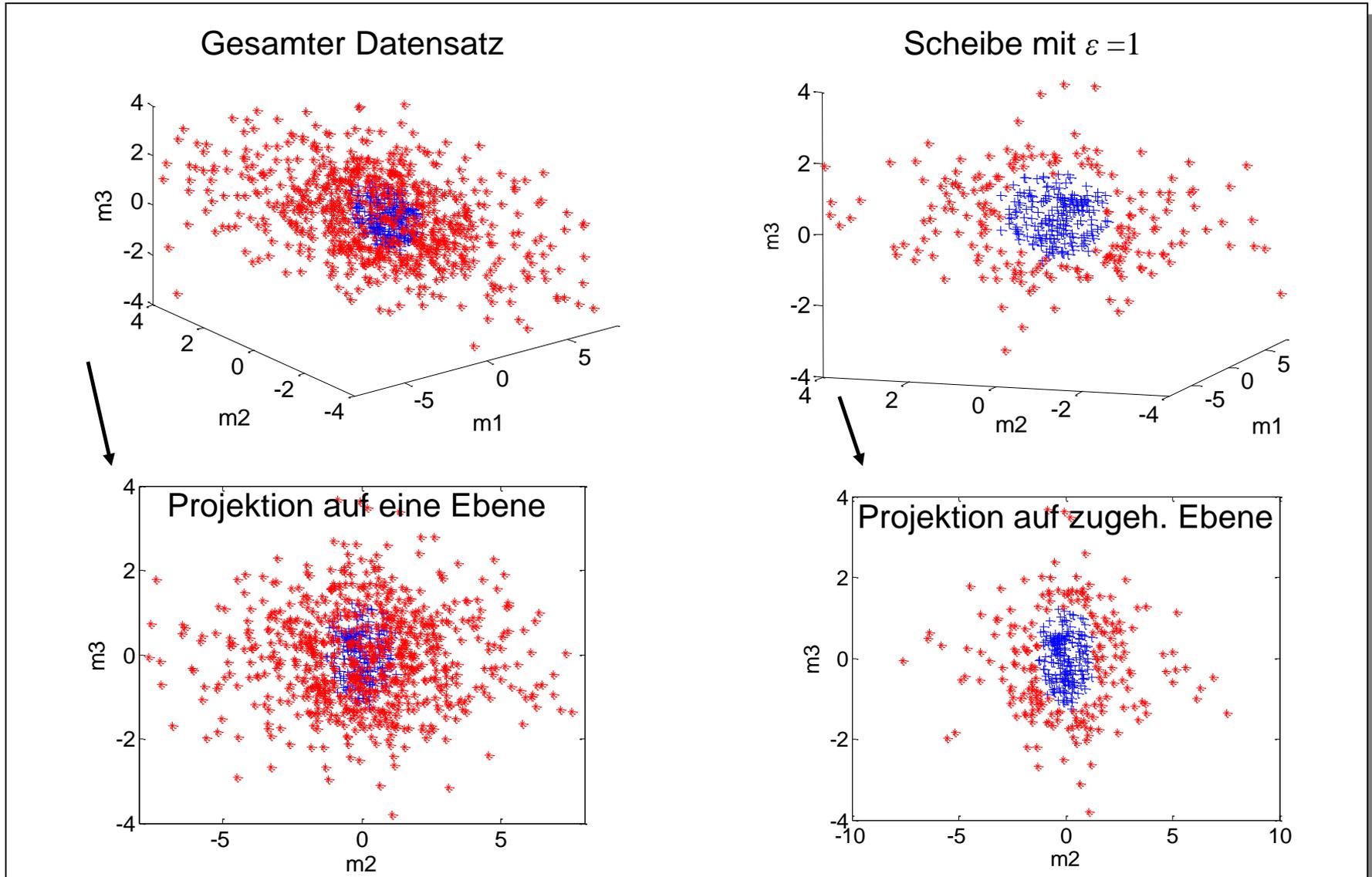
2.2. Sichtung des Merkmalsraumes

Beispiel: Eine Klasse umschließt eine andere.



2.2. Sichtung des Merkmalsraumes

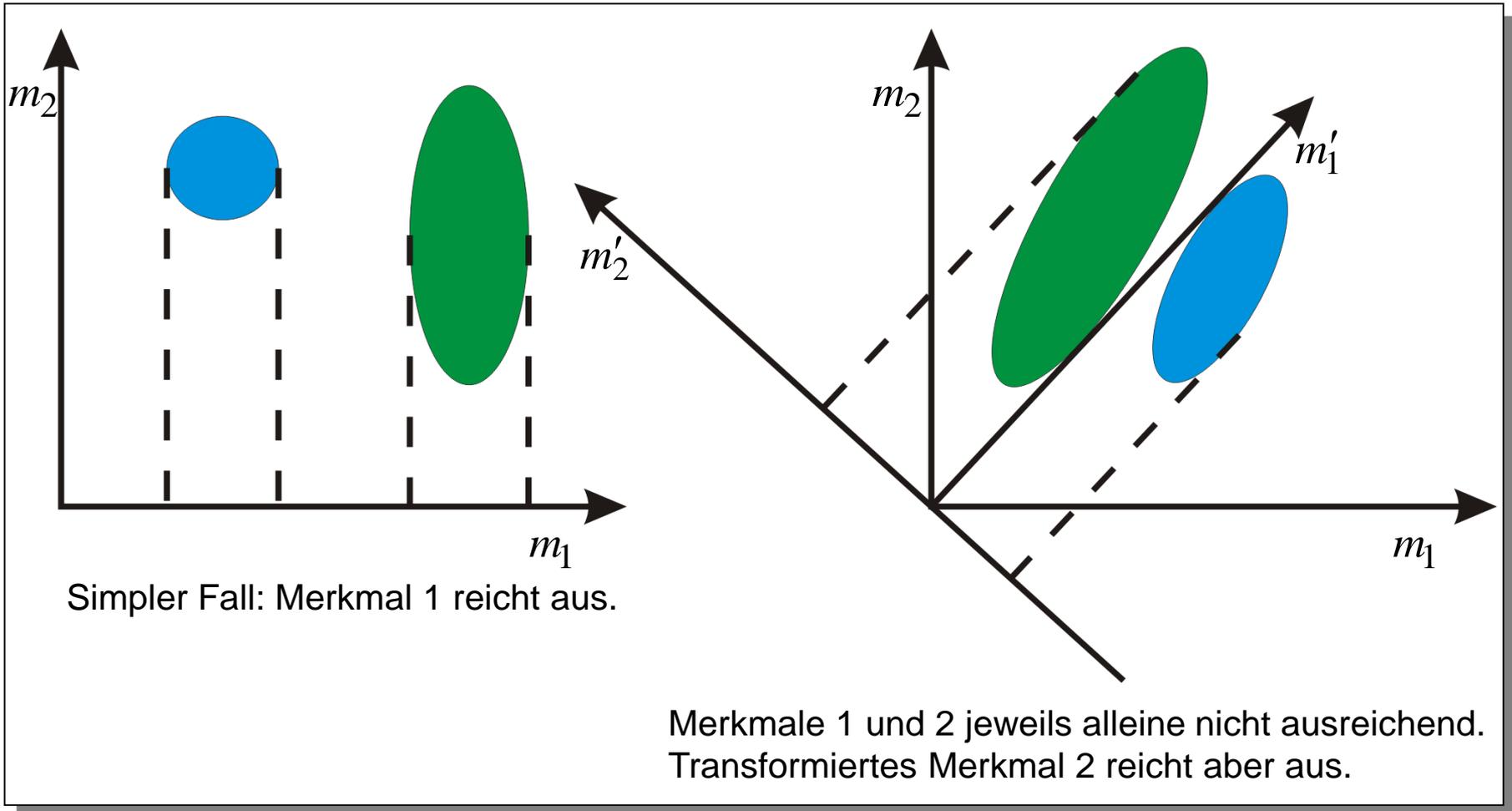
Beispiel: Eine Klasse umschließt eine andere.



2.3. Transformation der Merkmale

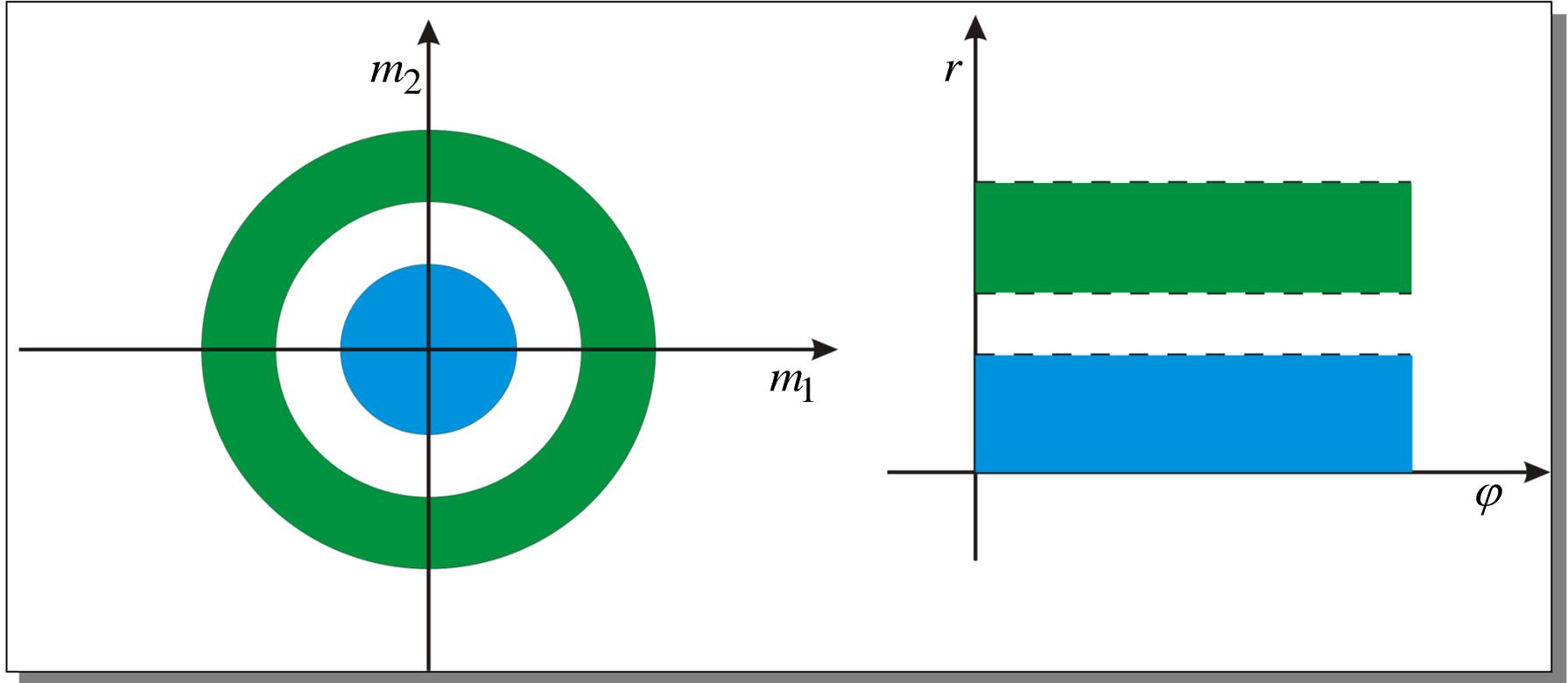
Transformation $\mathbf{m}' = \mathbf{T}\{\mathbf{m}\}$ zur Vereinfachung der Clusterstruktur

Beispiel: Lineare Koordinatentransformation



2.3. Transformation der Merkmale

Beispiel: Nichtlineare Koordinatentransformation.



$$m_1' := r = (m_1^2 + m_2^2)^{1/2}$$

$$m_2' := \varphi = \arctan(m_2 / m_1)$$

2.4. Abstandsmessung im Merkmalsraum

Wunsch: Quantitative Messung des Abstandes (des Unterschiedes) zwischen den Merkmalsvektoren zweier Objekte

Probleme:

- unterschiedliche **Einheiten** der Merkmale
- unterschiedliche **Bedeutungen** der Merkmale
- unterschiedliche **Skalen** der Merkmale
- unterschiedliche **Größenordnungen** der Merkmale

Möglicher Ansatz:

$$D(\mathbf{m}, \mathbf{m}') = \sum_{i=1}^d \alpha_i D_i(m_i, m'_i) \quad \text{mit} \quad \alpha_i > 0, \quad \sum_{i=1}^d \alpha_i = 1$$

Viele Freiheitsgrade (Design-Parameter)

2.4. Abstandsmessung im Merkmalsraum

Normaxiome:

Positivität:

$$\|\mathbf{m}\| \geq 0$$

$$\|\mathbf{m}\| = 0 \iff \mathbf{m} = \mathbf{0}$$

Homogenität:

$$\|a\mathbf{m}\| = |a| \cdot \|\mathbf{m}\|$$

Dreiecksungleichung:

$$\|\mathbf{m} + \mathbf{m}'\| \leq \|\mathbf{m}\| + \|\mathbf{m}'\|$$

Nur für mindestens verhältnisskalierte Merkmale!

2.4. Abstandsmessung im Merkmalsraum

Normen für quantitativ skalierte Merkmale:

- Minkowski Norm

$$\|\mathbf{m}\|_M = \left(\sum_{i=1}^d |m_i|^r \right)^{\frac{1}{r}}$$

- Euklidische Norm

$$\|\mathbf{m}\|_e = \sqrt{\sum_{i=1}^d (m_i)^2}$$

- Chebychev Norm

$$\|\mathbf{m}\|_C = \max_i \{|m_i|\}$$

- Mahalanobis Norm

$$\|\mathbf{m}\|_m = (\mathbf{m}^T \mathbf{A} \mathbf{m})^{1/2}$$

A positiv definit

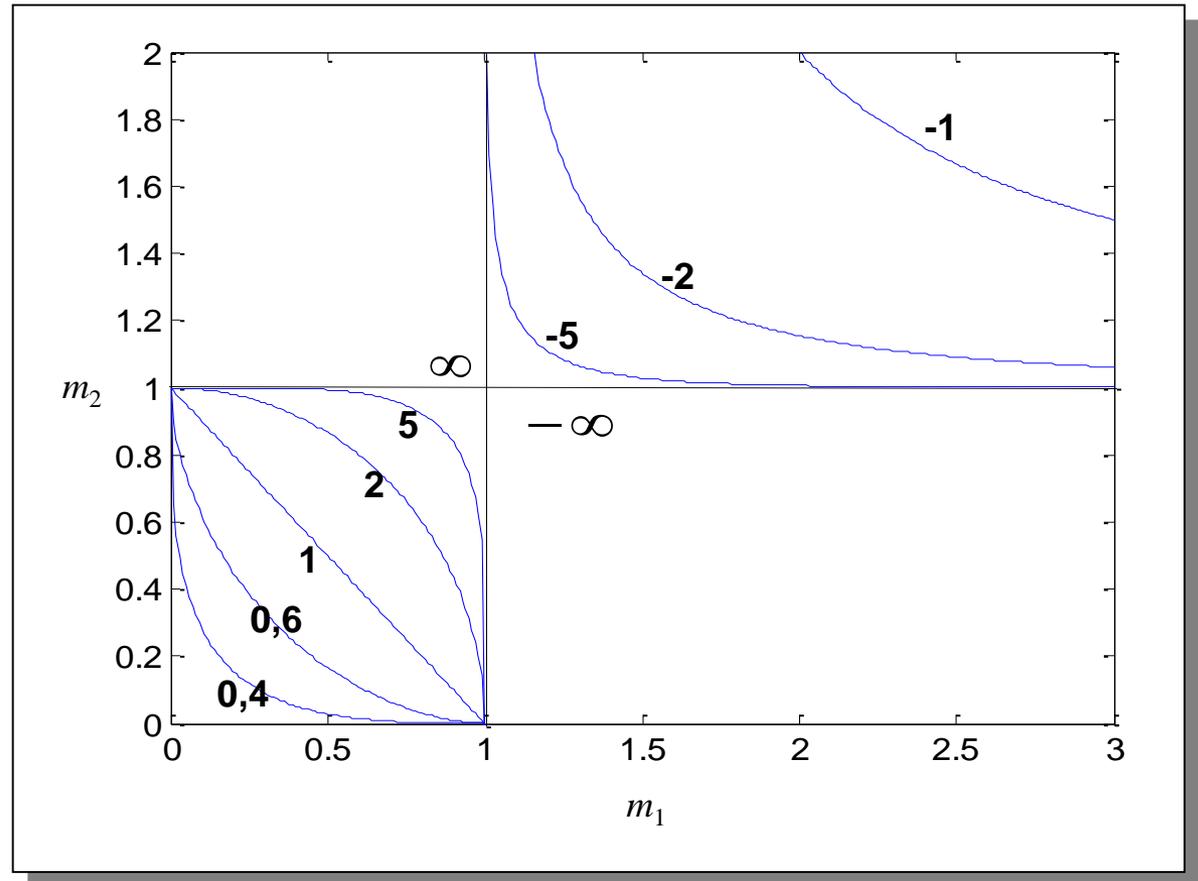
Ergänzen: Mahalanobis Norm ist **A** eine Kovarianzmatrix

2.4. Abstandsmessung im Merkmalsraum

Minkowski Norm für verschiedene Parameter r

Bsp.: $d = 2$, Punkte mit Norm = 1 d.h. „Einheitskreise“

$$\|\mathbf{m}\|_M = \left(\sum_{i=1}^2 |m_i|^r \right)^{1/r}$$



Quelle: J. P. Marques de Sá, Pattern Recognition

2.4. Abstandsmessung im Merkmalsraum

Metriken:

- Positivität $D(\mathbf{x}, \mathbf{y}) \geq 0$
- Reflexivität $D(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$
- Symmetrie $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y}, \mathbf{x})$
- Dreiecksungleichung $D(\mathbf{x}, \mathbf{z}) \leq D(\mathbf{x}, \mathbf{y}) + D(\mathbf{y}, \mathbf{z})$

Methode zur Definition einer Metrik D auf der Basis einer Norm $\| \cdot \|$:

$$D(\mathbf{m}_i, \mathbf{m}_j) := \|\mathbf{m}_i - \mathbf{m}_j\|$$

2.4. Abstandsmessung im Merkmalsraum

Tanimoto Metrik als Distanz zwischen zwei Mengen.

$$D_{Tanimoto}(S_1, S_2) = \frac{N_1 + N_2 - 2N_{12}}{N_1 + N_2 - N_{12}}$$

S_1, S_2 : Mengen

N_1, N_2 : Zahl der Mengenelemente

N_{12} : Zahl der gleichen Elemente

- Geeignet für nominale Merkmale.
- Wertebereich: rationale Zahlen im Intervall $[0, 1]$

Beispiel:

Die Arbeitsgruppe 1 (AG 1) hat die Mitglieder: Petra, Fritz, Franz und Hans.

Die AG 2 hat die Mitglieder: Anne, Barbara, Fritz, Hans und Klaus.

$$N_1 = 4$$

$$N_2 = 5$$

$$N_{12} = 2$$

$$D_{Tanimoto}(AG\ 1, AG\ 2) = 5/7$$

2.4. Abstandsmessung im Merkmalsraum

Tanimoto „Metrik“ erfüllt die Dreieckungleichung

$$D(S_1, S_3) \leq D(S_1, S_2) + D(S_2, S_3)$$

Trivial:

$$D(S_1, S_3) \leq D(S_2, S_3) \text{ oder}$$

$$D(S_1, S_3) \leq D(S_1, S_2) \text{ oder}$$

$$N_2 = 0$$

zu beweisen wenn: $D(S_1, S_3) > D(S_1, S_2)$ und $D(S_1, S_3) > D(S_2, S_3)$ und $N_2 > 0$

$$D(S_m, S_n) = \frac{N_m + N_n - 2N_{mn}}{N_m + N_n - N_{mn}} = 1 - \frac{N_{mn}}{N_m + N_n - N_{mn}} = 1 - M_{mn} \quad m, n \in \{1, 2, 3\}$$

$$M_{mn} = \frac{N_{mn}}{N_m + N_n - N_{mn}} \iff (N_m + N_n - N_{mn})M_{mn} = N_{mn} \iff$$

$$\iff (N_m + N_n)M_{mn} = N_{mn}(1 + M_{mn}) \iff N_{mn} = \frac{M_{mn}}{1 + M_{mn}}(N_m + N_n) \quad (1)$$

$$D(S_1, S_3) > D(S_1, S_2) \iff 1 - M_{13} > 1 - M_{12} \iff M_{12} > M_{13} \quad (2)$$

$$D(S_1, S_3) > D(S_2, S_3) \iff 1 - M_{13} > 1 - M_{23} \iff M_{23} > M_{13} \quad (3)$$

2.4. Abstandsmessung im Merkmalsraum

$$\left. \begin{array}{l} N_{13} \leq \min\{N_1, N_3\} \\ N_{12} \leq \min\{N_1, N_2\} \\ N_{23} \leq \min\{N_2, N_3\} \end{array} \right\} \Rightarrow N_2 - N_{23} - N_{12} + N_{13} \geq 0 \quad (1) \quad \Rightarrow$$

$$N_2 - \frac{M_{23}}{1+M_{23}}(N_2 + N_3) - \frac{M_{12}}{1+M_{12}}(N_1 + N_2) + \frac{M_{13}}{1+M_{13}}(N_1 + N_3) \geq 0 \Leftrightarrow$$

$$\Leftrightarrow N_2 \left(1 - \frac{M_{23}}{1+M_{23}} - \frac{M_{12}}{1+M_{12}} \right) \geq \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_1 + \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_3 \quad (4)$$

$$N_1 \geq N_{12} \Leftrightarrow N_1 \geq \frac{M_{12}}{1+M_{12}}(N_1 + N_2) \Leftrightarrow N_1 \left(1 - \frac{M_{12}}{1+M_{12}} \right) \geq \frac{M_{12}}{1+M_{12}} N_2 \Leftrightarrow$$

$$\Leftrightarrow N_1 \frac{1}{1+M_{12}} \geq \frac{M_{12}}{1+M_{12}} N_2 \Leftrightarrow N_1 \geq M_{12} N_2 \quad (5)$$

$$\text{analog } N_3 \geq N_{23} \Leftrightarrow N_3 \geq M_{23} N_2 \quad (6)$$

2.4. Abstandsmessung im Merkmalsraum

$$\left. \begin{array}{l} (5) \\ (2) \end{array} \right\} \Rightarrow \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_1 \geq M_{12} \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_2$$

$$\left. \begin{array}{l} (6) \\ (3) \end{array} \right\} \Rightarrow \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_3 \geq M_{23} \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_2$$

$$\left. \begin{array}{l} \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_1 + \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_3 \geq \\ \geq M_{12} \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_2 + M_{23} \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_2 \end{array} \right\} \Rightarrow$$

(4)

$$\left(1 - \frac{M_{12}}{1+M_{12}} - \frac{M_{23}}{1+M_{23}} \right) N_2 \geq M_{12} \left(\frac{M_{12}}{1+M_{12}} - \frac{M_{13}}{1+M_{13}} \right) N_2 + M_{23} \left(\frac{M_{23}}{1+M_{23}} - \frac{M_{13}}{1+M_{13}} \right) N_2$$

$$\Leftrightarrow 1 \geq \frac{M_{12} + M_{12}^2}{1+M_{12}} + \frac{M_{23} + M_{23}^2}{1+M_{23}} - M_{13} \left(\frac{M_{12} + M_{23}}{1+M_{13}} \right) \Leftrightarrow 1 + M_{13} \geq M_{12} + M_{23}$$

$$\Leftrightarrow 1 - M_{13} \leq 1 - M_{12} + 1 - M_{23} \Leftrightarrow D(S_1, S_3) \leq D(S_1, S_2) + D(S_2, S_3) \quad \text{q.e.d.}$$

2.4. Abstandsmessung im Merkmalsraum

Eine Metrik für ordinale Merkmale

A B C D E F G H I ...

$m, m' \in \{A, B, C, D, E, F, \dots\}$

$D(m, m') :=$ Minimale Zahl der Vertauschungen von Nachbarn, um von m auf m' zu kommen.

Beispiele: $D(A, D) = 3$

$D(A, A) = 0$

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz

Für kontinuierliche Merkmale:

$$D(p \parallel p') = \int \dots \int_{\text{supp}\{p(\mathbf{m})\}} p(\mathbf{m}) \log \left(\frac{p(\mathbf{m})}{p'(\mathbf{m})} \right) d\mathbf{m}$$

$p(\cdot)$, $p'(\cdot)$: Wahrscheinlichkeitsdichtefunktionen

Für diskrete Merkmale:

$$D(P \parallel P') = \sum_{\text{supp}\{P(\mathbf{m})\}} P(\mathbf{m}) \log \frac{P(\mathbf{m})}{P'(\mathbf{m})}$$

$P(\cdot)$, $P'(\cdot)$: Wahrscheinlichkeiten

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz (KL-Distanz)

Eigenschaften:

- Gültig für **nominal-, ordinal-, intervall-, verhältnis- und absolut skalierte Merkmale**
- Werte $\in [0, \infty)$
- Asymmetrie $D(p \parallel p') \neq D(p' \parallel p)$
- Dreiecksungleichung wird nicht erfüllt.
- $D(p \parallel p') = 0 \iff p(m) = p'(m)$ fast überall
- majorisiert Minkowski Metrik für $r = 1$ im Sinne von:

$$D(p \parallel p') \geq \frac{1}{2 \ln e} \left(\int_{\text{supp}\{p\} \cup \text{supp}\{p'\}} \dots \int |p(\mathbf{m}) - p'(\mathbf{m})| d\mathbf{m} \right)^2$$

- invariant bezüglich bijektiver Abbildungen des Merkmalsvektors

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz

$$D(p \parallel p') \geq 0$$

Beweisskizze:

$$Y_p = \{\mathbf{m} \mid p(\mathbf{m}) > 0\} =: \text{supp}\{p(\mathbf{m})\}$$

Jensensche Ungleichung und
 $-\log(\mathbf{m})$ ist konvex. 

Jensensche Ungleichung:

für $f(\cdot)$ konvex gilt:

$$E\{f(\mathbf{m})\} \geq f(E\{\mathbf{m}\})$$

$$\begin{aligned} D(p \parallel p') &= \int_{Y_p} p(\mathbf{m}) \log \frac{p(\mathbf{m})}{p'(\mathbf{m})} d\mathbf{m} = \\ &= \int_{Y_p} p(\mathbf{m}) \left(-\log \frac{p'(\mathbf{m})}{p(\mathbf{m})} \right) d\mathbf{m} \geq \\ &\geq -\log \int_{Y_p} p(\mathbf{m}) \frac{p'(\mathbf{m})}{p(\mathbf{m})} d\mathbf{m} = \\ &= -\log \int_{Y_p} p'(\mathbf{m}) d\mathbf{m} \geq \\ &\geq -\log \int_{Y_{p'}} p'(\mathbf{m}) d\mathbf{m} = \\ &= -\log 1 = 0 \end{aligned}$$

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz

Konvexität: $D(\lambda p_1 + (1-\lambda)p_2 \parallel \lambda p'_1 + (1-\lambda)p'_2) \leq \lambda D(p_1 \parallel p'_1) + (1-\lambda)D(p_2 \parallel p'_2)$

Beweisskizze:

$\forall \lambda \in [0,1]$

$f(t) = t \log t$ konvex für $\forall t \in (0, \infty)$

$$\begin{aligned} \implies f(\mu t_1 + (1-\mu)t_2) &\leq \mu f(t_1) + (1-\mu)f(t_2) \\ (\mu t_1 + (1-\mu)t_2) \log(\mu t_1 + (1-\mu)t_2) &\leq \mu t_1 \log(t_1) + (1-\mu)t_2 \log(t_2) \quad \forall \mu \in [0,1] \end{aligned}$$

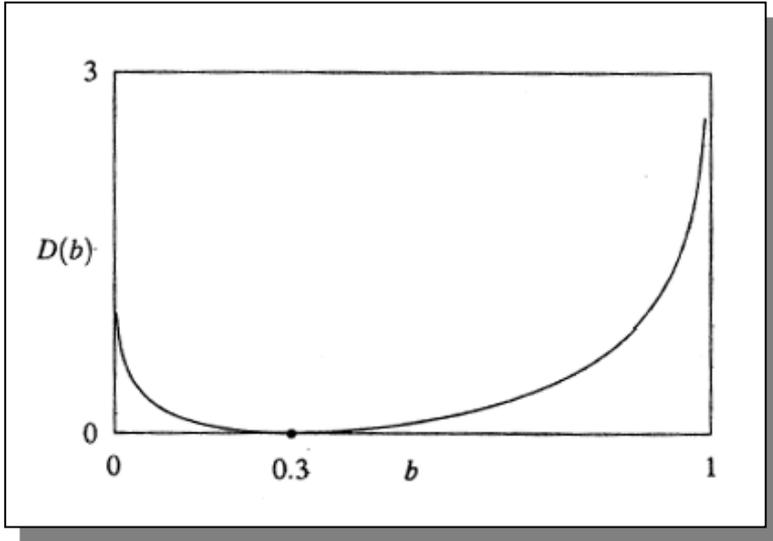
$$t_1 = \frac{\lambda p_1(\mathbf{m})}{\lambda p'_1(\mathbf{m})}, \quad t_2 = \frac{(1-\lambda)p_1(\mathbf{m})}{(1-\lambda)p'_1(\mathbf{m})}, \quad \mu = \frac{\lambda p'_1(\mathbf{m})}{\lambda p'_1(\mathbf{m}) + (1-\lambda)p'_2(\mathbf{m})} \quad \forall \lambda \in [0,1]$$

$$\begin{aligned} \implies (\lambda p_1(\mathbf{m}) + (1-\lambda)p_2(\mathbf{m})) \log \frac{\lambda p_1(\mathbf{m}) + (1-\lambda)p_2(\mathbf{m})}{\lambda p'_1(\mathbf{m}) + (1-\lambda)p'_2(\mathbf{m})} &\leq \\ \leq \lambda p_1(\mathbf{m}) \log \frac{p_1(\mathbf{m})}{p'_1(\mathbf{m})} + (1-\lambda)p_2(\mathbf{m}) \log \frac{p_2(\mathbf{m})}{p'_2(\mathbf{m})} &\quad \int \end{aligned}$$

q.e.d.

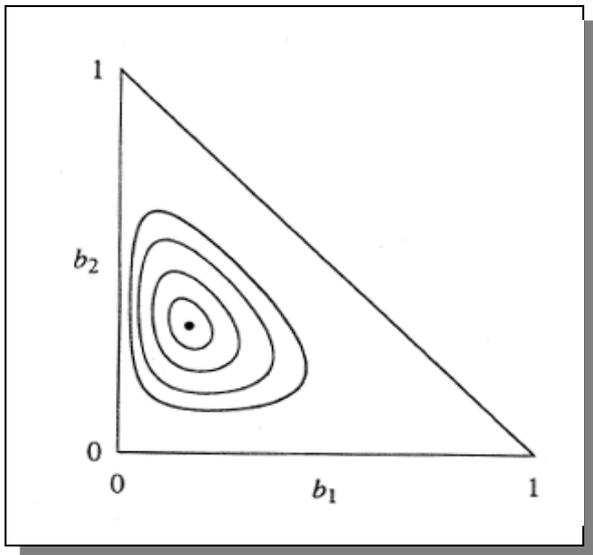
2.4. Abstandsmessung im Merkmalsraum

Kullback-Leibler Distanz, diskrete Beispiele:



$$D(\{a, 1-a\} \parallel \{b, 1-b\}) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}$$

$$a = 0,3$$



$$D(\{a_1, a_2, a_3\} \parallel \{b_1, b_2, b_3\}) = a_1 \log \frac{a_1}{b_1} + a_2 \log \frac{a_2}{b_2} + a_3 \log \frac{a_3}{b_3}$$

$$a_1 = 1/6$$

$$a_2 = 2/6$$

$$a_3 = 3/6$$

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz, kontinuierliches Beispiel:

$$p(m|\omega_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(m-\mu_1)^2\right) \quad p(m|\omega_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(m-\mu_2)^2\right)$$

$$D(p(m|\omega_1) \| p(m|\omega_2)) = \int p(m|\omega_1) \log \frac{(2\pi\sigma_1^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_1^2}(m-\mu_1)^2\right)}{(2\pi\sigma_2^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_2^2}(m-\mu_2)^2\right)} dm =$$

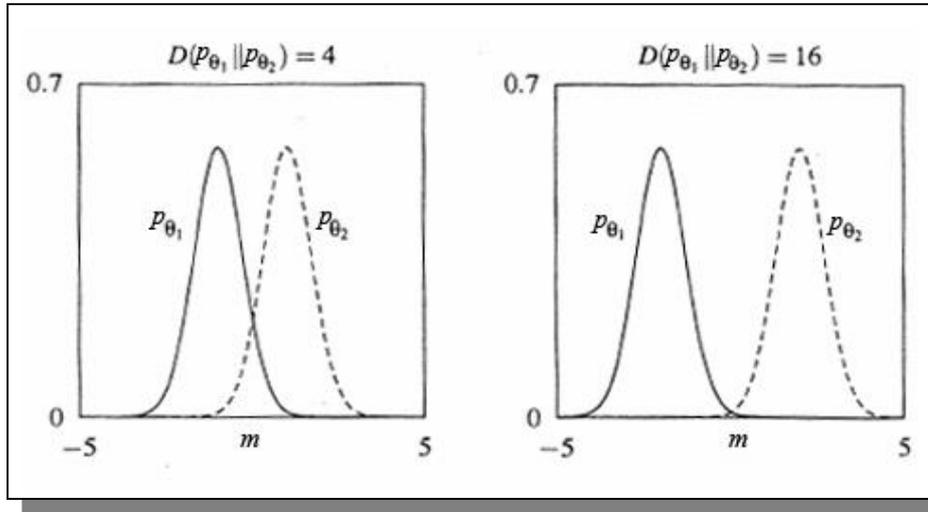
$$= -\frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2} - \frac{1}{2\sigma_1^2} \int p(m|\omega_1)(m-\mu_1)^2 dm + \frac{1}{2\sigma_2^2} \int p(m|\omega_1)(m-\mu_2)^2 dm =$$

$$= -\frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2} - \frac{1}{2\sigma_1^2} \sigma_1^2 + \frac{1}{2\sigma_2^2} (\sigma_1^2 + (\mu_1 - \mu_2)^2) =$$

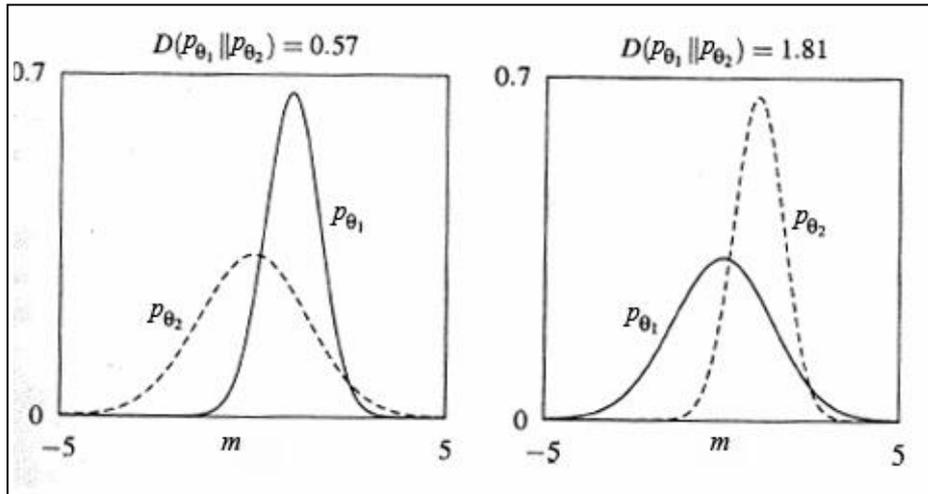
$$= \frac{1}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} - \log \frac{\sigma_1^2}{\sigma_2^2} - 1 \right) + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}$$

2.4. Abstandsmessung im Merkmalsraum

Die Kullback-Leibler Distanz, kontinuierliches Beispiel:



$$\sigma_1^2 = \sigma_2^2 \Rightarrow D(p_{\theta_1} \| p_{\theta_2}) = D(p_{\theta_2} \| p_{\theta_1})$$

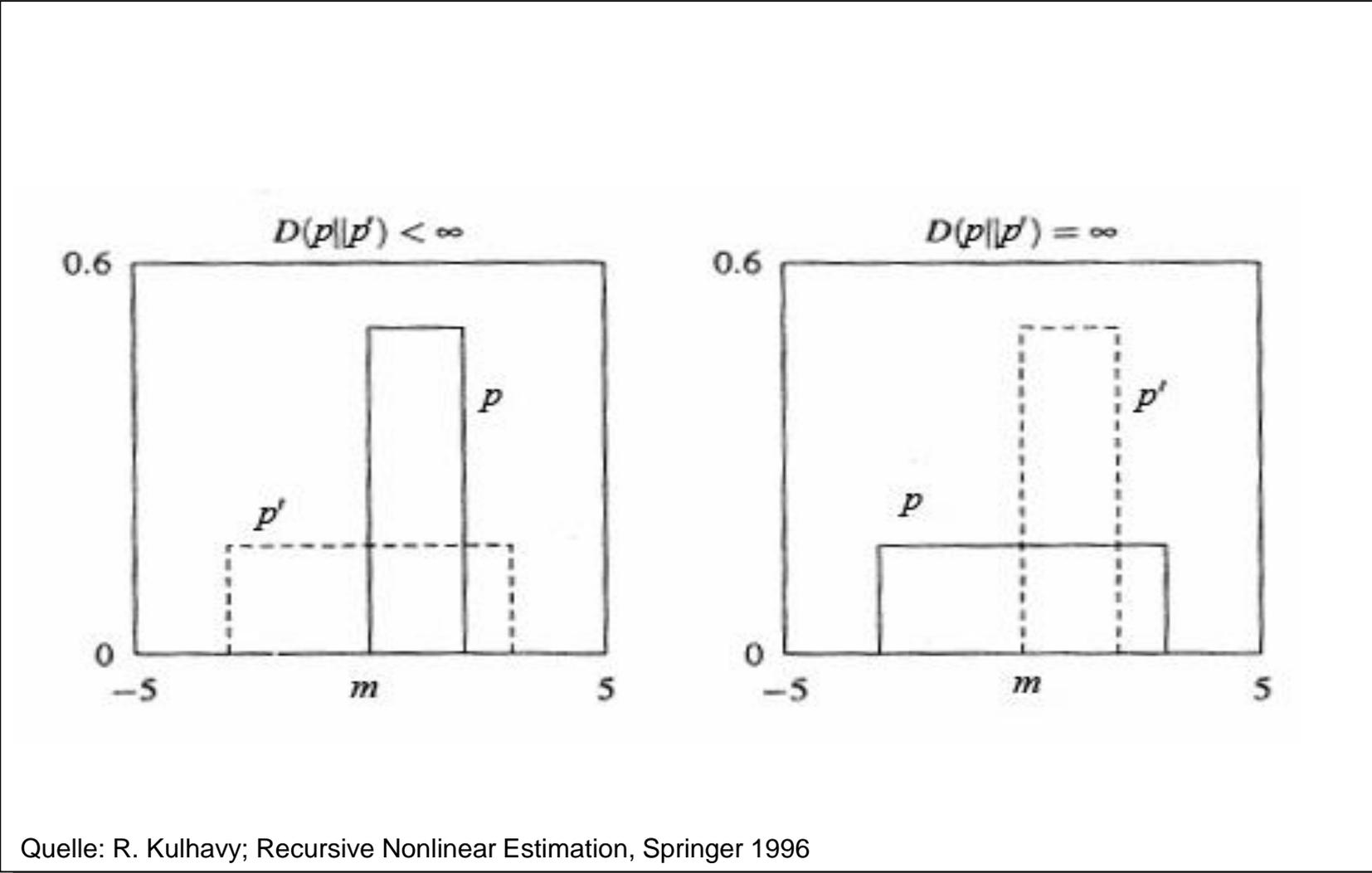


$$\sigma_1^2 \neq \sigma_2^2$$

Quelle: R. Kulhavy; Recursive Nonlinear Estimation, Springer 1996

2.4. Abstandsmessung im Merkmalsraum

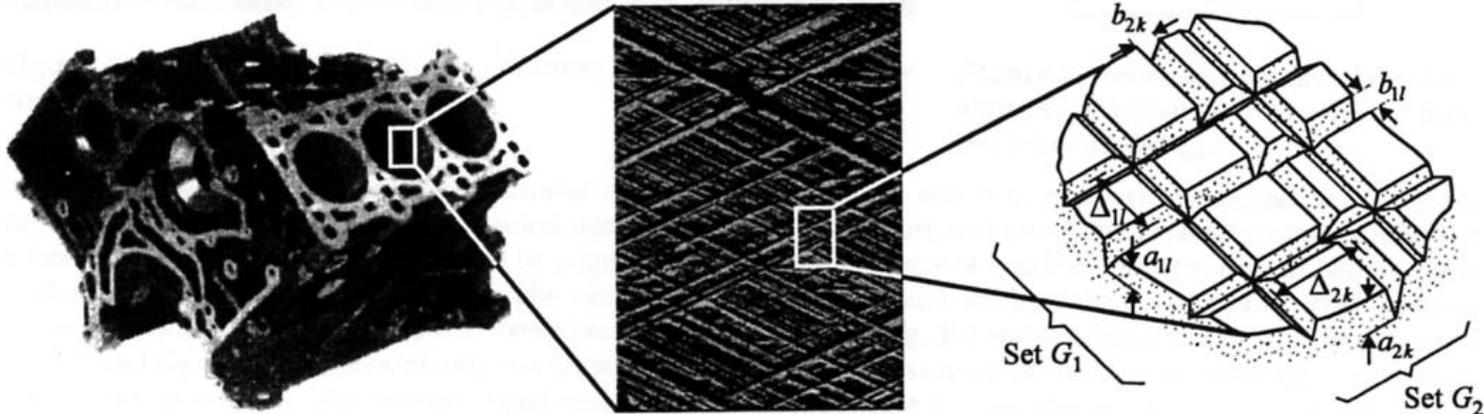
Die Kullback-Leibler Distanz, kontinuierliches Beispiel:



Quelle: R. Kulhavy; Recursive Nonlinear Estimation, Springer 1996

2.4. Abstandsmessung im Merkmalsraum

Beispiel: Konstruktion eines Abstandsmaßes mit Hilfe der KL-Distanz



Quelle: D. Krahe, J. Beyerer: A Parametric Method to Quantify the Balance of Groove Sets of Honed Cylinder Bores

$$\mathbf{u} := \begin{pmatrix} a \\ b \end{pmatrix} \quad \begin{array}{l} a: \text{Amplitude} \\ b: \text{Breite} \end{array} \quad \Delta: \text{Riefenabstand}$$

Modell für die Verteilung der Größen a , b , Δ für beide Riefenscharen:

$$p_i(\mathbf{u}, \Delta) = \underbrace{\lambda_i \exp(-\lambda_i \Delta)}_{\text{Exponentialverteilung}} \underbrace{\frac{1}{2\pi \sqrt{\det(\mathbf{C}_i)}} \exp\left(-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1}(\mathbf{u} - \boldsymbol{\mu}_i)\right)}_{\text{Gaussverteilung}} \quad i = 1, 2$$

Exponentialverteilung

Gaussverteilung

2.4. Abstandsmessung im Merkmalsraum

Beispiel: Konstruktion eines Abstandsmaßes mit Hilfe der KL-Distanz

$$\boldsymbol{\mu}_i = \mathbf{E}_i \{ \mathbf{u} \} = \int \int \int \mathbf{u} p_i(\mathbf{u}, \Delta) d\mathbf{u} d\Delta = \begin{pmatrix} \mu_{a_i} \\ \mu_{b_i} \end{pmatrix}$$

$$\mathbf{C}_i = \mathbf{E}_i \left\{ (\mathbf{u} - \boldsymbol{\mu}_i)(\mathbf{u} - \boldsymbol{\mu}_i)^T \right\} = \begin{pmatrix} \sigma_{a_i}^2 & \rho_i \sigma_{a_i} \sigma_{b_i} \\ \rho_i \sigma_{a_i} \sigma_{b_i} & \sigma_{b_i}^2 \end{pmatrix}$$

$$\rho_i = \frac{\mathbf{E} \left\{ (a_{ij} - \mu_{a_i})(b_{ij} - \mu_{b_i}) \right\}}{\sigma_{a_i} \sigma_{b_i}}$$

$$\lambda_i = \frac{1}{\mathbf{E} \{ \Delta_{ij} \}} : \text{Riefendichte}$$

2.4. Abstandsmessung im Merkmalsraum

Beispiel: Konstruktion eines Abstandsmaßes mit Hilfe der KL-Distanz

KL-Distanz:

$$D(p_1 \parallel p_2) = \int \int \int p_1(\mathbf{u}, \Delta) \ln \frac{p_1(\mathbf{u}, \Delta)}{p_2(\mathbf{u}, \Delta)} d\mathbf{u} d\Delta$$

Das Maß soll symmetrisch sein:

$$D_{ab\Delta} := D(p_1 \parallel p_2) + D(p_2 \parallel p_1) = \int \int \int p_1 \ln \frac{p_1}{p_2} d\mathbf{u} d\Delta + \int \int \int p_2 \ln \frac{p_2}{p_1} d\mathbf{u} d\Delta$$

$$D_{ab\Delta} = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) +$$

Mahalanobis Distanz zw. $\boldsymbol{\mu}_1$ u. $\boldsymbol{\mu}_2$

$$+ \left(\frac{1}{2} \text{tr}(\mathbf{C}_1^{-1} \mathbf{C}_2 + \mathbf{C}_2^{-1} \mathbf{C}_1) - 2 \right) +$$

Distanz der Kovarianzmatrizen

$$+ \left(\frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} - 2 \right)$$

Distanz der Riefendichten

2.4. Abstandsmessung im Merkmalsraum – Tangentendistanz

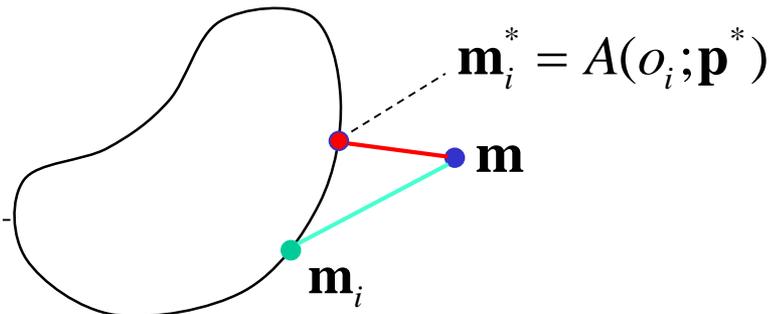
Ziel: Abstandsmaß, das bezüglich **kleiner**, systematischer Variationen robust, d.h. unempfindlich ist.

$\mathbf{m}_i \in D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$ Lernstichprobe

Ansatz: $\mathbf{m}_i = A(o_i; \mathbf{p})$ $\mathbf{p} \in \Pi \subseteq \mathbb{R}^q$ $\mathbf{m}_i \in \mathbb{R}^d$ o_i : Beispielobjekt Nr. i

q -dimensionale
differenzierbare
Mannigfaltigkeit

$\{A(o_i; \mathbf{p}) \mid \mathbf{p} \in \Pi\}$



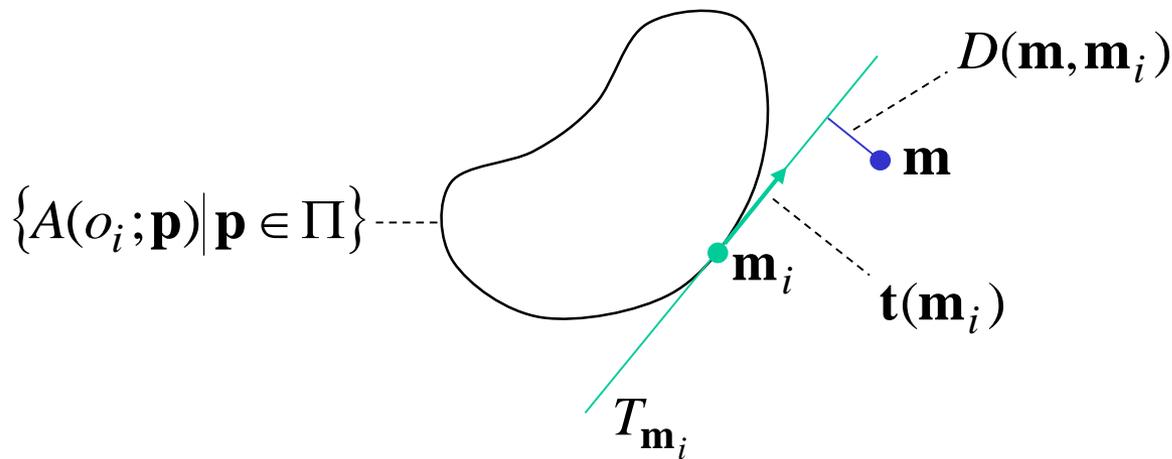
Ansatz: Messung des Abstandes zur mit \mathbf{m}_i assoziierten Mannigfaltigkeit anstatt zum \mathbf{m}_i Punkt direkt.

2.4. Abstandsmessung im Merkmalsraum – Tangentendistanz

Ziel: Abstandsmaß, das bezüglich kleiner, systematischer Variationen robust, d.h. unempfindlich ist.

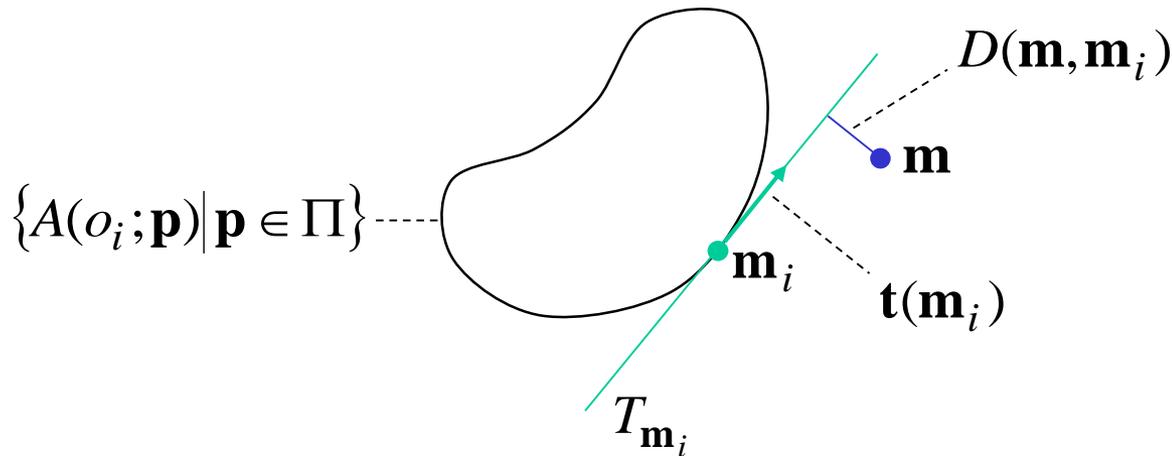
$\mathbf{m}_i \in D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$ Lernstichprobe

Ansatz: $\mathbf{m}_i = A(o_i; \mathbf{p}) \quad \mathbf{p} \in \Pi \subseteq \mathbb{R}^q \quad \mathbf{m}_i \in \mathbb{R}^d \quad o_i$: Beispielobjekt Nr. i



Tangentialraum: $T_{\mathbf{m}_i} = \mathbf{m}_i + \text{span}(\mathbf{t}_1(\mathbf{m}_i), \dots, \mathbf{t}_q(\mathbf{m}_i))$

2.4. Abstandsmessung im Merkmalsraum – Tangentendistanz



Tangentendistanz:

$$D(\mathbf{m}, \mathbf{m}_i) := \min_{\{\mathbf{a} \mid \|\mathbf{a}\| \leq \varepsilon\}} \|\mathbf{m} - (\mathbf{m}_i + \mathbf{T}_{\mathbf{m}_i} \mathbf{a})\|$$

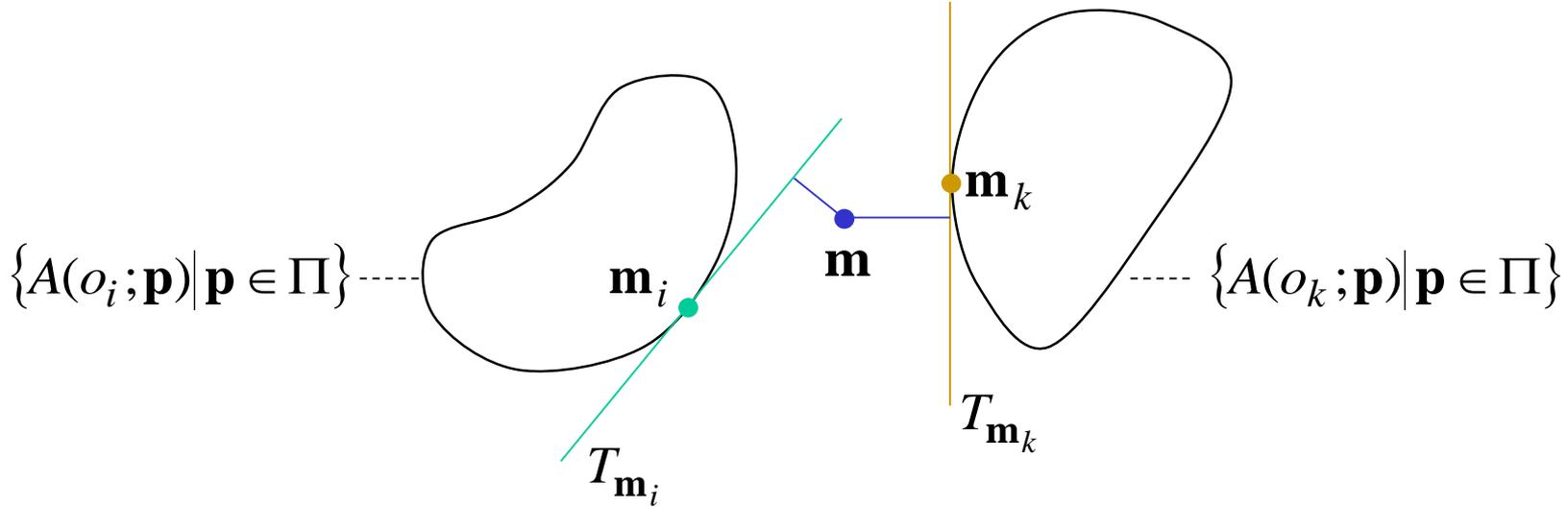
$$\mathbf{T}_{\mathbf{m}_i} := [\mathbf{t}_1(\mathbf{m}_i), \mathbf{K}, \mathbf{t}_q(\mathbf{m}_i)] \quad \det(\mathbf{T}_{\mathbf{m}_i}^T \mathbf{T}_{\mathbf{m}_i}) \neq 0$$

Verwendet man für $\|\cdot\|$ die Euklidische Norm, führt die Minimierung bezüglich \mathbf{a} auf ein quadratisches Optimierungsproblem, das mit Standardverfahren gelöst werden kann.

$$\text{Sekantennäherung: } \hat{\mathbf{t}}_j(\mathbf{m}_i) = \mathbf{m}_i - A(o_i; \mathbf{p} + \Delta \mathbf{p}_j) \quad \det[\Delta \mathbf{p}_1, \mathbf{K}, \Delta \mathbf{p}_q] \neq 0$$

2.4. Abstandsmessung im Merkmalsraum – Tangentendistanz

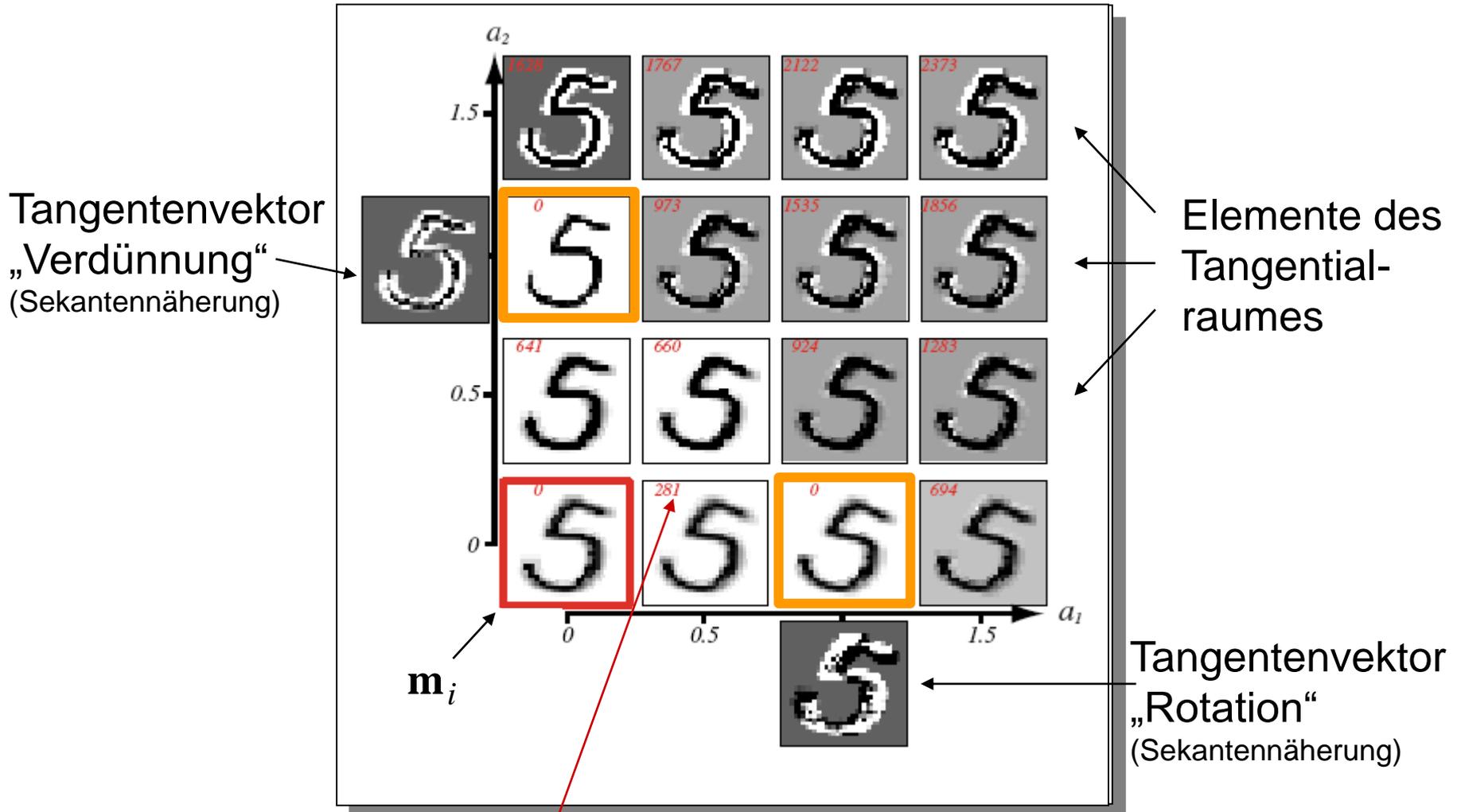
Veranschaulichung:



$$D(\mathbf{m}, \mathbf{m}_i) < D(\mathbf{m}, \mathbf{m}_k) \quad \|\mathbf{m} - \mathbf{m}_i\| > \|\mathbf{m} - \mathbf{m}_k\|$$

2.4. Abstandsmessung im Merkmalsraum – Tangentendistanz

Beispiel: Grauwertbild mit 32×32 Pixeln =: Merkmalsvektor \mathbf{m}



$$\left\| (\mathbf{m}_i + \hat{\mathbf{T}}_{\mathbf{m}_i} \mathbf{a}_{\min}) - A(o_i; \mathbf{p} + \Delta \mathbf{p}) \right\|$$

Quelle: R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification

2.5. Normalisierung

Ziele:

- Elimination irrelevanter Störeinflüsse in Mustern
- Elimination irrelevanter Variationen von Mustern
- Elimination irrelevanter Variationen von Merkmalen

Beispiele:

- Lageausgleich in Bildern
- Beleuchtungsausgleich in Bildern, z.B. Histogrammtransformation
- Amplitudenausgleich bei Audiosignalen z.B. AGC
- Entzerrung der Zeitachse bei Zeitsignalen (z.B. Sprachsignalen)
- Beseitigung geometrischer, durch die optische Abbildung bedingter Verzerrungen in Bildern

Eliminiert man Variationen der Muster durch Normalisierung, so sind aus den normalisierten Muster abgeleitete Merkmale unabhängig von diesen Mustervariationen.

2.5. Normalisierung

Beispiele:

- Zentrierung
- Abstraktion von der physikalischen Einheit
- Egalisierung der Größenverhältnisse

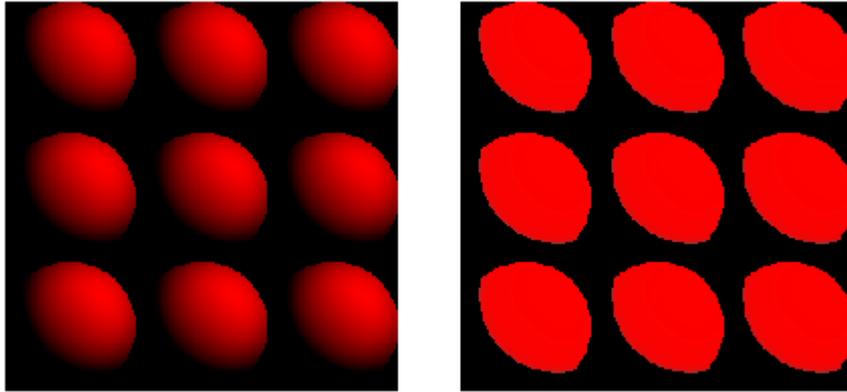
$$\mathbf{m}' := \mathbf{m} - \bar{\mathbf{m}} \quad \text{mit} \quad \bar{\mathbf{m}} = \frac{1}{N} \sum_{i=1}^N \mathbf{m}_i \approx E\{\mathbf{m}\}$$

$$\mathbf{m}'' := \left(\frac{m'_1}{s_1}, \quad \dots, \quad \frac{m'_d}{s_d} \right)^T$$

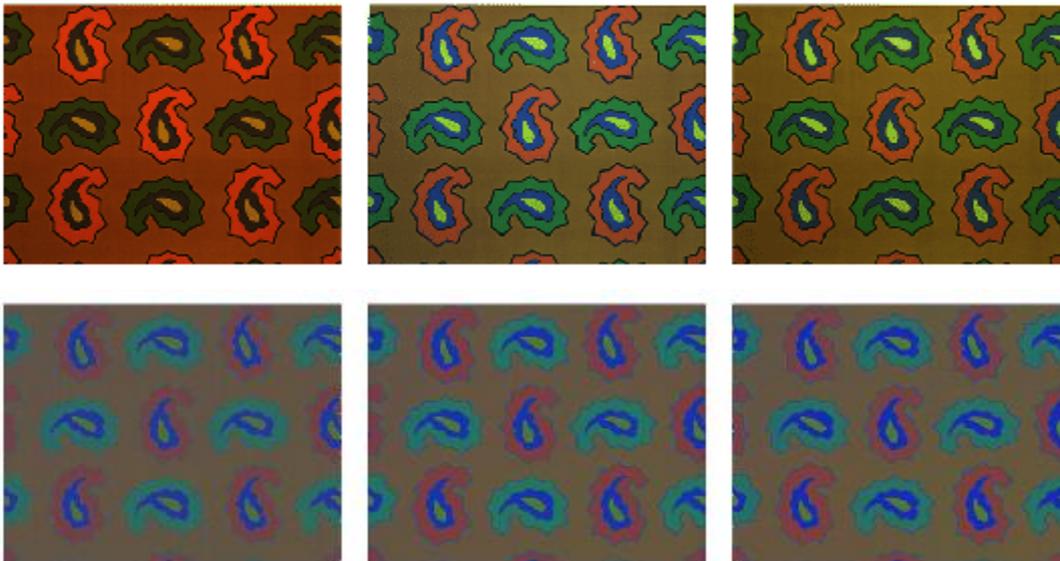
$$\text{mit} \quad s_j := \sqrt{\frac{1}{N-1} \sum_{i=1}^N (m'_{i,j})^2} \approx \sqrt{\text{Var}\{m_j\}} \Rightarrow \text{Var}\{m''_j\} \approx 1$$

2.5. Normalisierung

Beispiel zum Beleuchtungsausgleich in Bildern



Chromatizitätsnormalisierung
eines Bildes mit Schatten.

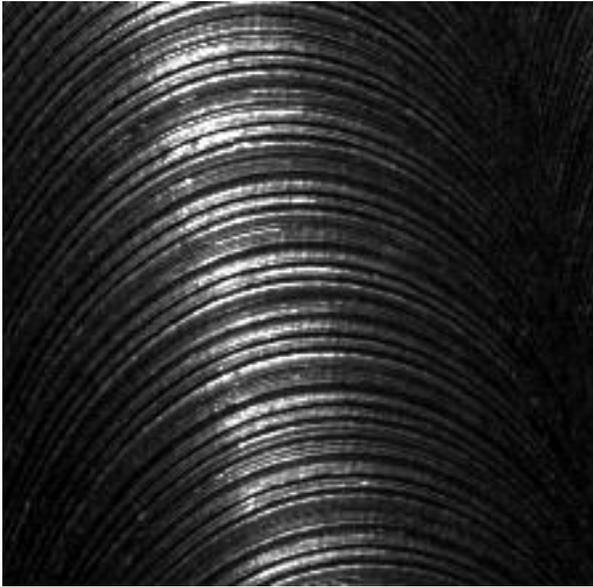


Normalisierung von Bildern,
die mit unterschiedlichen
Beleuchtungen aufgenommen
wurden.

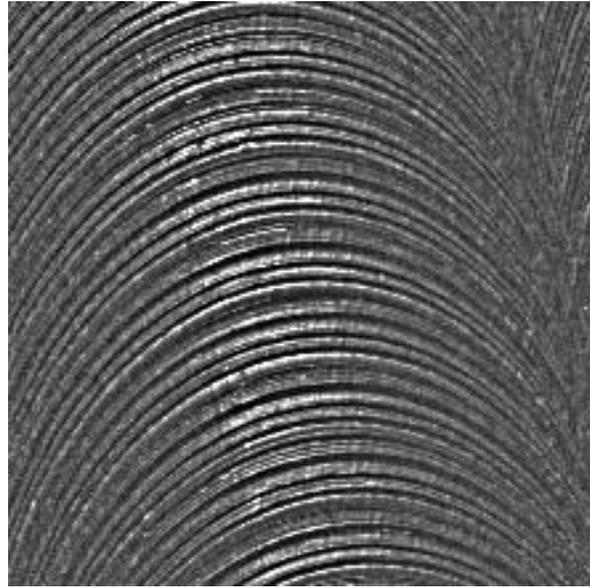
<http://www.color-research.com/>

2.5. Normalisierung

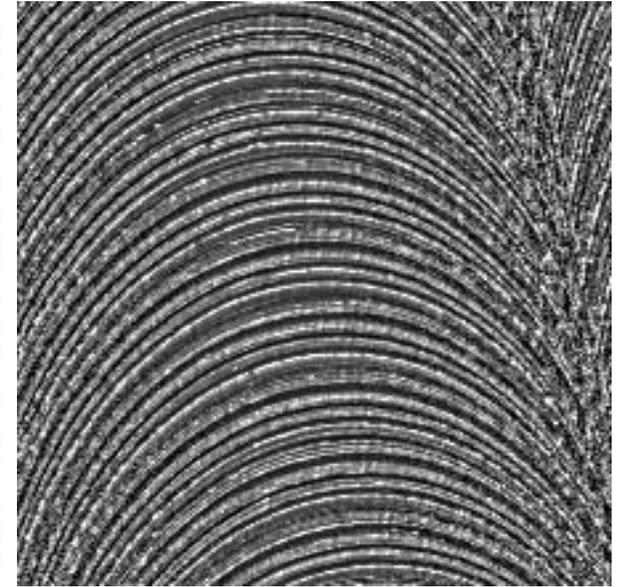
Beispiel zum Beleuchtungsausgleich in Bildern



Originalbild Frästextur



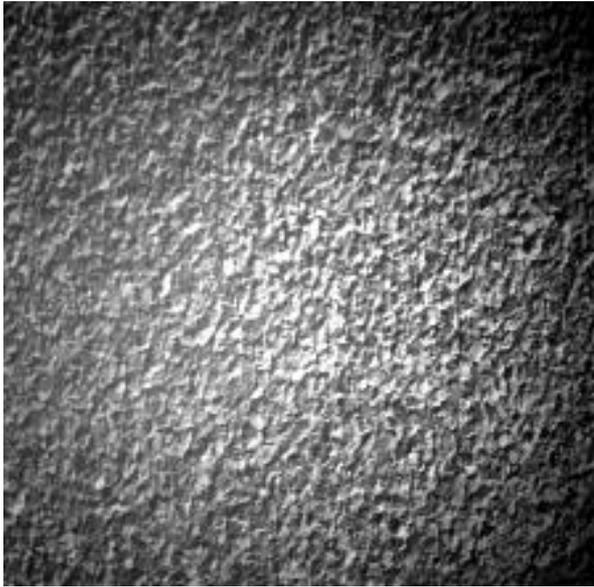
Homomorphe
Filterung



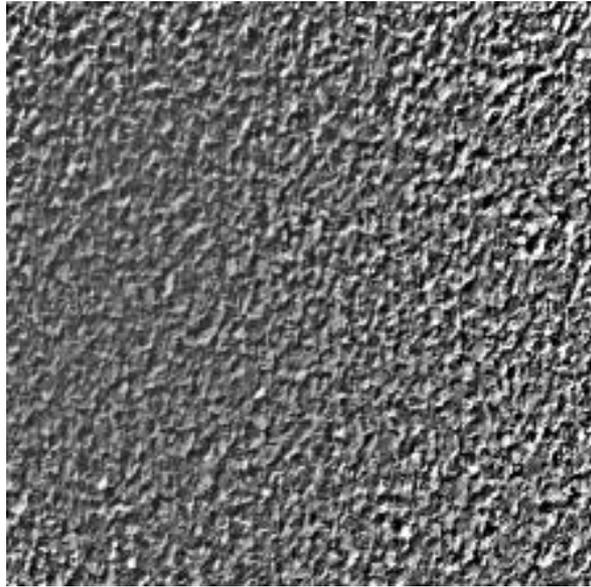
Homogenisierung von
Mittelwert u. Kontrast

2.5. Normalisierung

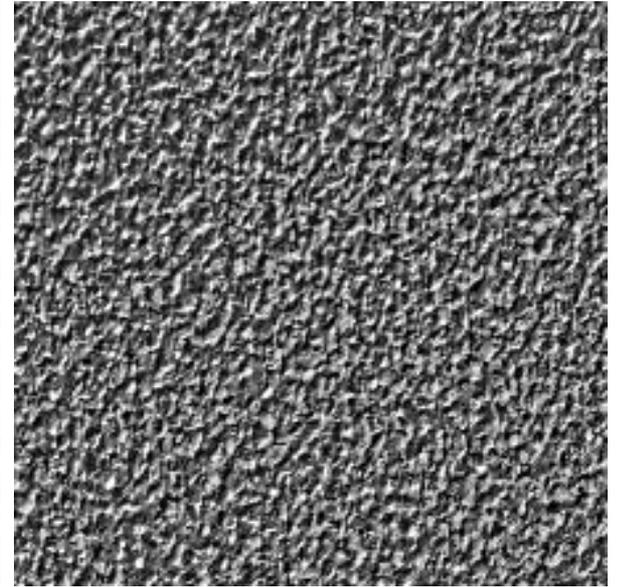
Beispiel zum Beleuchtungsausgleich in Bildern



Presskorktextur



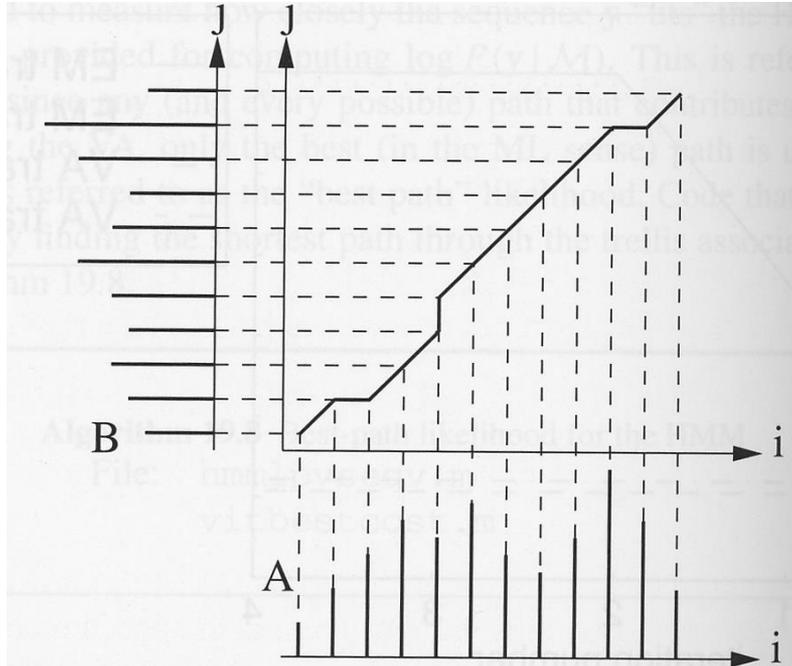
Homomorphe
Filterung



Homogenisierung von
Mittelwert u. Kontrast

2.5. Normalisierung

Bsp.: Zeitverzerrungsausgleich: Dynamic warping, z.B. bei Sprachsignalen



Quelle: T. K. Moon, W. C. Stirling: Mathematical Methods and Algorithms

$$x(t) : \mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_M\} \quad y(t) = x(v(t))$$

$$y(t) : \mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$$

Die Zeitpunkte von \mathbf{a}_i und \mathbf{b}_i stimmen nicht überein, Ordnung ist jedoch erhalten.

$$v(t) : \mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_K\} \quad \mathbf{c}_k = (i_k, j_k)$$

Indexpaare minimaler Abweichung:

$$\mathbf{C}_{opt} = \arg \min_{\mathbf{c}} \left\{ \sum_{k=1}^K \|\mathbf{a}_{i_k} - \mathbf{b}_{j_k}\| \right\}$$

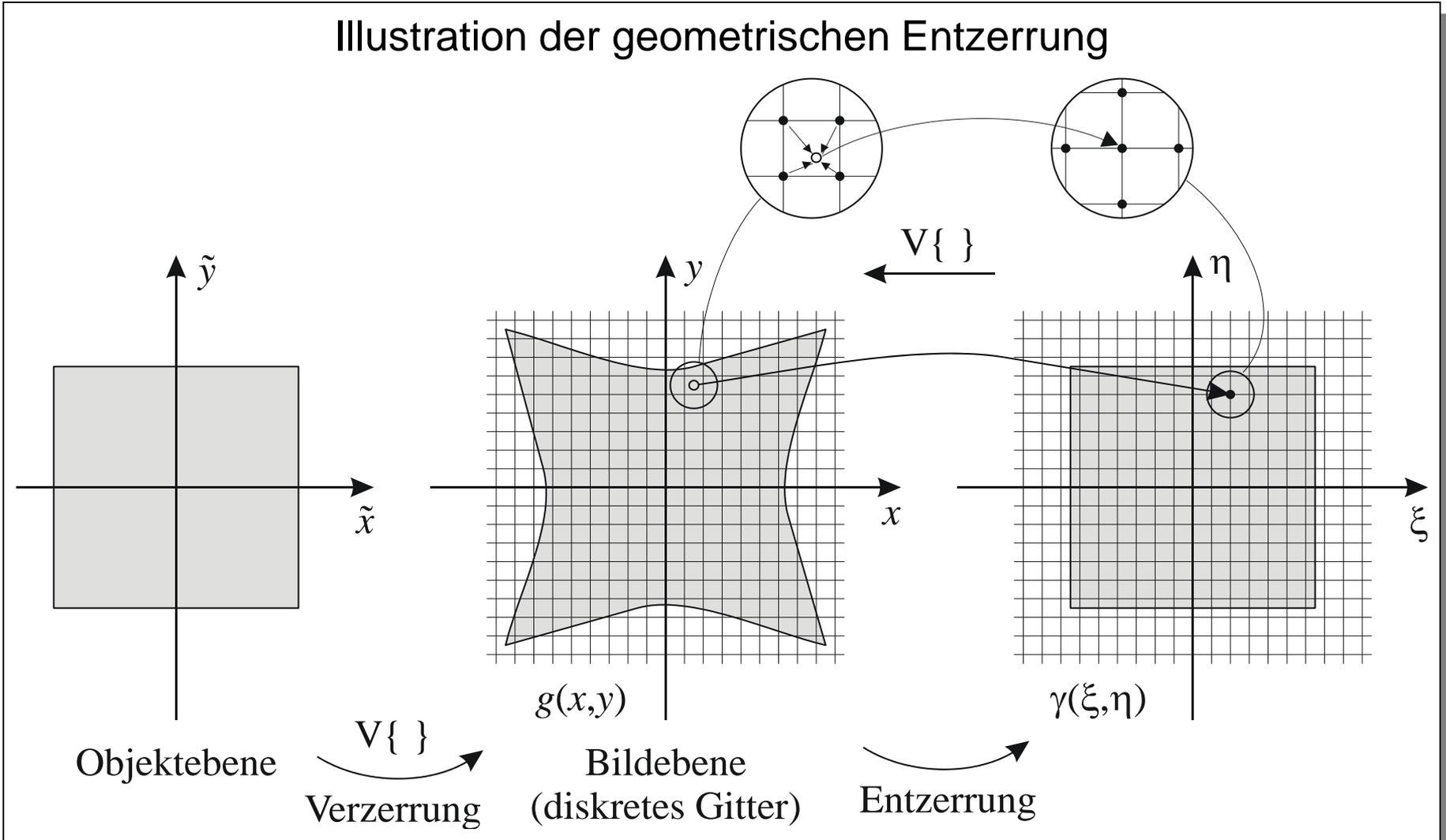
Eigenschaften von $v(t)$:

- streng monoton wachsend
- die Endpunkte müssen übereinstimmen: $i_1 = j_1 = 1, i_k = M, j_k = N$
- alle Punkten müssen aufgenommen werden: $i_k - i_{k-1} \leq 1, j_k - j_{k-1} \leq 1$

2.5. Normalisierung

Beispiel zur geometrischen Entzerrung von Bildern

Illustration der geometrischen Entzerrung



2.6. Auswahl und Konstruktion von Merkmalen

2.6.1. Deskriptive Merkmalen

Definition von Merkmalen, die offensichtliche Eigenschaften von Objekten bzw. ihren Klassen beschreiben.

Typisch: Heuristisch motivierte Merkmale mit anschaulicher Bedeutung.

2.6.1 Deskriptive Merkmale

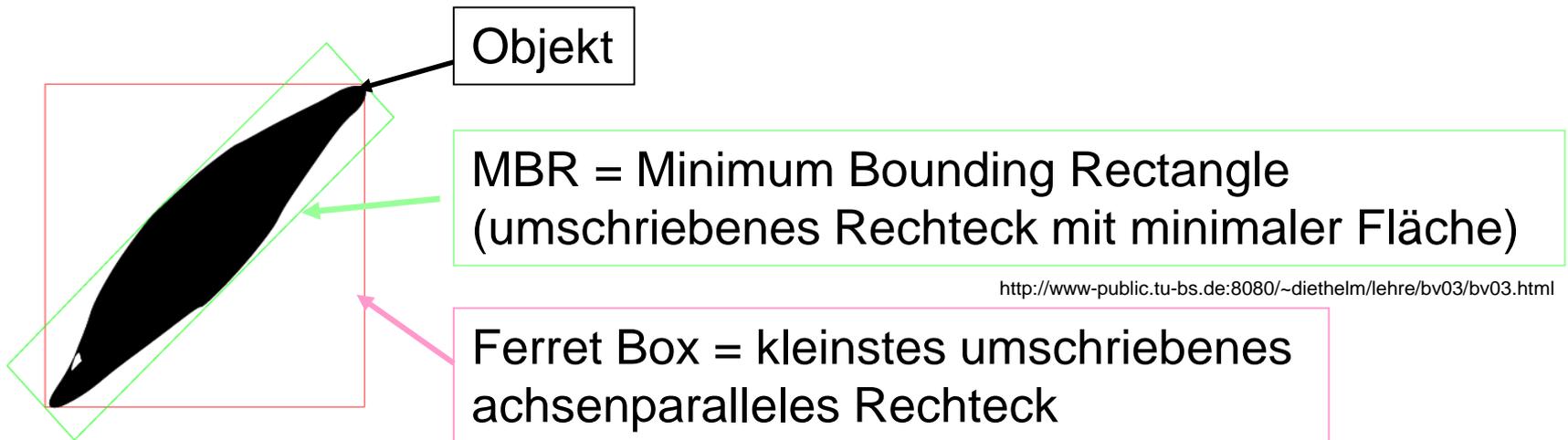
Beispiel: Geometrische Merkmale

Füllungsgrad:

$$m = \frac{\text{Fläche des kleinsten umschreibenden Rechtecks}}{\text{Fläche des Objekts}}$$

= 1 falls Objekt rechteckig

= $4/\pi$ falls Objekt kreisförmig

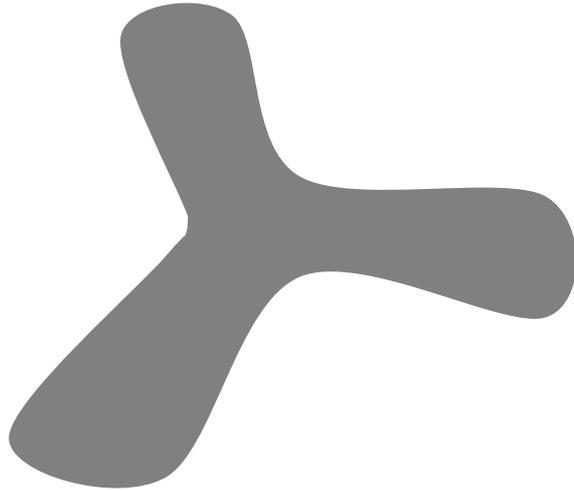


Diskussion:

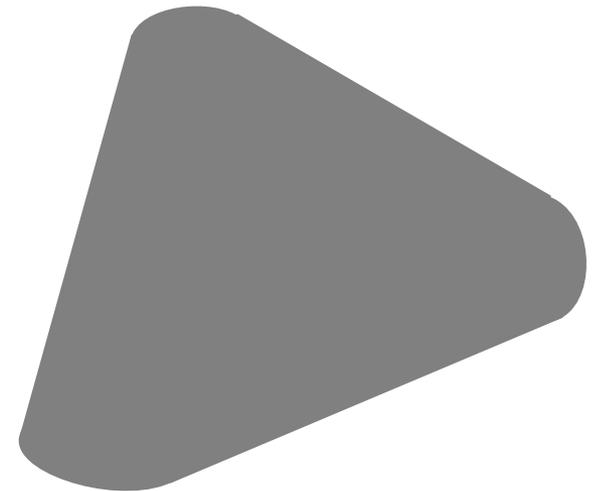
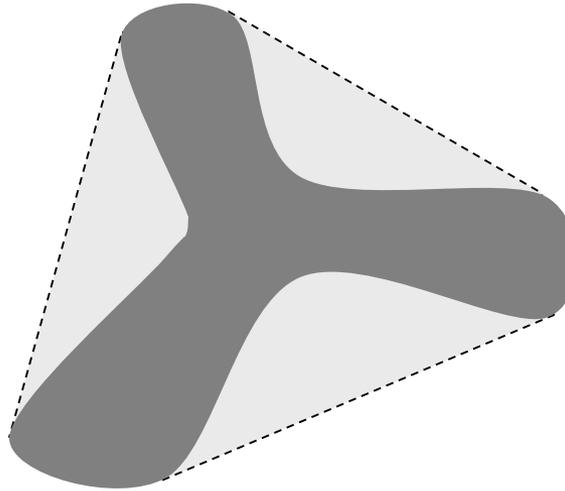
- lage-, rotations- und skalierungsinvariant?

2.6.1 Deskriptive Merkmale

Beispiel: Geometrische Merkmale



Objekt



Konvexe Hülle

Konvexitätsgrad:

$$m = \frac{\text{Fläche}(\textit{Objekt})}{\text{Fläche}(\text{Konvexe Hülle}\{\textit{Objekt}\})}$$

Diskussion:

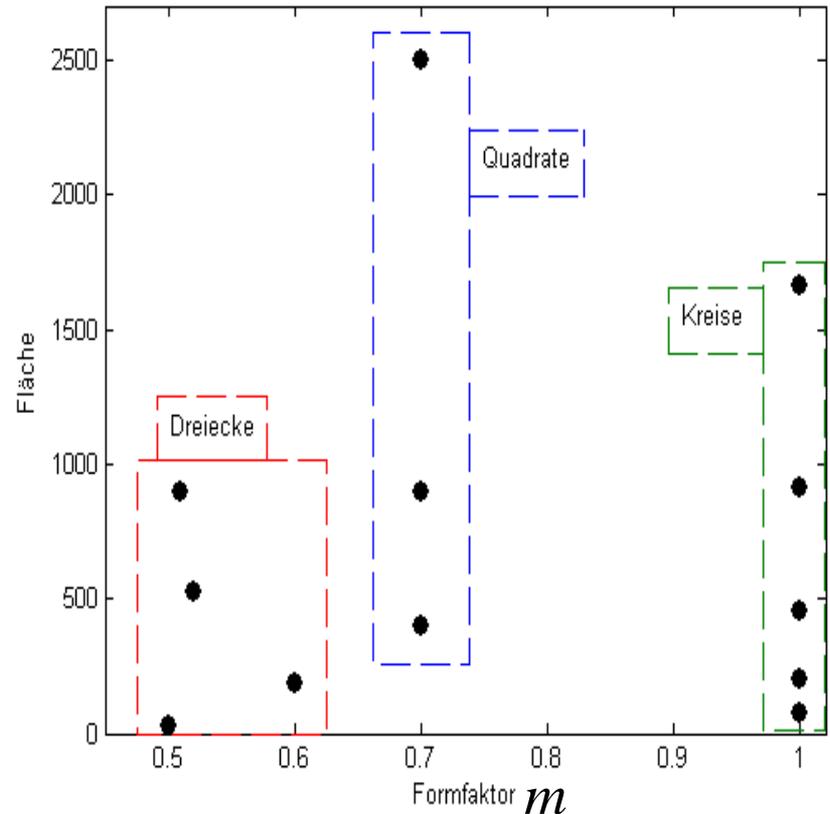
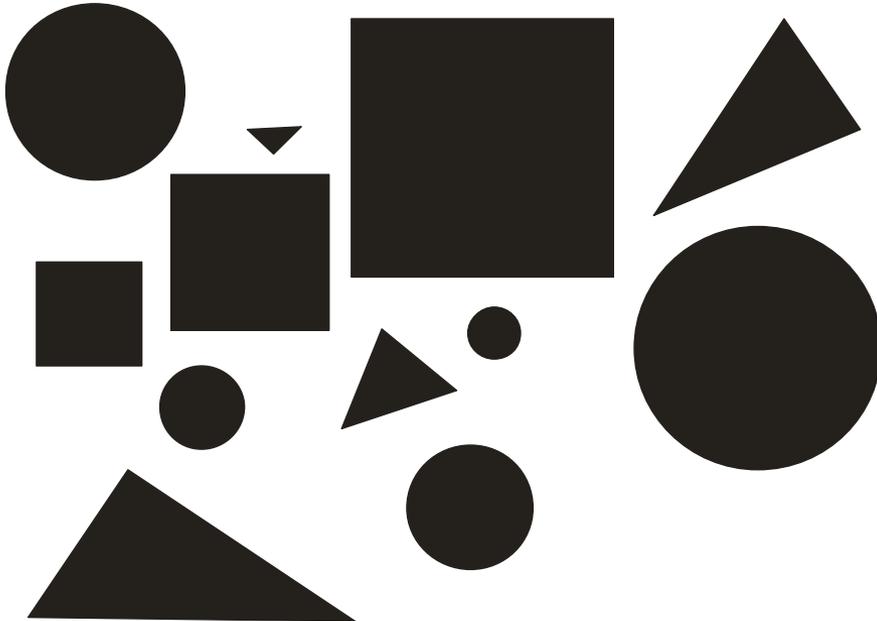
- lage-, rotations- und skalierungsinvariant

2.6.1 Deskriptive Merkmale

Beispiel: Geometrische Merkmale

Kompaktheit des Objektes (Formfaktor):

$$m = \frac{4 \pi \cdot \text{Fläche}}{\text{Umfang}^2}$$



Diskussion: m ist lage-, rotations- und skalierungsinvariant.

2.6.1 Deskriptive Merkmale

Beispiel: Topologische Merkmale

Eulersche Zahl: $m = E = B - L$

B = Anzahl der verbundenen Objektteile

L = Anzahl der Löcher

A B C D E F

A D

$$m = 1 - 1 = 0$$

B

$$m = 1 - 2 = -1$$

C E F

$$m = 1 - 0 = 1$$

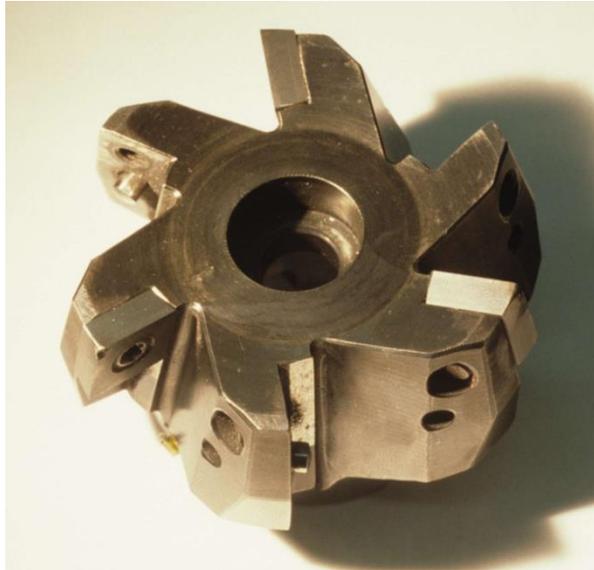
Diskussion: m ist invariant bezüglich „Gummituchverzerrungen“.

2.6.1 Deskriptive Merkmale

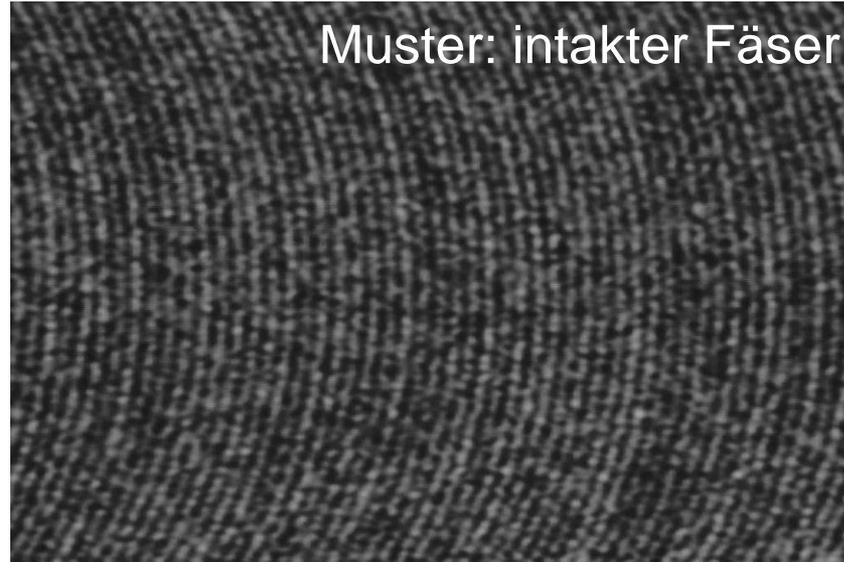
Bsp.: Detektion von Werkzeugbrüchen beim Stirnplanfräsen:

ω_1 : intakter Fräser

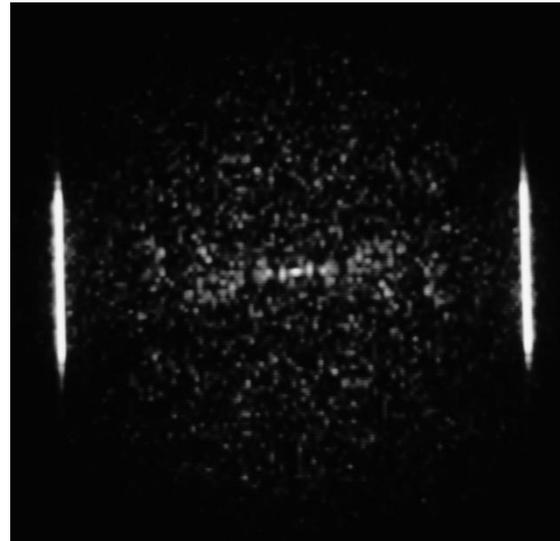
ω_2 : defekter Fräser



Fräser mit 3 Zähnen



Muster: intakter Fäser



Periodogramm =
Betragsquadrat
der Fouriertrans-
formierten

2.6.1 Deskriptive Merkmale

Bsp.: Detektion von Werkzeugbrüchen beim Stirnplanfräsen:

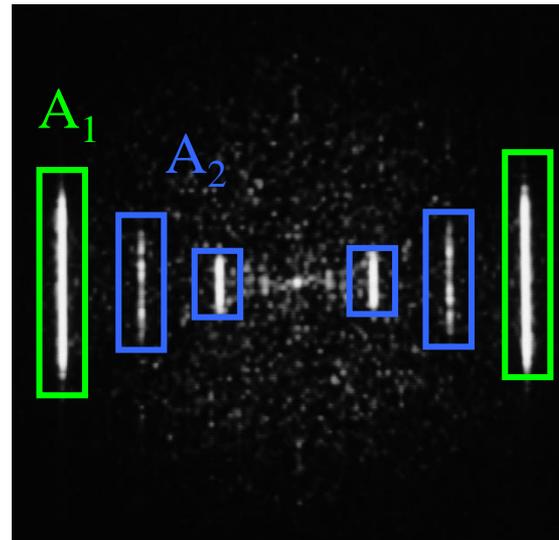
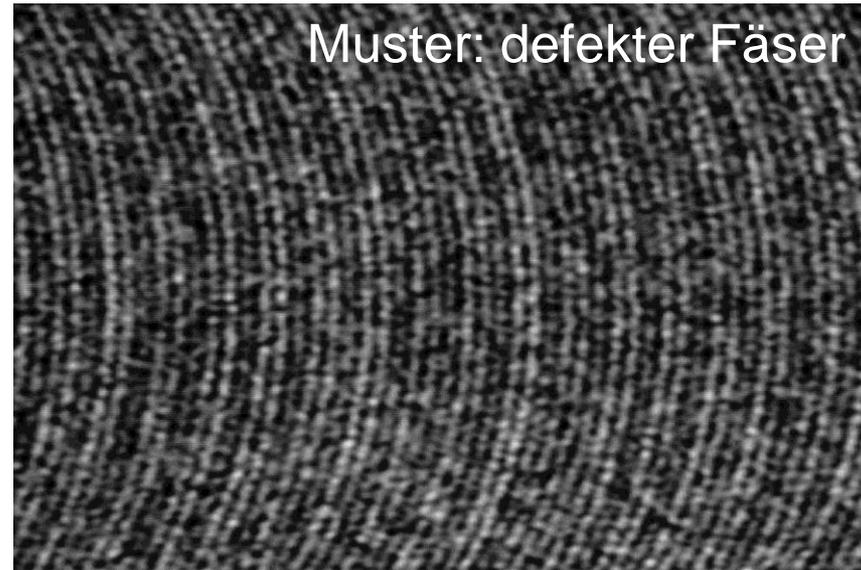
ω_1 : intakter Fräser

ω_2 : defekter Fräser



Zahn mit Schneidenbruch

- 1. „Harmonische“
- „Subharmonische“



Periodogramm

Merkmals:

$$m := \frac{\iint_{A_2} |F\{t(x, y)\}|^2 d f_x d f_y}{\iint_{A_1} |F\{t(x, y)\}|^2 d f_x d f_y}$$

2.6.2. Modellparameter als Merkmale

Idee:

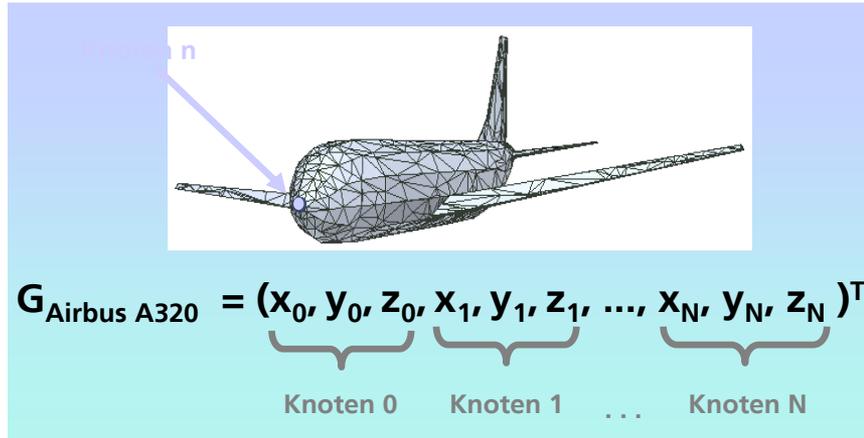
Erzeugendes Modell für die Objekte bzw. für die Muster erstellen, dessen Parameter als Merkmale verwendet werden können.

Vorteile:

- Modellgestützte Merkmale i.d.R. **anschaulich interpretierbar**.
- Modellgestützte Vorgehensweise erlaubt die **Einbringung von Expertenwissen**.
- Modellgestützte Merkmale sind sehr **problemspezifisch**.
- Modellgestütztes Vorgehen führt i.d.R. auf **niedrige Dimension** des Merkmalsraumes.
- **Evaluation** der Merkmale **durch Synthese** von Mustern.

2.6.2. Modellparameter als Merkmale

Beispiel: Erkennung von 3D-Objekten

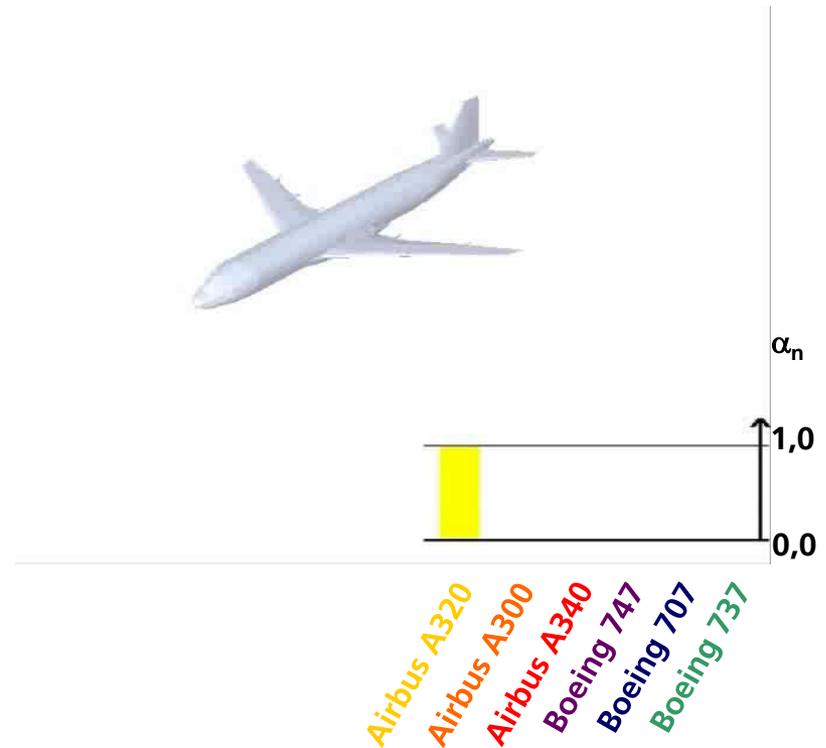


$$\begin{aligned}
 G_{\text{aktuell}} &= \alpha_1 * G_{\text{Airbus A320}} + \alpha_2 * G_{\text{Airbus A300}} \\
 &+ \alpha_3 * G_{\text{Airbus A340}} + \alpha_4 * G_{\text{Boeing 747}} \\
 &+ \alpha_5 * G_{\text{Boeing 707}} + \alpha_6 * G_{\text{Boeing 737}}
 \end{aligned}$$

Parameter

$$\mathbf{m} := (\alpha_1, \dots, \alpha_6)^T$$

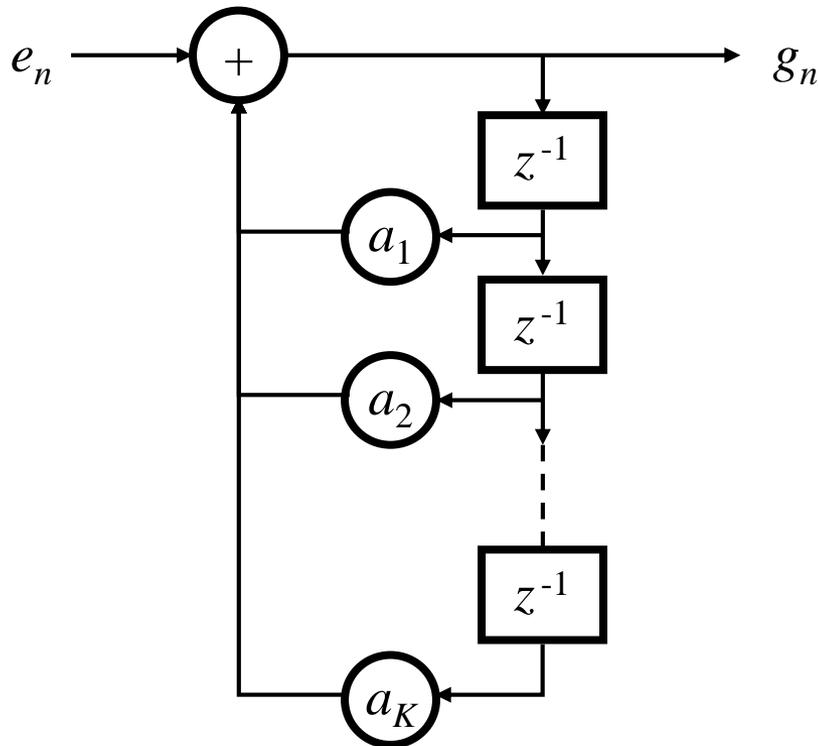
Quelle: Fraunhofer-IOSB Karlsruhe, Prof. Dr. A. Laubenheimer HS KA



2.6.2. Modellparameter als Merkmale

Bsp.: Autoregressives Signalmodell

- leistungsfähiges Modell zu Erzeugung von Zeitreihen
- effiziente Schätzverfahren für die Parameter



$$g_n = \sum_{i=1}^K a_i g_{n-i} + e_n$$

g_n : schwach stationär,

$$E\{g_n\} = 0$$

e_n : schwach stationäres,
weißes Rauschen,

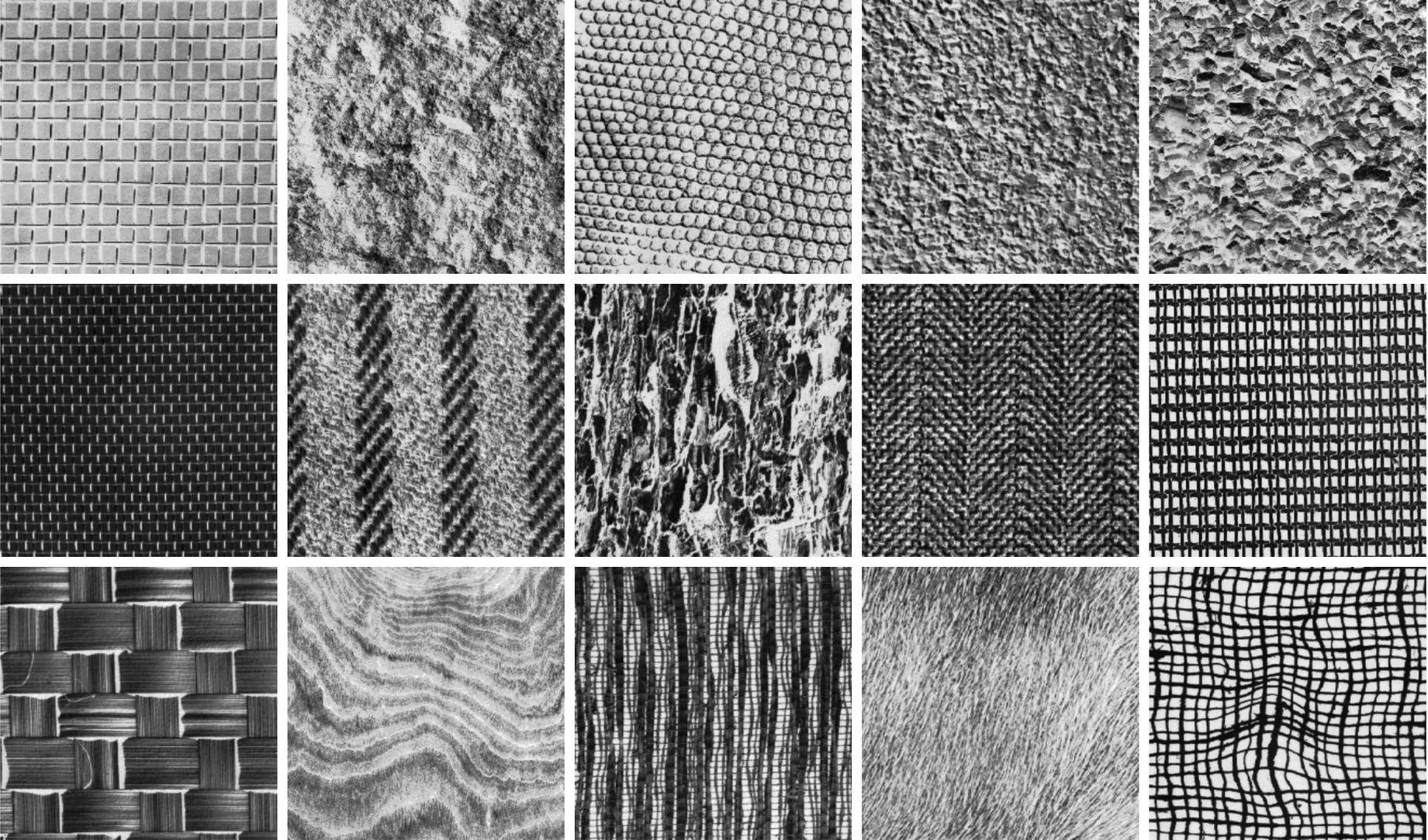
$$E\{e_n\} = 0,$$

$$E\{e_n e_{n+i}\} = \sigma^2 \delta_i^0$$

a_i : AR-Koeffizienten

2.6.2. Modellparameter als Merkmale

Beispiele für Texturen (Quelle: Brodatz)



© 2018 Lehrstuhl für Interaktive Echtzeitsysteme, KIT, alle Rechte einschließlich Kopier- und Weitergaberechte bei uns.

2.6.2. Modellparameter als Merkmale

Beispiel: Autoregressives (AR) Signalmodell zur Texturanalyse

Def. AR-Modell:

$$g_{mn} = \sum_{(k,l) \in U} a_{kl} g_{m-k,n-l} + e_{mn}$$

Zahl der Elemente von $U =: |U|$ (Ordnung des AR-Modells)

g_{mn} : schwach stationär, $E\{g_{mn}\} = 0$

e_{mn} : schwach stationär, weißes Rauschen,

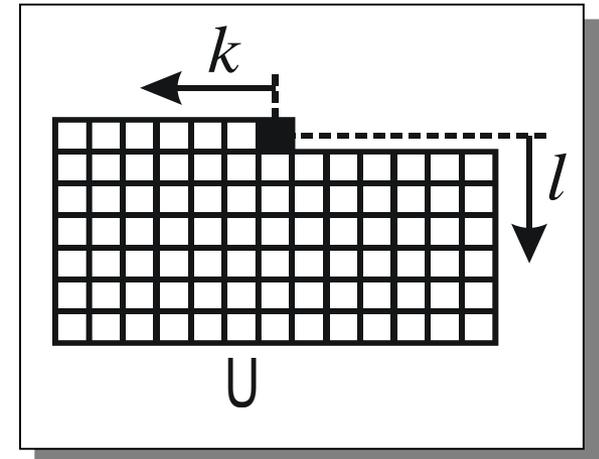
$$E\{e_{mn}\} = 0,$$

$$E\{e_{mn} e_{m+i,n+j}\} = \sigma^2 \delta_i^0 \delta_j^0$$

a_{kl} : AR-Koeffizienten

U : „kausale“ Umgebung des Ursprungs $(0,0) \notin U$

→ rekursive Auswertbarkeit



2.6.2. Modellparameter als Merkmale

Def.:

$$\boldsymbol{\gamma}_{mn} := (\mathbf{K}, g_{m-k, n-l}, \mathbf{K})^T \quad (k, l) \in \mathbf{U}$$

$$\mathbf{a} := (\mathbf{K}, a_{kl}, \mathbf{K})^T \quad (k, l) \in \mathbf{U}$$

$$g_{mn} = \mathbf{a}^T \boldsymbol{\gamma}_{mn} + e_{mn}$$

$$\boldsymbol{\gamma}_{mn}, \mathbf{a} \in \mathbf{R}^{|\mathbf{U}|}$$

$|\mathbf{U}|$: Mächtigkeit von \mathbf{U}

Modellierung einer gegebenen Textur g_{mn} mit den Parametern: \mathbf{a} und σ^2

Merkmalsvektor für die Textur:

$$\mathbf{m} := \begin{pmatrix} \mathbf{a} \\ \sigma \end{pmatrix}$$

Parameterschätzung:

$\mathbf{a}^T \boldsymbol{\gamma}_{mn}$ sagt (prädiziert) aus „Vergangenheitswerten“ $g_{m-k, n-l}$, $(k, l) \in \mathbf{U}$, den aktuellen Wert g_{mn} vorher.

Das AR-Modell ist optimal angepasst, wenn der Vorsagefehler

$g_{mn} - \mathbf{a}^T \boldsymbol{\gamma}_{mn} = e_{mn}$ minimale Varianz hat.

2.6.2. Modellparameter als Merkmale

$$\begin{aligned} E\{e_{mn}^2\} = \sigma^2 &= E\{g_{mn}^2 - 2\mathbf{a}^T \boldsymbol{\gamma}_{mn} g_{mn} + \mathbf{a}^T \boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T \mathbf{a}\} \rightarrow \text{Min.} \\ &= E\{g_{mn}^2\} - 2\mathbf{a}^T E\{\boldsymbol{\gamma}_{mn} g_{mn}\} + \mathbf{a}^T E\{\boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T\} \mathbf{a} \rightarrow \text{Min.} \end{aligned} \quad (*)$$

Notwendig:

$$\text{grad}_{\mathbf{a}} E\{e_{mn}^2\} = -2E\{\boldsymbol{\gamma}_{mn} g_{mn}\} + 2E\{\boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T\} \mathbf{a} \stackrel{!}{=} 0$$

$$\begin{aligned} \Leftrightarrow \mathbf{a} &= \underbrace{E\{\boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T\}^{-1}}_{\substack{\text{über möglichst} \\ \text{großen Bereich} \\ \text{mitteln}}} \cdot E\{\boldsymbol{\gamma}_{mn} g_{mn}\} \\ &\approx \left(\frac{1}{K} \sum_{m,n} \boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T \right)^{-1} \approx \frac{1}{K} \sum_{m,n} \boldsymbol{\gamma}_{mn} g_{mn} \end{aligned}$$

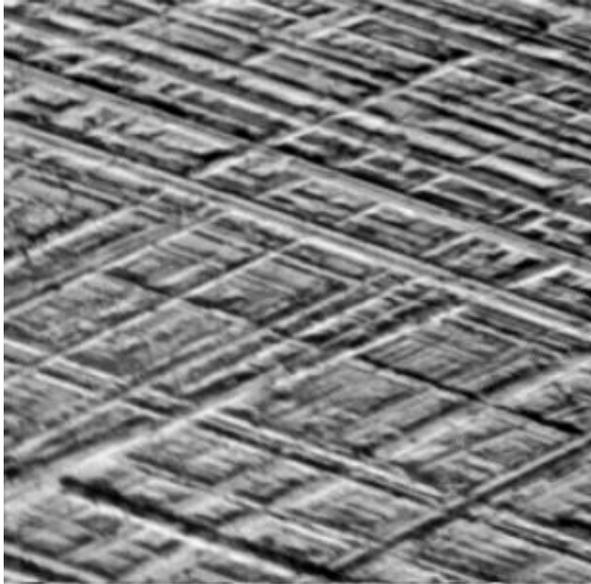
Bei Ergodizität: Schätzung der Kovarianzmatrizen durch örtliche Mittelung

$$\Rightarrow \hat{\mathbf{a}} = \left(\sum_{m,n} \boldsymbol{\gamma}_{mn} \boldsymbol{\gamma}_{mn}^T \right)^{-1} \cdot \sum_{m,n} \boldsymbol{\gamma}_{mn} g_{mn}$$

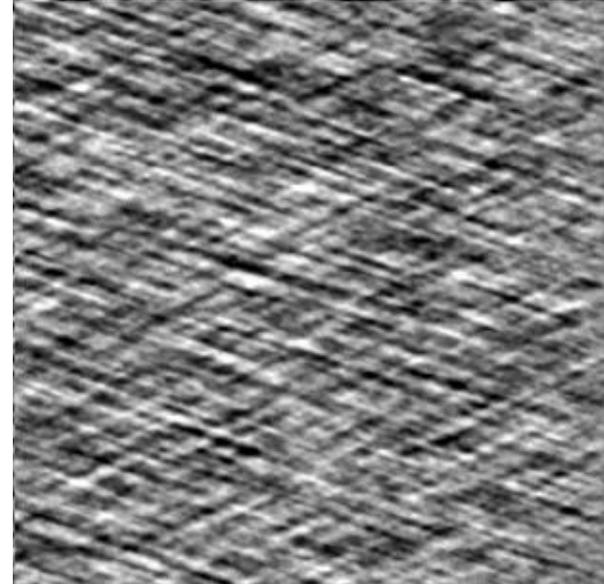
$$\hat{\sigma}^2 \text{ durch Einsetzen von } \hat{\mathbf{a}} \text{ und } E\{g_{mn}^2\} \approx \frac{1}{K} \sum_{m,n} g_{mn}^2 \text{ in } (*)$$

2.6.2. Modellparameter als Merkmale

Beispiel: Synthetische Hontexturen



Original

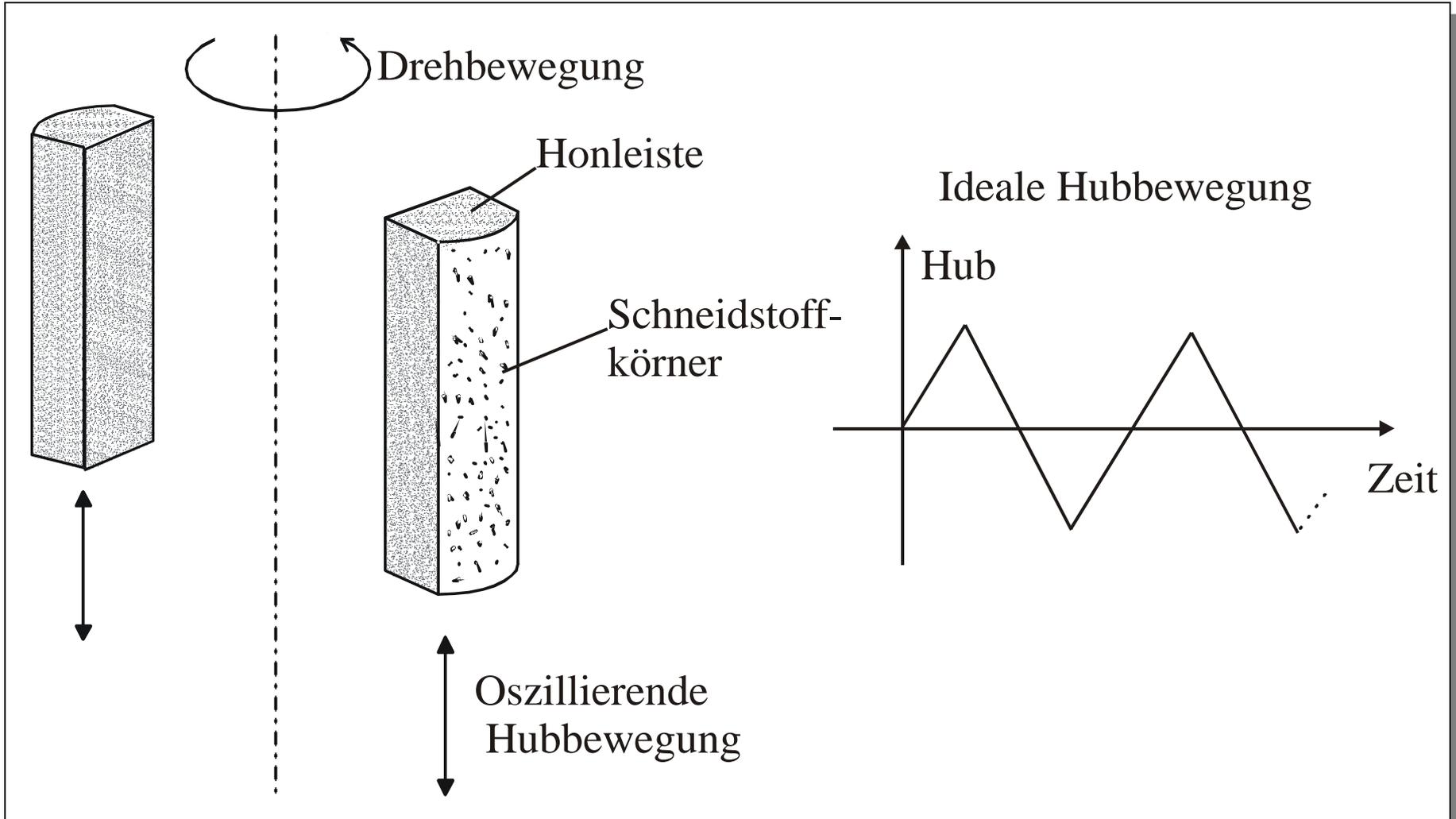


Synthese mit
AR-Modell der
Ordnung 84
(85 Parameter = Merkmale)

2.6.2. Modellparameter als Merkmale

Beispiel: Hontextur

Physikalisch motiviertes Modell



2.6.2. Modellparameter als Merkmale

Beispiel: Hontextur – Ein physikalisch motiviertes Modell

Modell:

$$t(\mathbf{x}) = \sum_{\nu=-\infty}^{\infty} g_{\nu}(\mathbf{x}^T \mathbf{e}_{\nu} - d_{\nu})$$

$g_{\nu}(\cdot)$: zufällige, i. i. d.* Riefenprofile (\rightarrow Texel)

$\mathbf{e}_{\nu} = (\cos \alpha_{\nu}, \sin \alpha_{\nu})^T$: Normalenvektor der ν -ten Riefe

$\alpha_{\nu} \in [0, \pi)$: zufällige, i. i. d. Riefenwinkel

ideal:
$$p(\alpha_{\nu}) = \frac{1}{2}(\delta(\alpha_{\nu} - \beta_1) + \delta(\alpha_{\nu} - \beta_2))$$

d_{ν} : zufällige, gleichverteilte Ursprungsabstände der Riefen;

Zahl q der Riefen in einem Intervall der Länge L :

$$P(q) = e^{-\lambda L} \frac{(\lambda L)^q}{q!} \quad q = 0, K, \infty$$

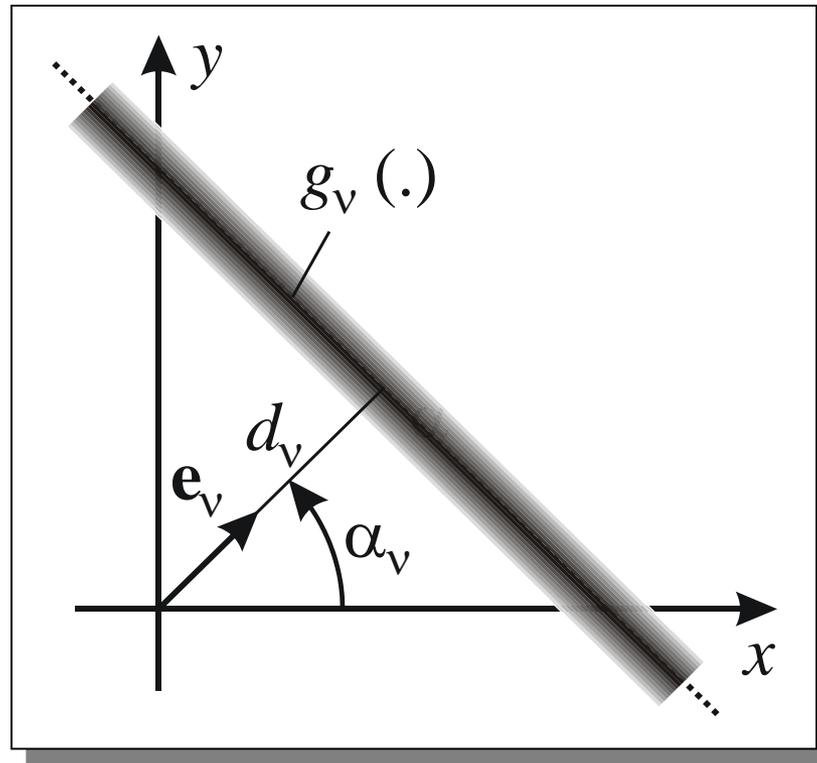
Poisson-
Verteilung

*) i. i. d.: *independent, identically distributed* (statistisch unabhängig, identisch verteilt)

2.6.2. Modellparameter als Merkmale

Beispiel: Hontextur – Ein physikalisch motiviertes Modell

$\alpha_v, d_v \rightarrow$ stochastisches Anordnungsschema

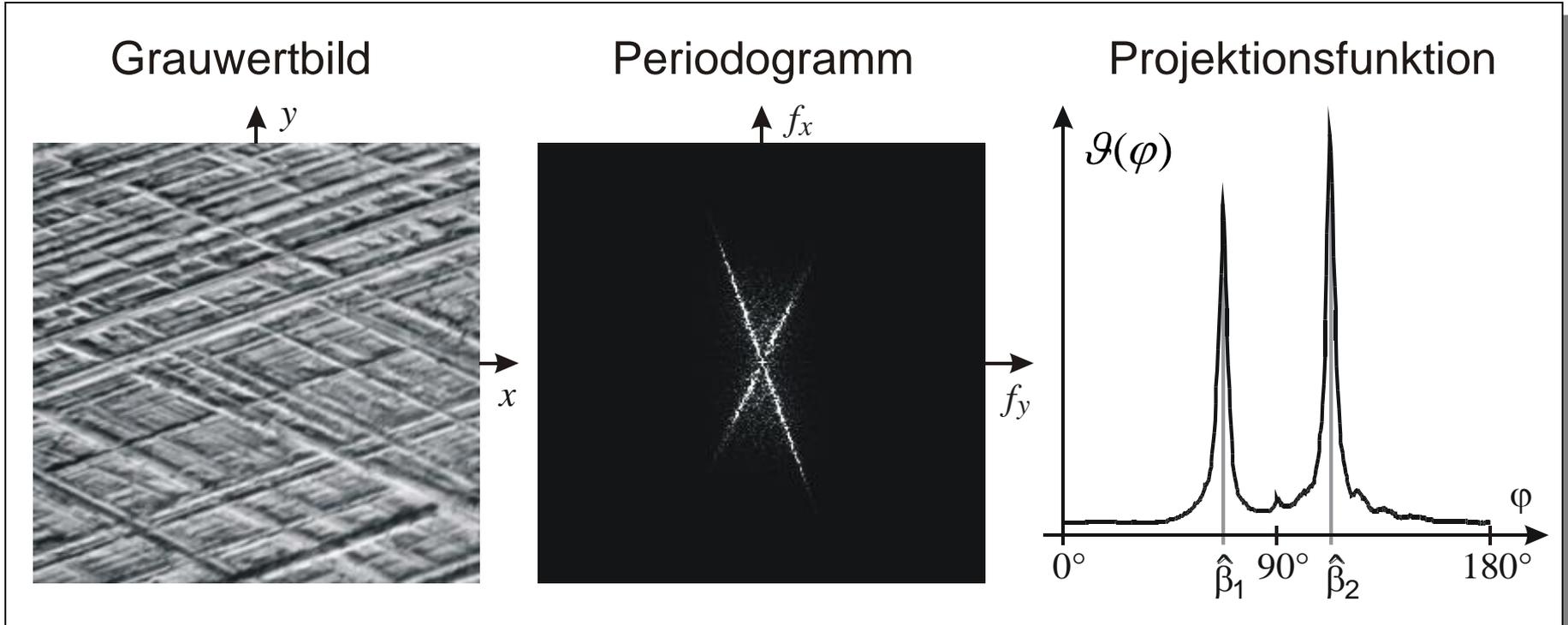


Modellparameter: $\beta_1, \beta_2, \lambda, E\{g_v(\cdot)\} =$ Merkmale

2.6.2. Modellparameter als Merkmale

Beispiel: Hontextur – Ein physikalisch motiviertes Modell

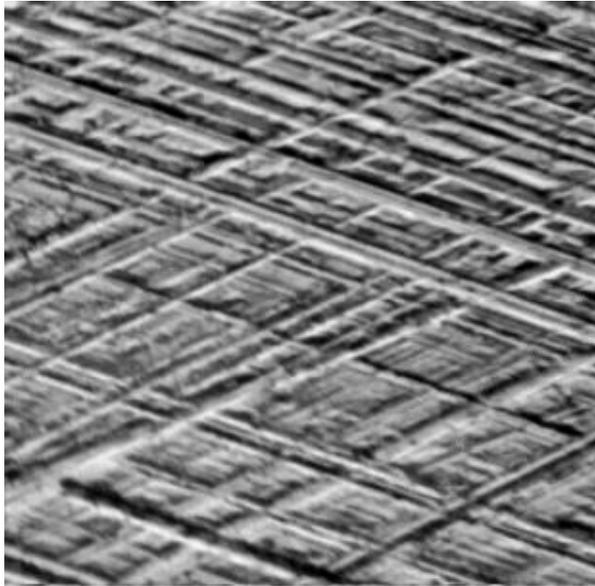
Bsp.: Berechnung (Schätzung) der Merkmale: Vorzugsrichtungswinkel



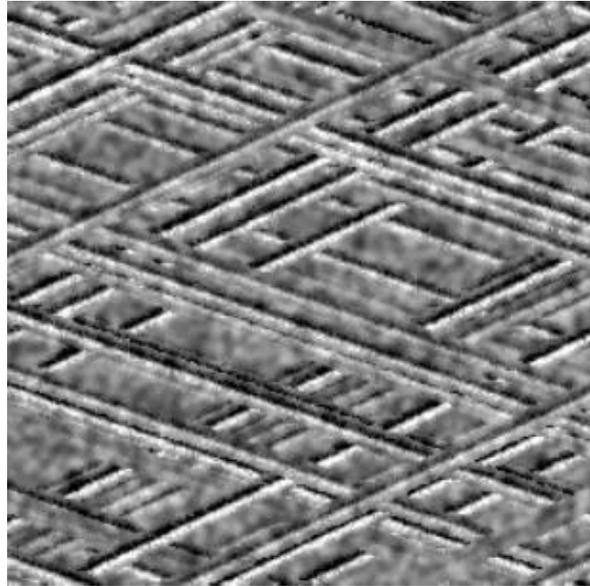
$$\mathcal{G}(\varphi) := \int_{f_y=0}^{\infty} \int_{f_x=-\infty}^{\infty} |T(\mathbf{f})|^2 \delta((-\sin \varphi, \cos \varphi) \mathbf{f}) d\mathbf{f}$$

2.6.2. Modellparameter als Merkmale

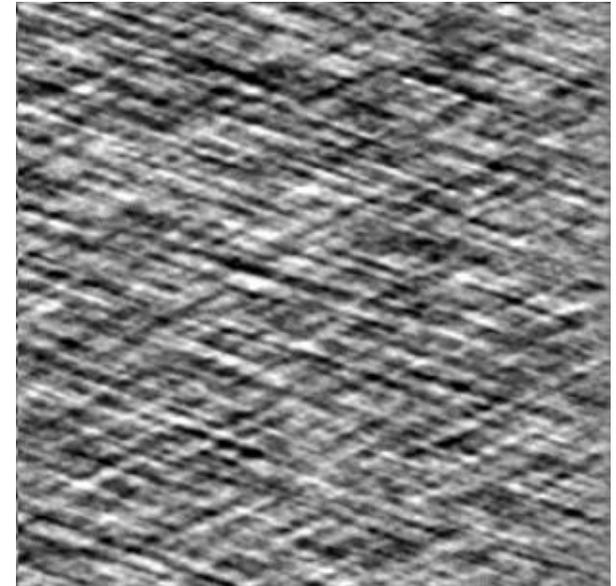
Beispiel: Synthetische Hontexturen



Original



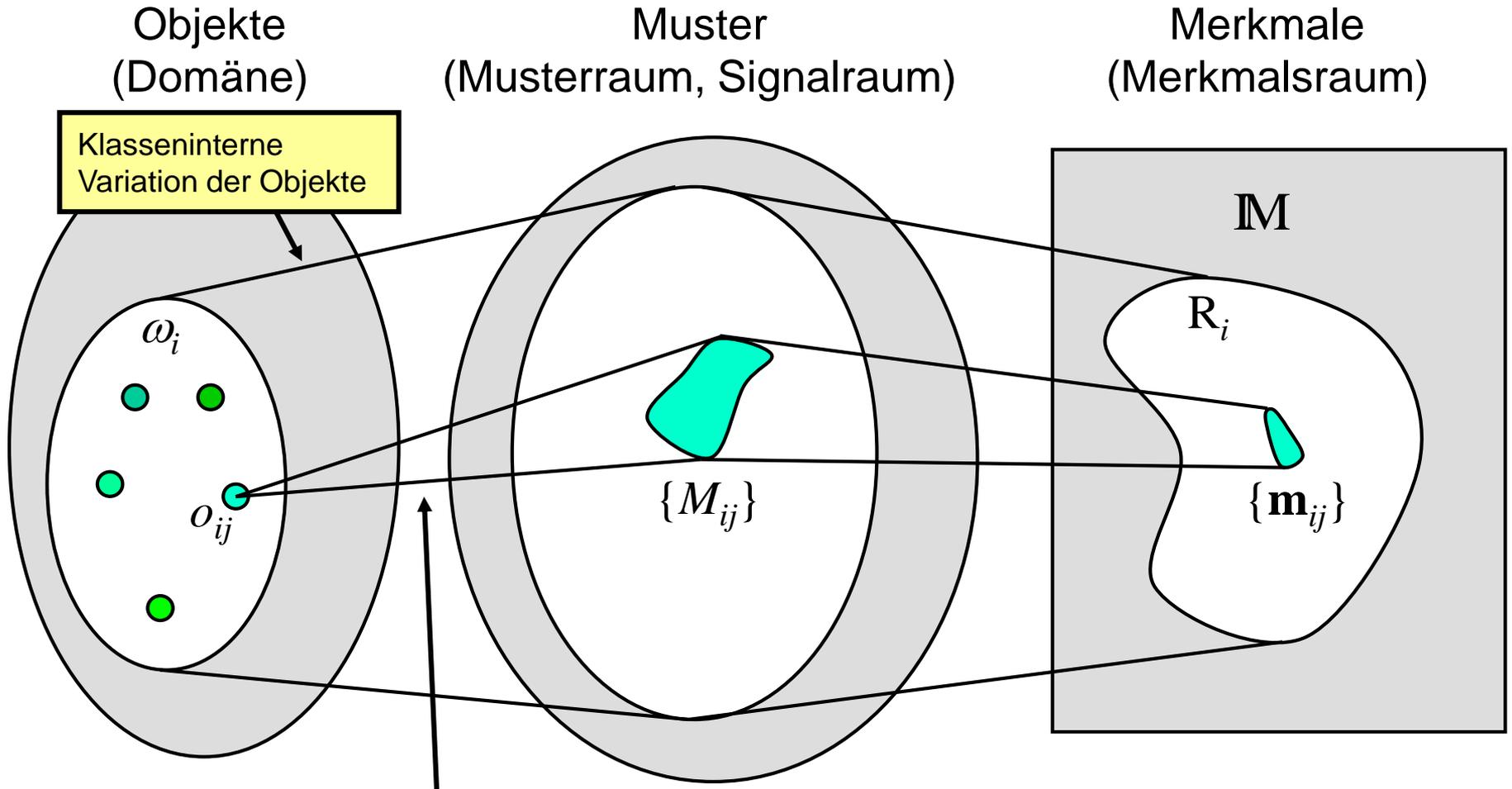
Synthese mit
strukturell-statistischem
Modell
(14 Parameter)



Synthese mit
AR-Modell der
Ordnung 84
(85 Parameter)

2.6.3. Konstruktion invarianter Merkmale

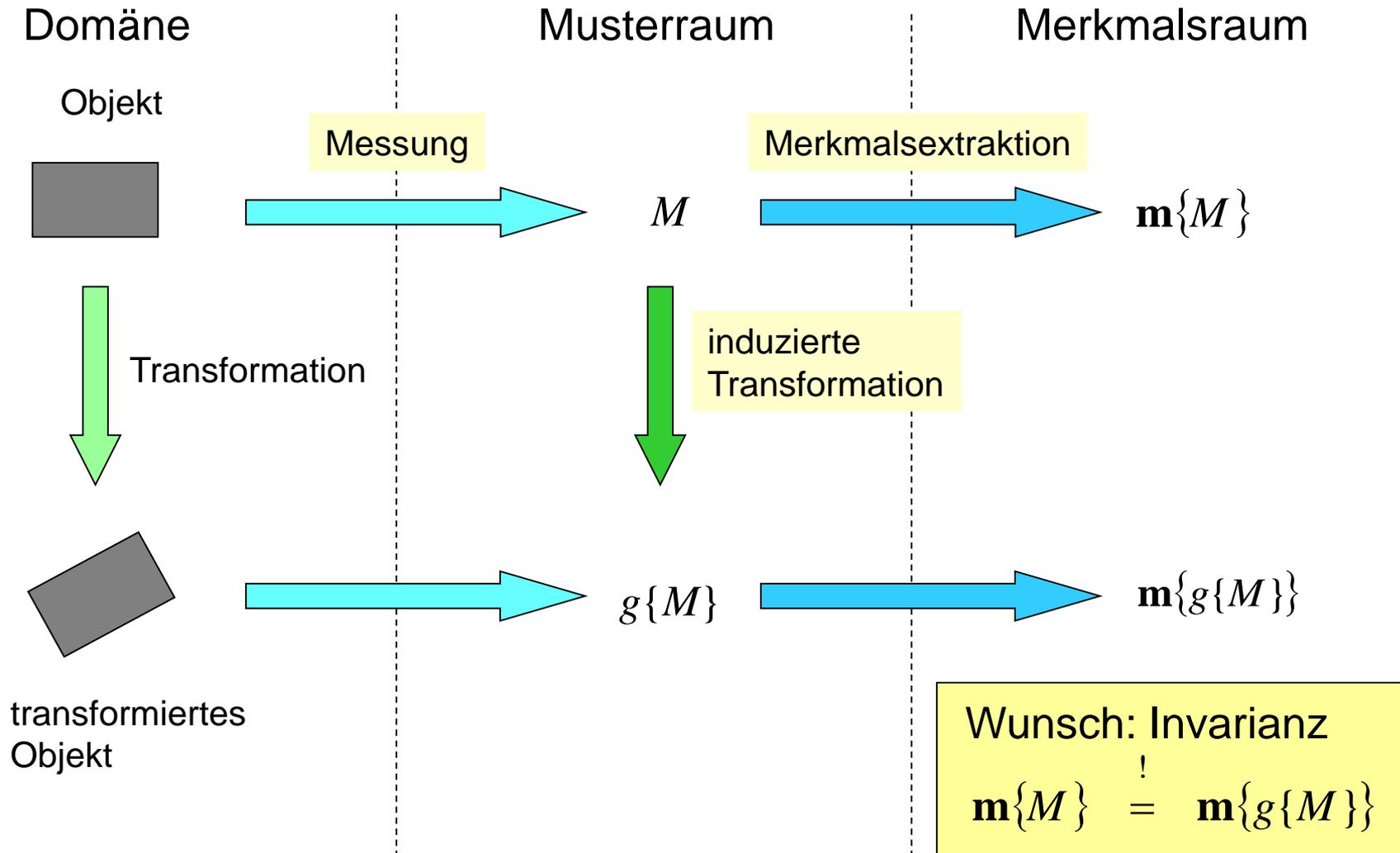
Variation von Mustern und Merkmalen



- Systematische Einflüsse der Messung, Beobachtung: Lage, Skalierung, geometrische Verzerrung, ...
- Störungen: Rauschen, Beleuchtung, ...

2.6.3. Konstruktion invarianter Merkmale

Transformation im Objekt- und Musterraum



2.6.3. Konstruktion invarianter Merkmale

Transformation durch Gruppen

Def.: Eine algebraische Struktur G mit einer dyadischen inneren Verknüpfung \bullet heißt Gruppe, wenn für beliebige Elemente $a, b, c \in G$ gilt:

(1) $a \bullet (b \bullet c) = (a \bullet b) \bullet c$ Assoziativität

(2) Es existiert ein Einselement $e \in G$ mit:

$$a \bullet e = e \bullet a = a \quad \forall a \in G$$

(3) Zu jedem a existiert ein inverses Element $a^{-1} \in G$ mit:

$$a \bullet a^{-1} = a^{-1} \bullet a = e$$

Aus (1)-(3) folgt, dass es genau ein Einselement und zu jedem $a \in G$ genau ein inverses Element $a^{-1} \in G$ gibt.

Ist die Verknüpfung \bullet kommutativ, nennt man G eine abelsche Gruppe.

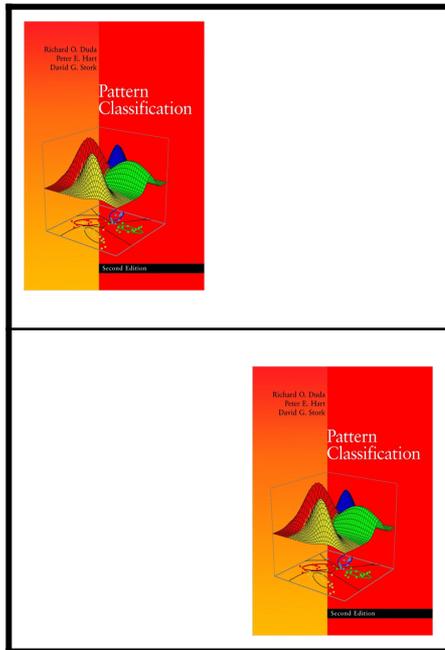
2.6.3. Konstruktion invarianter Merkmale

Beispiele für Transformationsgruppen

Translationsgruppe τ

$$\mathbf{x}' = \mathbf{x} + \mathbf{a} \quad (2 \text{ FG})$$

$$\mathbf{x} = (x, y)^T$$

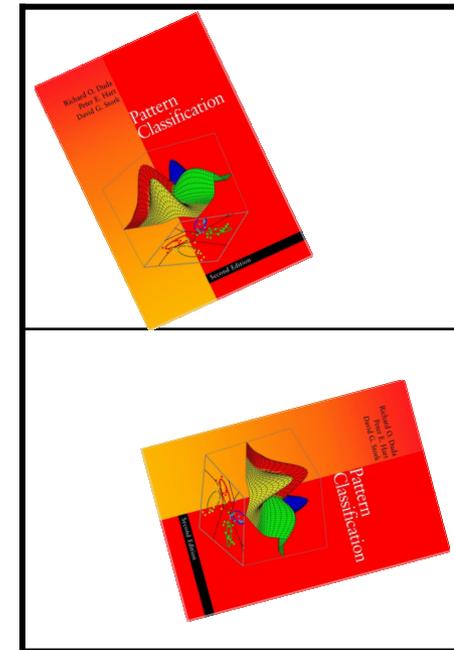


Verschiebung

Kongruenzgruppe C

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{a} \quad (3 \text{ FG})$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{I}$$



+ Drehung

2.6.3. Konstruktion invarianter Merkmale

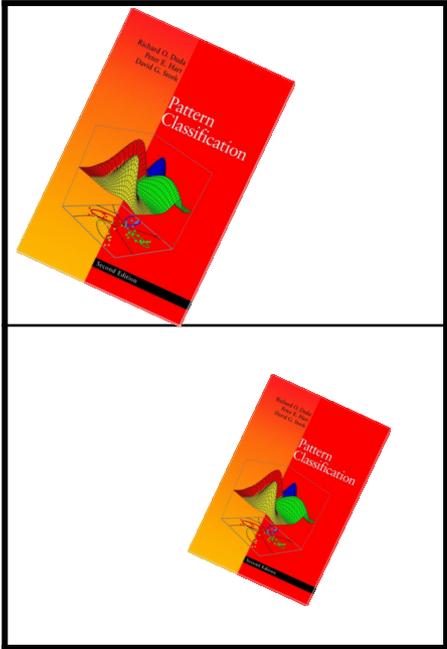
Beispiele für Transformationsgruppen

Ähnlichkeitsgruppe S

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{a}$$

$$\mathbf{A}^T \mathbf{A} = k\mathbf{I}$$

(4 FG)



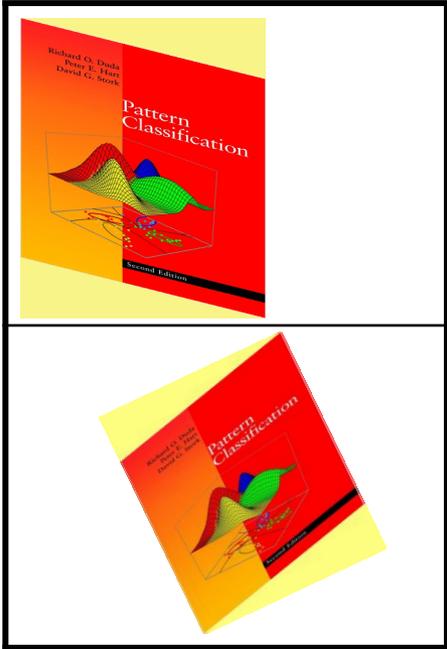
+ Skalierung

Affine Transformation A

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{a}$$

$$\det(\mathbf{A}) \neq 0$$

(6 FG)



2.6.3. Konstruktion invarianter Merkmale

„Brute Force“ Mustererkennung

Gruppe G sei parametrisiert durch den Parametervektor $\mathbf{p} \in P$:

$$G = \{g(\mathbf{p})\{.\}\mid \mathbf{p} \in P\}$$

$$(i, \mathbf{p}^*) = \arg \min_{j, \mathbf{p}} \|M - g(\mathbf{p})\{M_j\}\| \rightarrow \hat{\omega} = \omega_i \quad (*)$$

Entspricht Berechnung einer **verallgemeinerten Kreuzkorrelationsfunktion** mit **$\dim(P)$ Freiheitsgraden**. Aufwand wächst exponentiell mit $\dim(P)$.

Beispiel: Affine Transformation in der Ebene, $\dim(P) = 6$. Jede Klasse sei durch ein M_j repräsentiert. Sei jeder Parameter p_l , $l = 1, \dots, 6$ mit 10^3 Stufen quantisiert:

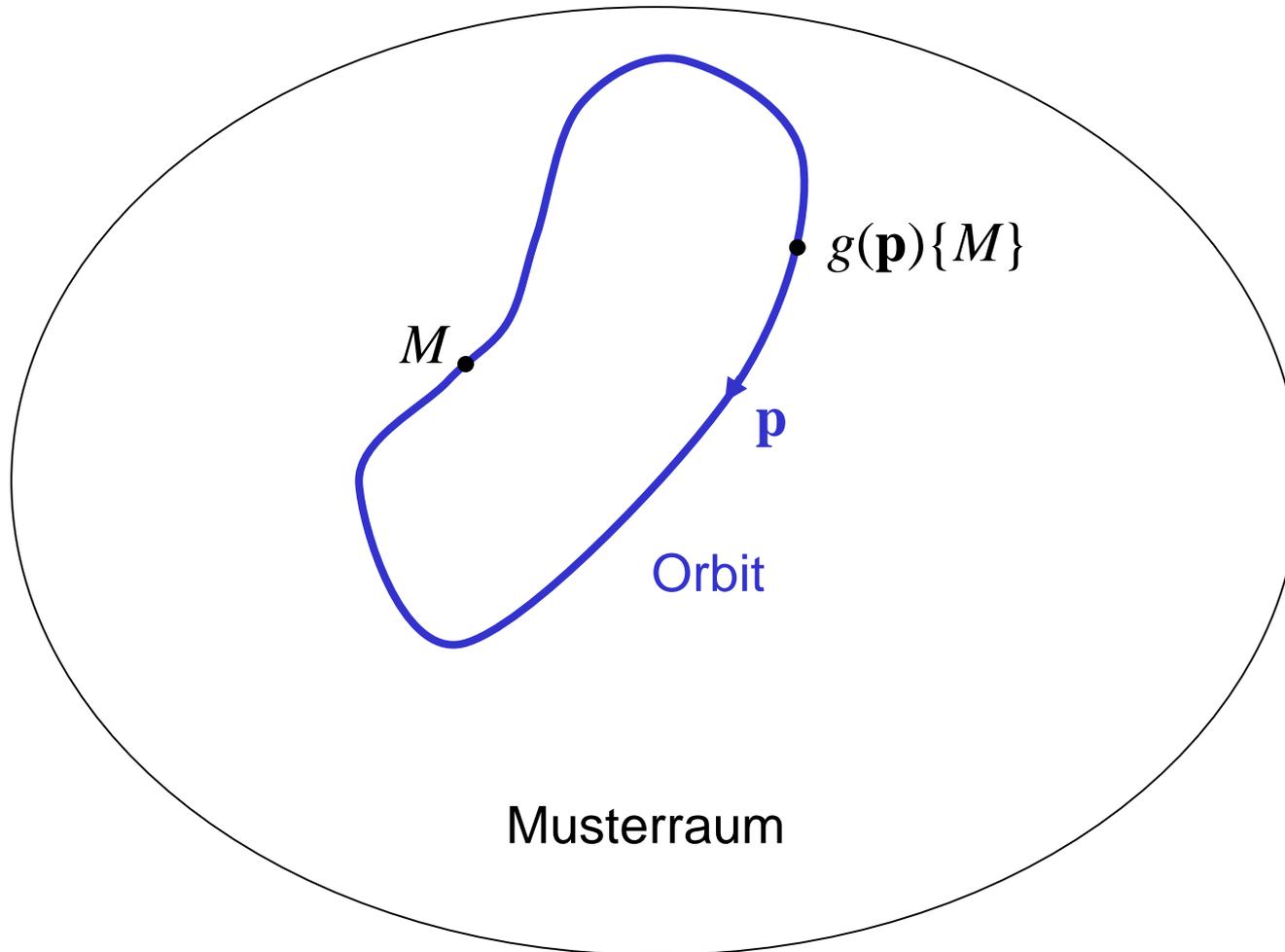
→ Berechnung von Gl. (*) für 10^{18} Werte von \mathbf{p} pro Klasse. Mit einer Maschine, die 10^9 solche Berechnungen/s durchführen kann, bräuchte man ca. 31,7 Jahre.

→ Wunsch nach **invarianten Merkmalen** bezüglich G

2.6.3. Konstruktion invarianter Merkmale

Definition: **Orbit** eines Musters bezüglich einer Gruppe G :

$$\{g\{M\} \mid g \in G\} = \{g(\mathbf{p})\{M\} \mid \mathbf{p} \in P\}$$



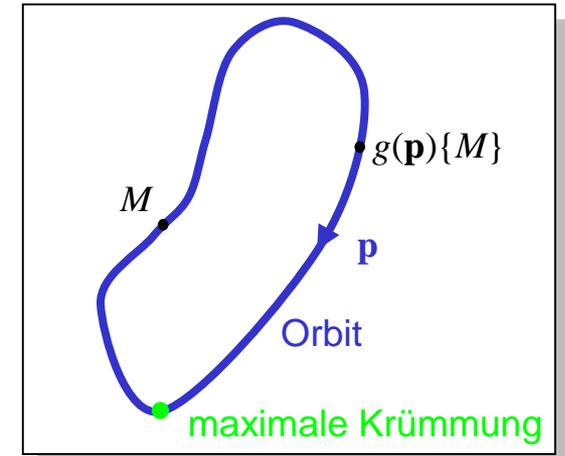
2.6.3. Konstruktion invarianter Merkmale

Drei fundamentale Ansätze zur invariante Merkmalsgewinnung

(1) **Integration über die Transformationsgruppe**, Haar-Integral, Hurwitz 1897

$$m = \int_G f(g\{M\}) d g = \int_{\mathbf{p} \in P} f(g(\mathbf{p})\{M\}) d \mathbf{p}$$

Funktion $f(\cdot)$: z.B. Polynome, Monome



(2) **Differentielle Methode**

$$\frac{\partial m(g(\mathbf{p})\{M\})}{\partial p_l} \equiv 0 \quad \text{für } l = 1, \dots, \dim(P) \text{ und } \forall \mathbf{p} \in P$$

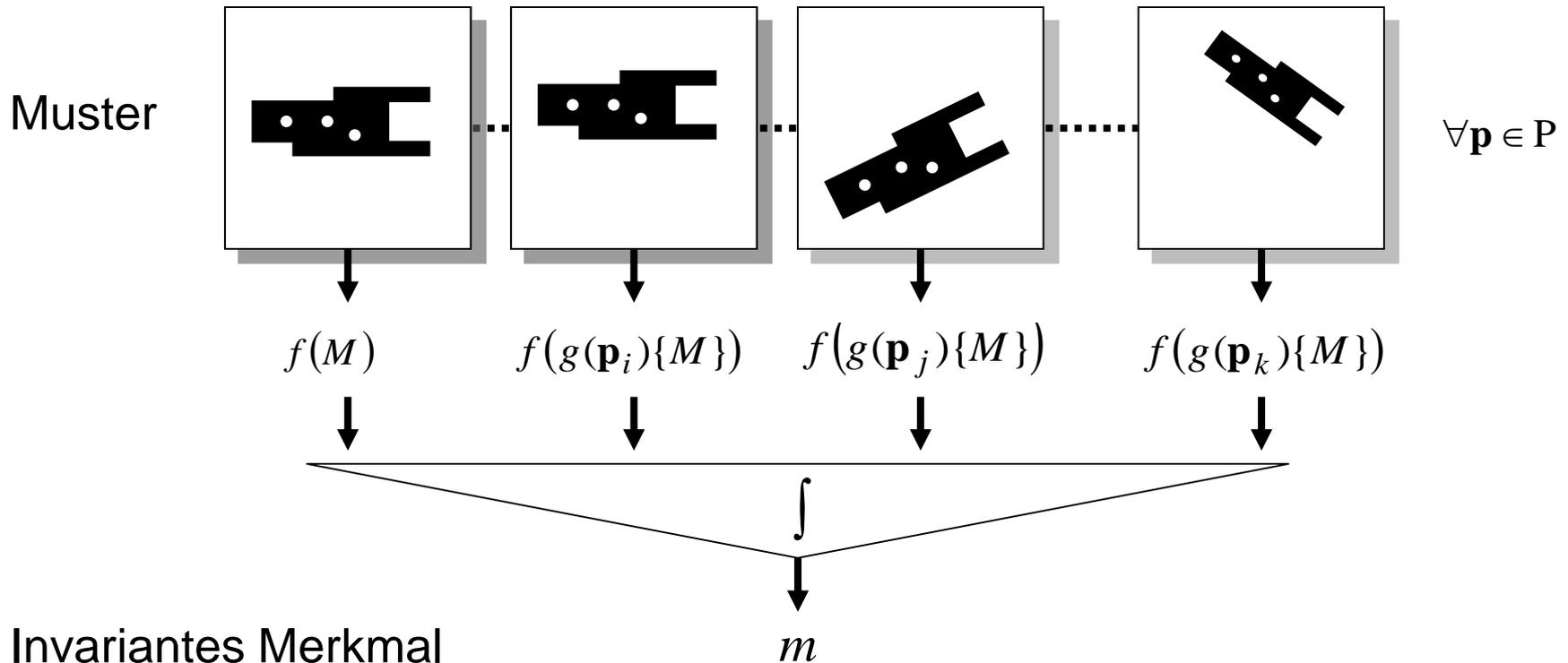
Lösung der daraus resultierenden partiellen DGI \rightarrow Lie-Theorie

(3) **Normalisierung**: Zurückführung auf spezielle Punkte des Orbits, z.B. den Punkt maximaler Krümmung. Bsp.: Fourierdeskriptoren

2.6.3. Konstruktion invarianter Merkmale

Invariante Merkmale durch Gruppenmittelung

$$m = \int_G f(g\{M\})d g = \int_{\mathbf{p} \in P} f(g(\mathbf{p})\{M\})d \mathbf{p}$$



2.6.3. Konstruktion invarianter Merkmale

Bsp.: Invariante Merkmale für die Kongruenzgruppe C (ebene Bewegungen) mit lokalen Funktionen

Muster M : $b(x, y)$ Grauwertbild mit $(x, y) \in [0, L_x] \times [0, L_y]$

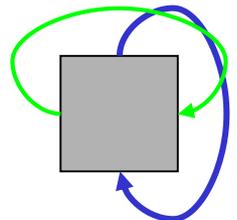
$$g(\mathbf{p})\{M\} = g(\tau_x, \tau_y, \varphi)\{b(x, y)\} = b(x', y')$$

mit:
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}$$

$$m = \frac{1}{2\pi L_x L_y} \int_{\tau_y=0}^{L_y} \int_{\tau_x=0}^{L_x} \int_{\varphi=0}^{2\pi} f(b(x', y')) d\varphi d\tau_x d\tau_y$$

$f(\cdot)$: „lokale Funktion“, z.B.: Monom benachbarter Bildpunkte

Problematik **Bildrand**: Bildkoordinaten x, y werden modulo der Bilddimension verstanden, d.h. Bild wird als „Torus“ behandelt.
→ Die Bewegungen werden zyklisch.



2.6.3. Konstruktion invarianter Merkmale

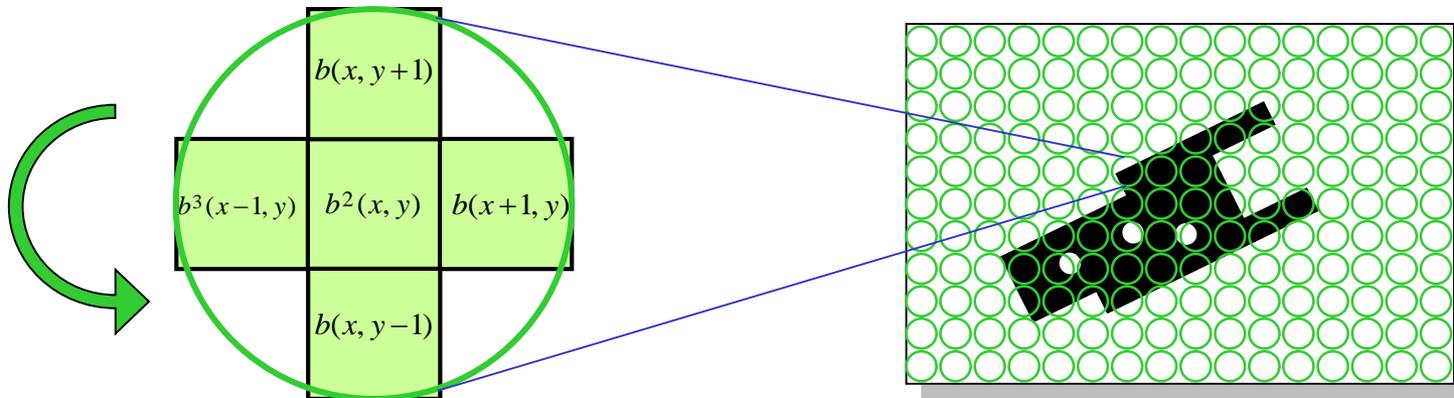
Bsp.: Invariante Merkmale für die Kongruenzgruppe C

Man kann zeigen, dass f und b vertauschbar sind. \rightarrow Aufwandsreduktion!

Anstatt das Bild über den ganzen Parameterraum zu variieren, transformiert man die lokale Funktion. D.h. man bewegt die lokale Funktion f gemäß der vorliegenden Gruppe über das ganze Bild (alle Verschiebungen und Drehungen) und integriert über die Funktionswerte.

Praktische Durchführung: Näherung der Integration durch Summation auf dem Pixelraster, Rotationen nur um eine endliche Anzahl von Winkeln, Interpolation der Zwischenwerte, z.B. durch bilineare Interpolation.

$f(\cdot)$: z.B.: Monom: $b^2(x, y)b(x, y-1)b^3(x-1, y)b(x, y+1)b(x+1, y)$



Auswertung der lokalen Funktion f in allen Rotationslagen an allen Bildpunkten und anschließende Summation $\rightarrow m$

2.6.3. Konstruktion invarianter Merkmale

Differentielle Methode

Gruppen \supset parametrische Gruppen \supset Lie-Gruppen

Parametrierung zusammen mit einer Metrik $D(\mathbf{p}, \mathbf{p}')$ auf dem Parameter-
raum P induziert einen Nachbarschaftsbegriff innerhalb der Gruppe G .

→ topologische Gruppe

Lie-Gruppe: parametrische Gruppe $G = \{g(\mathbf{p})\{.\} | \mathbf{p} \in P\}$, deren
Gruppenprodukt und inverse Elemente stetig von \mathbf{p} abhängen.

→ Infinitesimale Transf.: $g(d\mathbf{p})\{.\}$ verändern die Muster nur infinitesimal.

Eine Gruppe G heißt **zusammenhängend**, falls sich jedes Element g von G
durch einen Weg, der vollständig in G verläuft, mit dem Einselement e
verbinden lässt.

Es gilt: In einer zusammenhängenden Lie-Gruppe lässt sich jedes Element
 $g \in G$ als **Produkt von infinitesimalen Transformationen** darstellen.

Idee der differentiellen Methode: Bei zusammenhängenden Lie-Gruppen
folgt aus der Invarianz bzgl. infinitesimaler Transformationen die Invarianz
bzgl. der ganzen Gruppe.

2.6.3. Konstruktion invarianter Merkmale

Differentielle Methode

Es gilt: Bei Konstruktion invarianter Merkmale mit der **differentiellen Methode** bezüglich **zusammenhängender Lie-Gruppen** genügt die Betrachtung von $\mathbf{p} = \mathbf{0}$.

$$\left. \frac{\partial m(g(\mathbf{p})\{M\})}{\partial p_l} \right|_{\mathbf{p}=\mathbf{0}} = 0 \quad \text{für } l = 1, \dots, \dim(P)$$

$$M' = g(\mathbf{p})\{M\}: \quad M \rightarrow M$$

$$m(.): \quad M \rightarrow \mathbb{R}$$

2.6.3. Konstruktion invarianter Merkmale

Beispiel: Muster ist ein Bildsignal

$$b(x, y) \in \mathbf{M}: \Omega \rightarrow \mathbb{R}$$

$$\frac{\partial m(g(\mathbf{p})\{b(x, y)\})}{\partial p_l} = \frac{\partial m(b'(\xi, \eta; \mathbf{p}))}{\partial p_l} = \sum_{(\xi, \eta) \in \Omega} \frac{\partial m}{\partial b'(\xi, \eta; \mathbf{p})} \frac{\partial b'(\xi, \eta; \mathbf{p})}{\partial p_l}$$

Beispiel: Drehgruppe (ist eine zusammenhängende Lie-Gruppe.)

$$g(\mathbf{p})\{M\} = g(\varphi)\{b(x, y)\} \quad \text{mit} \quad \mathbf{p} = \varphi \in [0, 2\pi)$$

$$g(\varphi)\{b(x, y)\} = b(x \cos \varphi - y \sin \varphi, x \sin \varphi + y \cos \varphi) = b'(\xi(\varphi), \eta(\varphi)) = b(\xi, \eta)$$

$$\frac{\partial m(g(\varphi)\{b(x, y)\})}{\partial \varphi} = \frac{\partial m(b(\xi, \eta))}{\partial \varphi} = \sum_{(\xi, \eta) \in \Omega} \frac{\partial m}{\partial b(\xi, \eta)} \left(\frac{\partial b}{\partial \xi} \frac{\partial \xi}{\partial \varphi} + \frac{\partial b}{\partial \eta} \frac{\partial \eta}{\partial \varphi} \right)$$

2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

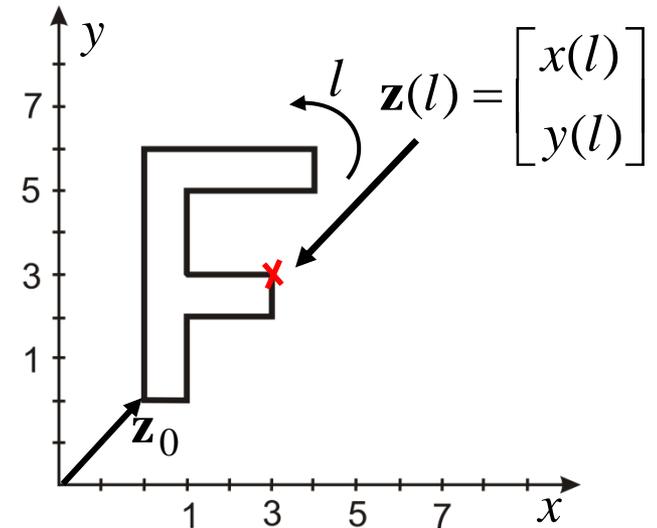
Bsp.: Fourierdeskriptoren für geschlossene ebene Konturen

$$\mathbf{z}(l) = \begin{bmatrix} x(l) \\ y(l) \end{bmatrix}$$

Parameter: $l \in [0, L)$

Periodische Funktion:

$$\mathbf{z}(l) = \mathbf{z}(l + L) \quad \forall l \in [0, L)$$



Periodische Funktionen lassen sich durch Fourierreihen darstellen.

$$\mathbf{z}(l) = \begin{bmatrix} x(l) \\ y(l) \end{bmatrix} = \sum_{n=-\infty}^{\infty} \mathbf{F}_n \exp(j2\pi nl / L) \quad n = 0, \pm 1, \pm 2, \dots, \infty$$

$$\mathbf{F}_n = \begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \frac{1}{L} \int_0^L \begin{bmatrix} x(l) \\ y(l) \end{bmatrix} \exp(-j2\pi nl / L) dl \quad X_n^* = X_{-n} \quad Y_n^* = Y_{-n}$$

2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

$$\mathbf{z}(l) = \begin{bmatrix} x(l) \\ y(l) \end{bmatrix} = \sum_{n=-\infty}^{\infty} \mathbf{F}_n \exp(j2\pi nl / L) \quad n = 0, \pm 1, \pm 2, \dots, \infty$$

$$\mathbf{F}_n = \begin{bmatrix} X_n \\ Y_n \end{bmatrix} \quad X_n^* = X_{-n} \quad Y_n^* = Y_{-n}$$

Komplexe Schreibweise:

$$z(l) = x(l) + jy(l) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi nl / L) \quad c_n = \frac{1}{L} \int_0^L z(l) \exp(-j2\pi nl / L) dl$$
$$c_n = X_n + jY_n$$

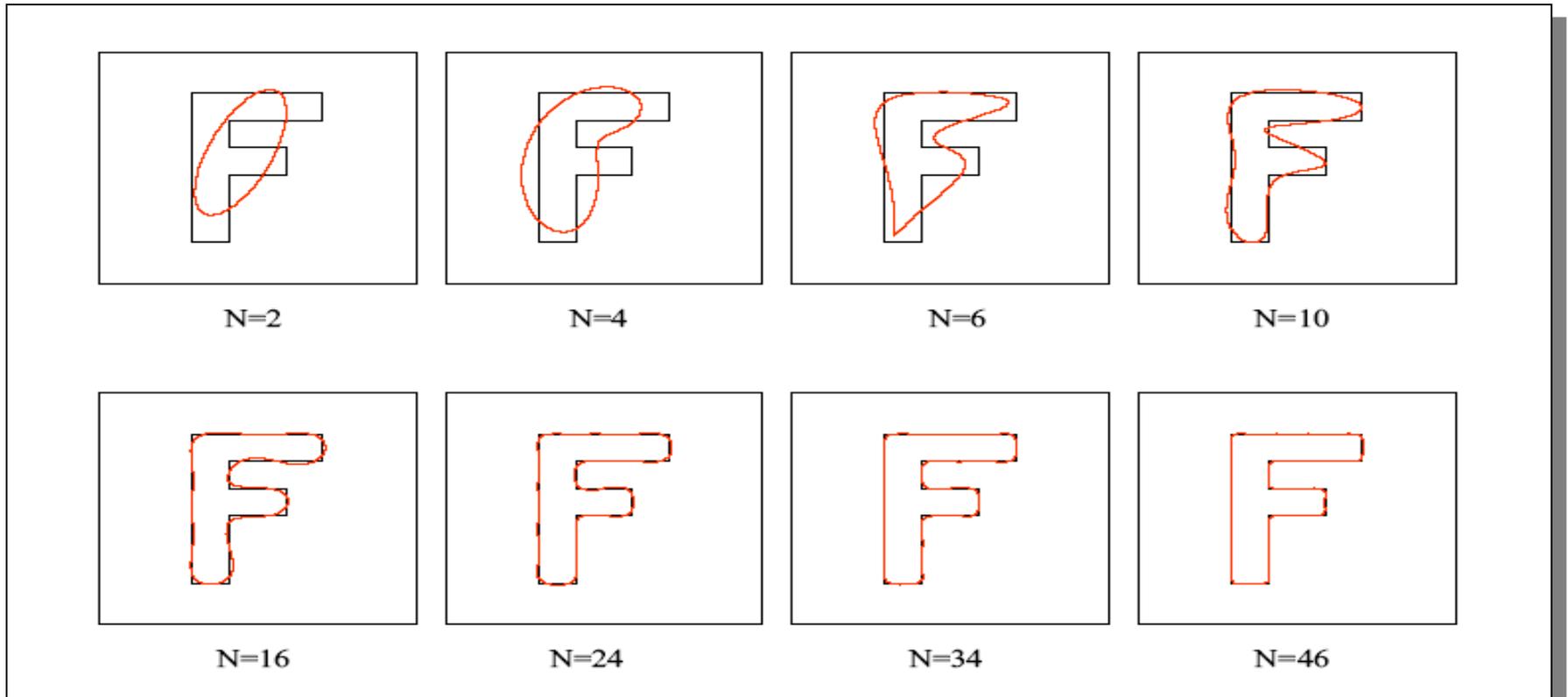
Konturschwerpunkt: c_0

Translationen wirken nur auf c_0 und lassen die $c_i, |i| > 0$ unverändert.

2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

Fouriersynthese einer geschlossenen Kontur

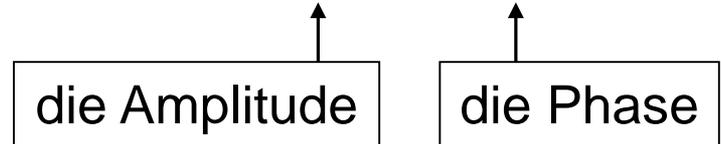


Quelle: <http://imb.informatik.uni-freiburg.de/lectures/mustererkennung/WS0405/material.html>

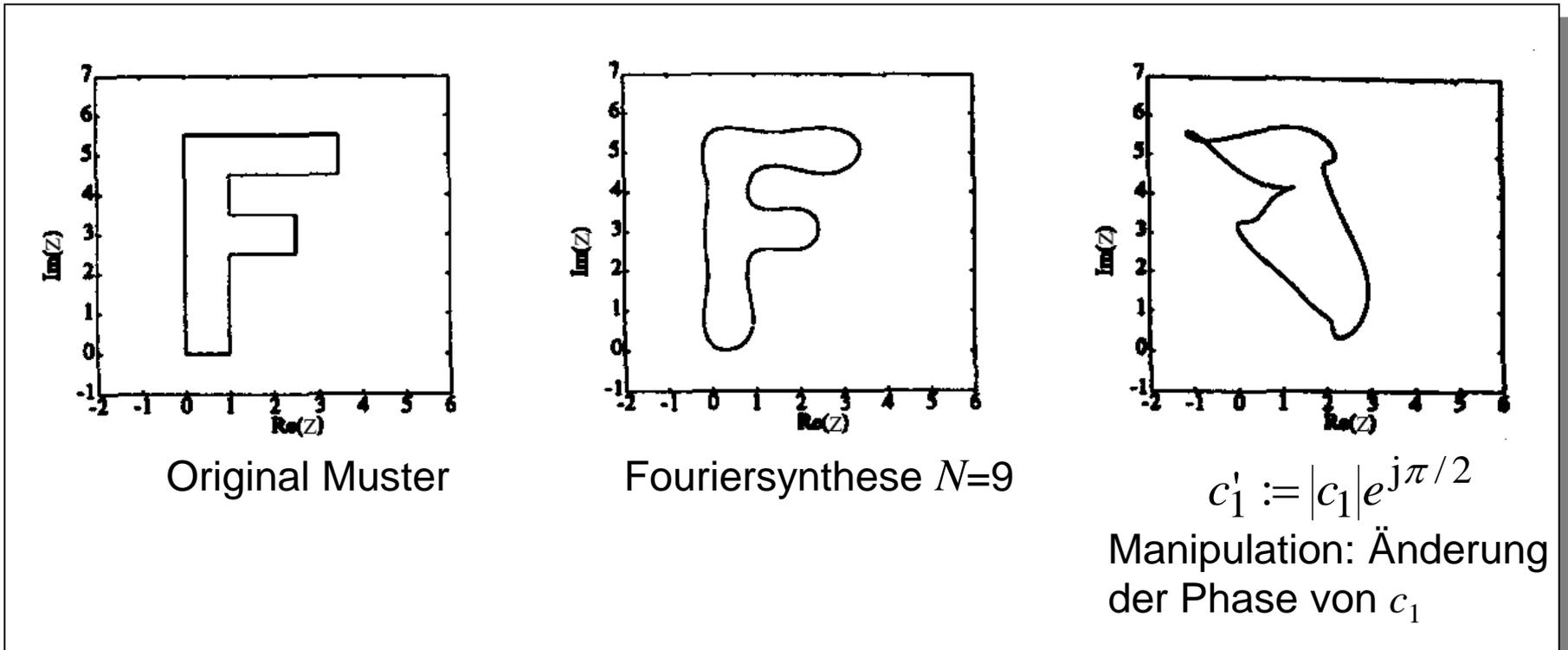
2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

Polare Schreibweise der Fourierkoeffizienten: $c_n = |c_n| \exp(j\phi_n)$



Beispiel zur Bedeutung der Phase:



Quelle: <http://imb.informatik.uni-freiburg.de/lectures/mustererkennung/WS0405/material.html>

2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

Fourierreihenoeffizienten: $c_n = \frac{1}{L} \int_{\text{primitive Periode}} z(l) \exp(-j2\pi nl/L) dl$

Ansätze für die Normalisierung:

$$z'(l) = z(l) + a \rightarrow c'_n = c_n + a\delta_0^n$$

→ **Translationsinvarianz**: Ignorieren von c_0

$$z'(l) = az(l) \rightarrow c'_n = ac_n$$

→ **Skalierungsinvarianz**: Berechnung von Quotienten c_n / c_m

$$z'(l) = z(l+a) \rightarrow c'_n = c_n \exp(j2\pi na/L)$$

→ **Rotationsinvarianz**: Berechnung von Phasendifferenzen $n\phi_m - m\phi_n$

2.6.3. Konstruktion invarianter Merkmale

Normalisierung, Bsp.: Fourierdeskriptoren für Konturen

Es gilt: Bezüglich geschlossener Konturen sind die Merkmale:

$$m_{nq} := \frac{|c_n|}{|c_q|} \exp(j(\phi_n + \alpha\phi_r - \beta\phi_q)), \quad n = \pm 1, \pm 2, \dots, \pm \infty$$

$$\alpha := \frac{q-n}{r-q}, \quad \beta := \frac{r-n}{r-q}, \quad r := q+s, \quad q \in \mathbb{N}^+$$

invariant bezüglich der Gruppe C der ebenen Bewegungen (Translationen und Rotationen) und der radialen Streckung (Skalierung).

s : Grad der Rotationssymmetrie

Def.: Geht ein Muster bei einer Rotation um seinem Schwerpunkt um $2\pi/s$ in sich selbst über, so liegt eine Rotationssymmetrie vom Grad s vor.

Gut geeignet für die Normierung ist der i.d.R. dominante Koeffizient c_1 .

Falls zudem keine Rotationssymmetrie vorliegt ist $s = 1$.

$$q = 1, s = 1 \Rightarrow r = 2, \alpha = 1 - n, \beta = 2 - n$$

2.7. Reduktion der Dimension des Merkmalsraumes

Problem: Hohe Dimension des Merkmalsraumes ist ungünstig.

Wunsch: Verringerung der Dimension des Merkmalsraumes bei möglichst geringem Verlust an klassifikationsrelevantem Informationsgehalt

Ansatz: Projektion des Merkmalsvektors der Dimension d auf einen Unterraum geringerer Dimension d' .

Methoden:

- **Hauptkomponentenanalyse:** Berechnet die Projektion, die gegebene Daten in Sinne der kleinsten Quadrate optimal repräsentiert.
 - Lineare PCA (Principal Component Analysis)
 - Kernelized PCA (nichtlinear)
- **Independent Component Analysis (ICA)**
- **Multiple Discriminant Analysis:** Berechnet die Projektion, die gegebene Daten in Sinne der kleinsten Quadrate optimal trennt.
- **Auswahl** einer Teilmenge von Merkmalen → Verwerfung von Merkmalen

2.7.1. Hauptkomponentenanalyse

HKA: Hauptkomponentenanalyse

Andere Bezeichnungen: Principal Component Analysis (PCA),
Karhunen-Loève-Transformation

Datensatz: $D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$

Erster Schritt: Welcher Punkt \mathbf{m}_0 im Merkmalsraum repräsentiert die Daten D mit minimalem quadratischem Fehler?

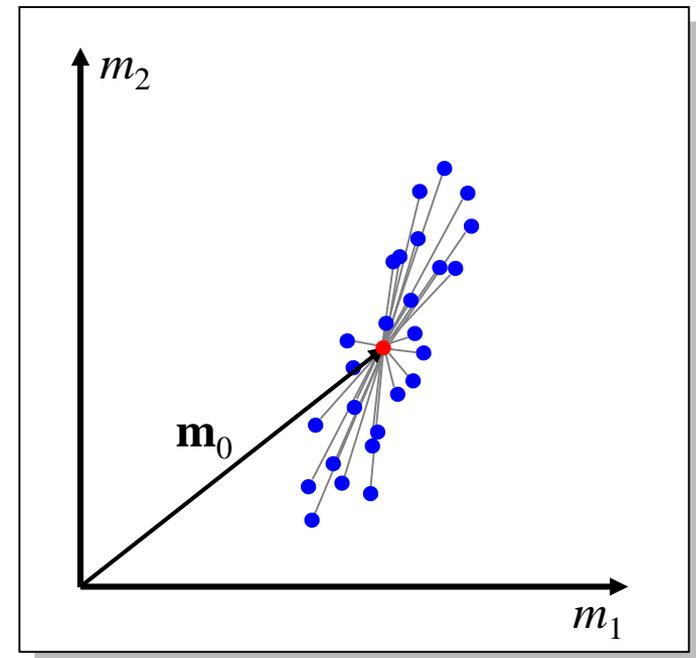
Fehlermaß:

$$J_0(\mathbf{m}) = \sum_{k=1}^N \|\mathbf{m} - \mathbf{m}_k\|^2$$

Gesucht: \mathbf{m}_0 , so dass $J_0(\mathbf{m}_0)$ minimal

Resultat:

$$\mathbf{m}_0 = \bar{\mathbf{m}} := \frac{1}{N} \sum_{k=1}^N \mathbf{m}_k$$



2.7.1. Hauptkomponentenanalyse

Beweis:

$$\begin{aligned} J_0(\mathbf{m}_0) &= \sum_{k=1}^N \|\mathbf{m}_0 - \mathbf{m}_k\|^2 = \sum_{k=1}^N \|(\mathbf{m}_0 - \bar{\mathbf{m}}) - (\mathbf{m}_k - \bar{\mathbf{m}})\|^2 = \\ &= \sum_{k=1}^N \|(\mathbf{m}_0 - \bar{\mathbf{m}})\|^2 - 2 \sum_{k=1}^N (\mathbf{m}_0 - \bar{\mathbf{m}})^T (\mathbf{m}_k - \bar{\mathbf{m}}) + \sum_{k=1}^N \|(\mathbf{m}_k - \bar{\mathbf{m}})\|^2 = \\ &= \sum_{k=1}^N \|(\mathbf{m}_0 - \bar{\mathbf{m}})\|^2 - 2(\mathbf{m}_0 - \bar{\mathbf{m}})^T \sum_{k=1}^N (\mathbf{m}_k - \bar{\mathbf{m}}) + \sum_{k=1}^N \|(\mathbf{m}_k - \bar{\mathbf{m}})\|^2 = \\ &= \left. \sum_{k=1}^N \|(\mathbf{m}_0 - \bar{\mathbf{m}})\|^2 + \sum_{k=1}^N \|(\mathbf{m}_k - \bar{\mathbf{m}})\|^2 \right\} \implies J_0(\mathbf{m}_0) = \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 \\ &\quad \text{Wird minimal für: } \mathbf{m}_0 = \bar{\mathbf{m}} \end{aligned}$$

2.7.1. Hauptkomponentenanalyse

Zweiter Schritt: Welche Gerade im Merkmalsraum repräsentiert die Daten D mit minimalem quadratischem Fehler?

Gerade h : $\mathbf{m} = \bar{\mathbf{m}} + a\mathbf{e}$, $a \in \mathbb{R}$, $\mathbf{e} \in \mathbb{R}^d$, $\|\mathbf{e}\| = 1$

$\hat{\mathbf{m}}_k = \bar{\mathbf{m}} + a_k \mathbf{e}$ mit $a_k =$ Parameterwert der orthogonalen Projektion des Punktes \mathbf{m}_k auf diese Gerade.

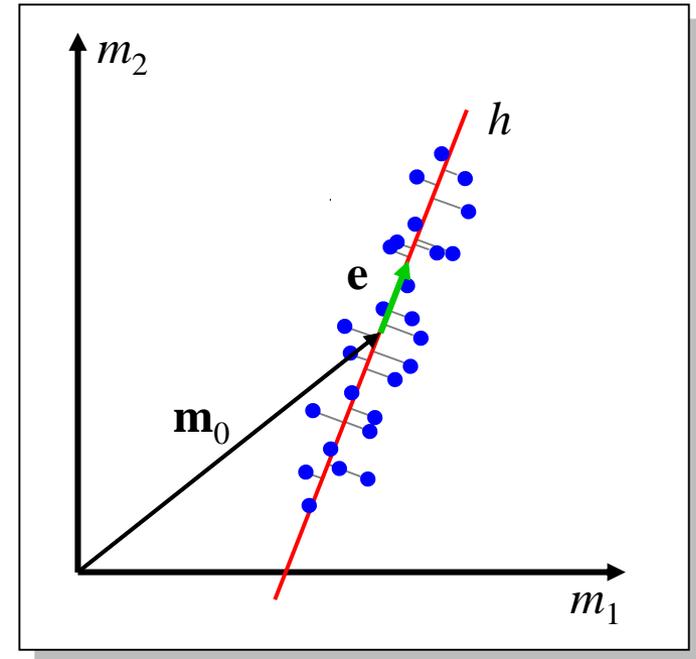
Fehlermaß:

$$J_1(a_1, \dots, a_N, \mathbf{e}) = \sum_{k=1}^N \|\bar{\mathbf{m}} + a_k \mathbf{e} - \mathbf{m}_k\|^2$$

Zunächst gesucht: a_1, \dots, a_N ,
so dass $J_1(a_1, \dots, a_N, \mathbf{e})$ minimal.

Resultat:

$$a_k = \mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}})$$



2.7.1. Hauptkomponentenanalyse

Beweis:

$$J_1(a_1, \dots, a_N, \mathbf{e}) = \sum_{k=1}^N \|(\bar{\mathbf{m}} + a_k \mathbf{e}) - \mathbf{m}_k\|^2 = \sum_{k=1}^N \|a_k \mathbf{e} - (\mathbf{m}_k - \bar{\mathbf{m}})\|^2 =$$

$$= \sum_{k=1}^N \|a_k \mathbf{e}\|^2 - 2 \sum_{k=1}^N a_k \mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}}) + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 =$$

$$= \sum_{k=1}^N a_k^2 \|\mathbf{e}\|^2 - 2 \sum_{k=1}^N a_k \mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}}) + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2$$

$$\frac{\partial J_1(a_1, \dots, a_N, \mathbf{e})}{\partial a_l} = 2a_l - 2\mathbf{e}^T (\mathbf{m}_l - \bar{\mathbf{m}}) \stackrel{!}{=} 0$$

$$\Leftrightarrow a_l = \mathbf{e}^T (\mathbf{m}_l - \bar{\mathbf{m}}) \quad \text{q.e.d.}$$

2.7.1. Hauptkomponentenanalyse

Berechnung des optimalen Richtungsvektors \mathbf{e} :

Streumatrix:
$$\mathbf{S} := \sum_{k=1}^N (\mathbf{m}_k - \bar{\mathbf{m}})(\mathbf{m}_k - \bar{\mathbf{m}})^T$$

Rechnung:

$$\begin{aligned} J_1(\mathbf{e}) &= \sum_{k=1}^N a_k^2 - 2 \sum_{k=1}^N a_k \mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}}) + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 = \\ &= \sum_{k=1}^N a_k^2 - 2 \sum_{k=1}^N a_k^2 + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 = - \sum_{k=1}^N a_k^2 + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 = \\ &= - \sum_{k=1}^N [\mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}})]^2 + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 \\ &= - \sum_{k=1}^N \mathbf{e}^T (\mathbf{m}_k - \bar{\mathbf{m}})(\mathbf{m}_k - \bar{\mathbf{m}})^T \mathbf{e} + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 = \\ &= -\mathbf{e}^T \mathbf{S} \mathbf{e} + \sum_{k=1}^N \|\mathbf{m}_k - \bar{\mathbf{m}}\|^2 \end{aligned}$$

→ $\mathbf{e}^T \mathbf{S} \mathbf{e}$ maximieren unter der Nebenbedingung: $\|\mathbf{e}\| = 1$

2.7.1. Hauptkomponentenanalyse

Rechnung:

$$f(\mathbf{e}, \lambda) = \mathbf{e}^T \mathbf{S} \mathbf{e} - \lambda(\mathbf{e}^T \mathbf{e} - 1) \quad \lambda: \text{Lagrange Multiplikator}$$

$$\nabla_{\mathbf{e}} f = 2\mathbf{S}\mathbf{e} - 2\lambda\mathbf{e} = \mathbf{0}$$

$$\mathbf{S}\mathbf{e} = \lambda\mathbf{e}$$

$$\mathbf{e}^T \mathbf{S} \mathbf{e} = \lambda \mathbf{e}^T \mathbf{e} = \lambda \Rightarrow \lambda \text{ muss maximal gew\u00e4hlt werden.}$$

Resultat:

Die **optimale Gerade** geht durch den Mittelwert $\bar{\mathbf{m}}$ und hat als Richtungsvektor \mathbf{e} den Eigenvektor zum gr\u00f6\u00dften Eigenwert λ der Streumatrix \mathbf{S} .

2.7.1. Hauptkomponentenanalyse

Dritter Schritt:

Welcher affine, d' -dimensionale Unterraum des Merkmalsraumes repräsentiert die Daten D mit minimalem quadratischem Fehler?

Ansatz: $\mathbf{m} = \bar{\mathbf{m}} + \sum_{i=1}^{d'} a_i \mathbf{e}_i$ $\{\mathbf{e}_i\}$ bilden eine orthonormale Basis

Fehlermaß:

$$J_{d'} = \sum_{k=1}^N \left\| \left(\bar{\mathbf{m}} + \sum_{i=1}^{d'} a_{ki} \mathbf{e}_i \right) - \mathbf{m}_k \right\|^2, \quad d' \leq d$$

Resultat:

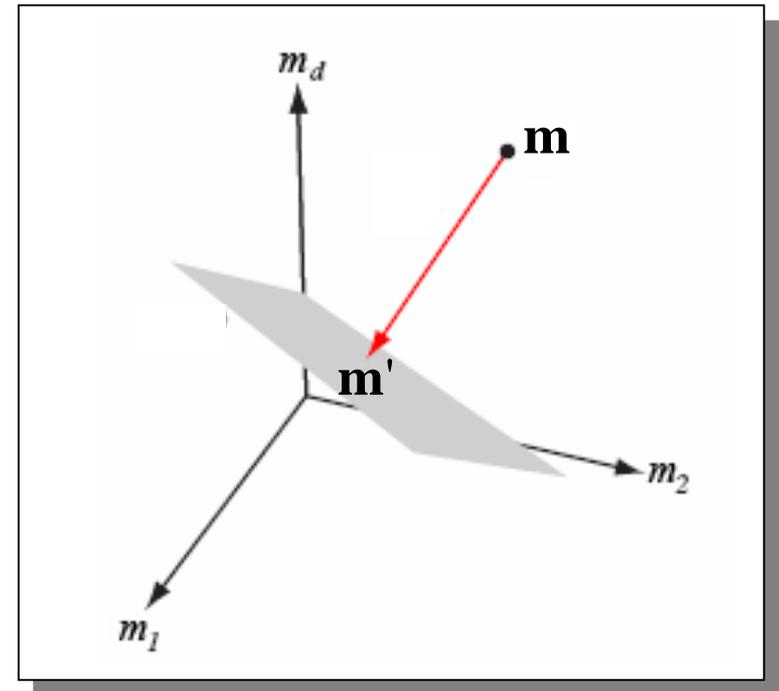
Der **optimale affine Unterraum** der Dimension d' geht durch den Mittelwert $\bar{\mathbf{m}}$ und wird von den Eigenvektoren \mathbf{e}_i zu den d' größten Eigenwerten λ_i der Streumatrix \mathbf{S} aufgespannt.

2.7.1. Hauptkomponentenanalyse

Vorgehensweise:

- Berechnung des Mittelwertvektors $\bar{\mathbf{m}}$ und der Streumatrix \mathbf{S}
- Berechnung der Eigenvektoren und Sortierung nach fallenden Eigenwerten
- Bildung einer Matrix \mathbf{A} mit den d' Eigenvektoren der d' größten Eigenwerte als Spalten

- Transformation $\mathbf{m}' = \mathbf{A}^T (\mathbf{m} - \bar{\mathbf{m}})$



Quelle: R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification

2.7.1. Hauptkomponentenanalyse

Diskussion HKA: Erwartungswert und Kovarianzmatrix

Sei \mathbf{m} ein Zufallsvektor.

Erwartungswert: $\boldsymbol{\mu} := \mathbf{E}\{\mathbf{m}\}$

Kovarianzmatrix: $\boldsymbol{\Sigma} := \text{Cov}\{\mathbf{m}\} = \mathbf{E}\{(\mathbf{m} - \boldsymbol{\mu})(\mathbf{m} - \boldsymbol{\mu})^T\}$

Eigenvektorzerlegung:

$$\boldsymbol{\Sigma} = \sum_{i=1}^d \kappa_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T = \mathbf{E} \begin{bmatrix} \kappa_1 & \cdot & & \\ \cdot & \cdot & 0 & \\ & 0 & \cdot & \\ & & & \kappa_d \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_1^T \\ \cdot \\ \cdot \\ \boldsymbol{\varepsilon}_d^T \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}^T$$

$\boldsymbol{\Lambda}$

Eigenvektoren:

$$\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_j = \delta_i^j$$

$$\mathbf{E}^T \mathbf{E} = \mathbf{I} \Rightarrow \mathbf{E}^T = \mathbf{E}^{-1} \Rightarrow \boldsymbol{\Lambda} = \mathbf{E}^T \boldsymbol{\Sigma} \mathbf{E}$$

Eigenwerte:

$$\kappa_i \geq 0 \quad i = 1, \dots, d \quad \text{o.B.d.A.: } \kappa_1 \geq \kappa_2 \geq \dots \geq \kappa_d$$

Praktisch: Kovarianzmatrix immer positiv definit: $\kappa_i > 0 \quad i = 1, \dots, d$

2.7.1. Hauptkomponentenanalyse

Diskussion HKA: Erwartungswert und Kovarianzmatrix

Karhunen-Loève Transformation: $\tilde{\mathbf{m}} := \mathbf{E}^T (\mathbf{m} - \boldsymbol{\mu})$

Erwartungswert: $E\{\tilde{\mathbf{m}}\} = \mathbf{0}$

Kovarianzmatrix:

$$\begin{aligned}\text{Cov}\{\tilde{\mathbf{m}}\} &:= E\{\mathbf{E}^T (\mathbf{m} - \boldsymbol{\mu})(\mathbf{m} - \boldsymbol{\mu})^T \mathbf{E}\} \\ &= \mathbf{E}^T E\{(\mathbf{m} - \boldsymbol{\mu})(\mathbf{m} - \boldsymbol{\mu})^T\} \mathbf{E} \\ &= \mathbf{E}^T \boldsymbol{\Sigma} \mathbf{E} = \boldsymbol{\Lambda}\end{aligned}$$

Varianz der Komponenten:

$$\kappa_i = \text{Var}\{\tilde{m}_i\}$$

Komponenten sind unkorreliert:

$$E\left\{(\tilde{m}_i - E\{\tilde{m}_i\})(\tilde{m}_j - E\{\tilde{m}_j\})\right\} = \kappa_i \delta_i^j$$

2.7.1. Hauptkomponentenanalyse

Diskussion HKA: Zusammenhang mit den empirischen Größen

Empirischer Mittelwert:
$$\bar{\mathbf{m}} := \frac{1}{N} \sum_{k=1}^N \mathbf{m}_k$$

Streumatrix:
$$\mathbf{S} := \sum_{k=1}^N (\mathbf{m}_k - \bar{\mathbf{m}})(\mathbf{m}_k - \bar{\mathbf{m}})^T$$

Erwartungstreue Schätzung des Erwartungswertes:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{m}}$$

$$\mathbb{E}\{\bar{\mathbf{m}}\} = \boldsymbol{\mu}$$

Erwartungstreue Schätzung der Kovarianzmatrix:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \mathbf{S}$$

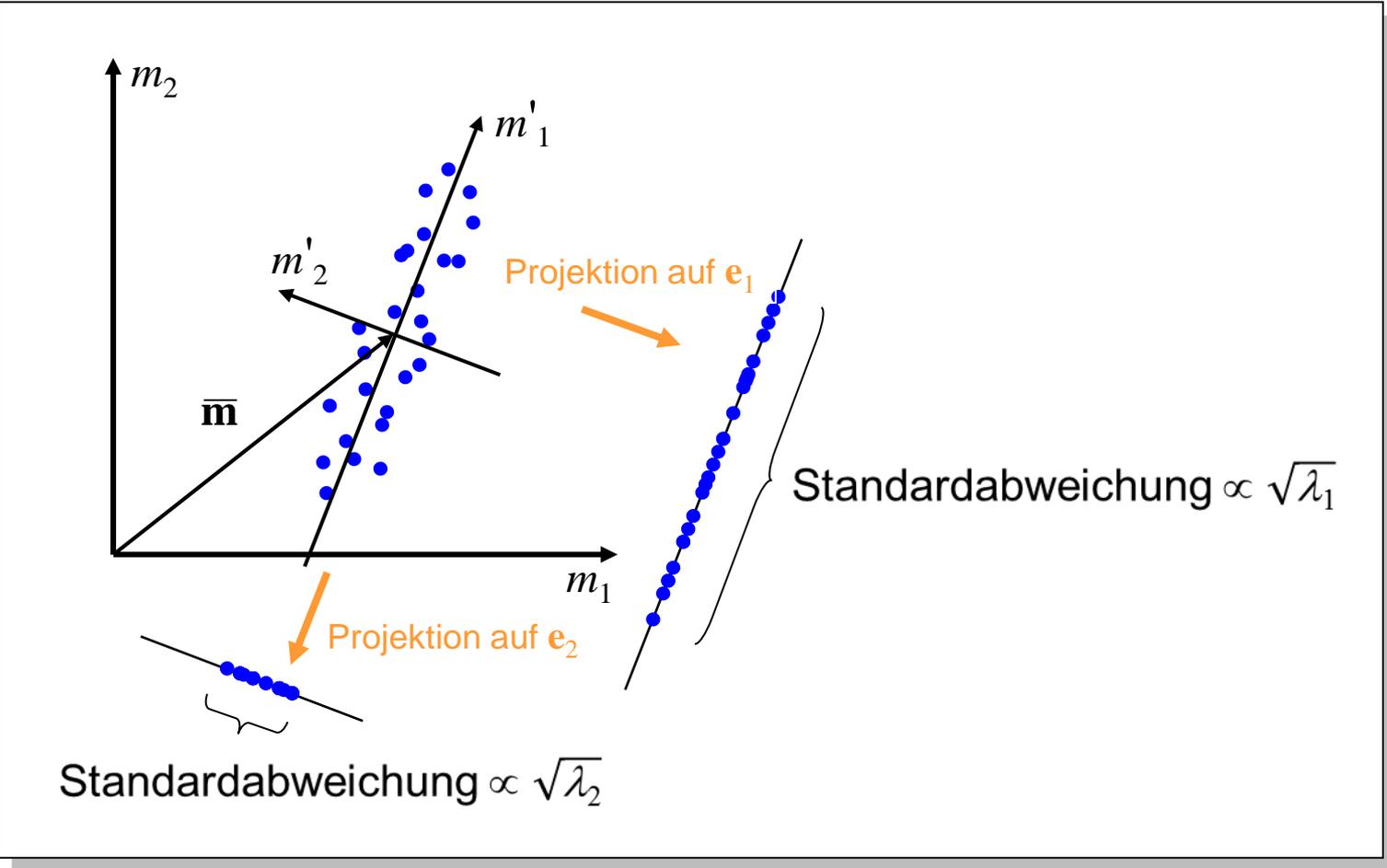
$$\mathbb{E}\left\{\frac{1}{N-1} \mathbf{S}\right\} = \boldsymbol{\Sigma} := \mathbb{E}\left\{(\mathbf{m} - \boldsymbol{\mu})(\mathbf{m} - \boldsymbol{\mu})^T\right\}$$

Schätzung der Varianz der transformierten Komponenten:

$$\hat{\kappa}_i = \frac{1}{N-1} \lambda_i$$

2.7.1. Hauptkomponentenanalyse

Diskussion HKA: Veranschaulichung



2.7.1. Hauptkomponentenanalyse

Diskussion HKA:

Greedy-Abspaltung der maximal variierenden Unterräume

Linear transformiertes Merkmal
mit maximaler „Varianz“ λ_1 :

$$m'_1 = \mathbf{e}_1^T (\mathbf{m} - \bar{\mathbf{m}})$$

Abspalten der von \mathbf{e}_1
aufgespannten Dimension:

$$\mathbf{m}_{[1]} := (\mathbf{I} - \mathbf{e}_1 \mathbf{e}_1^T) (\mathbf{m} - \bar{\mathbf{m}})$$

Linear transformiertes Merkmal
mit maximaler „Varianz“ λ_2 :

$$m'_2 = \mathbf{e}_2^T \mathbf{m}_{[1]}$$

Abspalten der von \mathbf{e}_2
aufgespannten Dimension:

$$\mathbf{m}_{[2]} := (\mathbf{I} - \mathbf{e}_2 \mathbf{e}_2^T) \mathbf{m}_{[1]}$$

Usw.

⋮

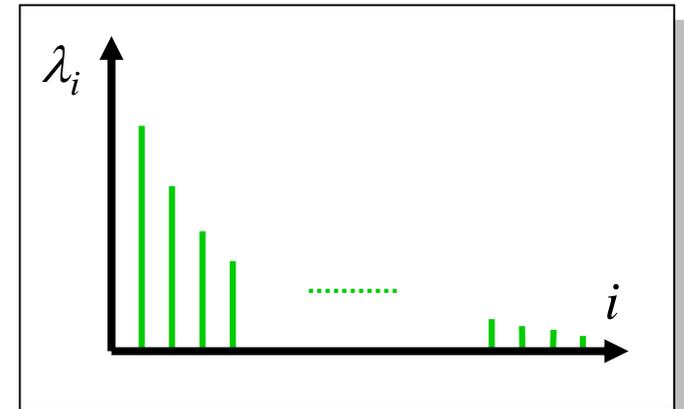
2.7.1. Hauptkomponentenanalyse

Diskussion HKA: Minimale Entropie der Varianzen, beste lineare Näherung

Die „Varianzen“ der transformierten Merkmale sind **maximal ungleichverteilt**: **minimale Entropie** der Varianzen.

$$H(\zeta_1, \dots, \zeta_d) = -\sum_{i=1}^d \zeta_i \log \zeta_i \text{ ist minimal.}$$

$$\zeta_i := \frac{\lambda_i}{\sum_{j=1}^d \lambda_j}$$



Die HKA liefert im Sinne der kleinsten quadratischen Abweichungen die **beste d' -dimensionale Approximation** der Daten D .

$$\sum_{k=1}^N \left\| (\mathbf{m}_k - \bar{\mathbf{m}}) - \sum_{i=1}^{d'} \mathbf{e}_i \mathbf{e}_i^T (\mathbf{m}_k - \bar{\mathbf{m}}) \right\|^2 = \sum_{i=d'+1}^d \lambda_i \text{ ist minimal.}$$

2.7.1. Hauptkomponentenanalyse

Diskussion HKA:

- Transformierte Merkmale sind paarweise **unkorreliert**.
- Varianzen der transformierten Merkmale sind **maximal ungleichverteilt**: minimale Entropie der „Varianzen“.
- Die HKA liefert im Sinne der kleinsten quadratischen Abweichungen die **beste d' -dimensionale Approximation** der Daten D .
- **HKA zielt nicht auf optimale Klassentrennbarkeit**, sondern auf die kleinste Summe der quadratischen Fehler im Hinblick auf Beschreibung der Gesamtstichprobe durch einen affinen Unterraum. Erfahrungsgemäß liefert die HKA meist trotzdem sehr gute Merkmalsräume verminderter Dimension.
- Ursprüngliche **Merkmale verlieren** durch HKA i. Allg. ihre **anschauliche Bedeutung**.
- HKA macht die Verteilung der transformierten **Merkmale** einer Gaußverteilung ähnlicher (Zentraler Grenzwertsatz).

2.7.1. Hauptkomponentenanalyse

Beispiel: Gesichtserkennung mit Eigengesichtern (Eigenfaces)

Grauwertbild eines Gesichtes: $b(x, y)$

Bild als Spaltenvektor: $\mathbf{m} := \text{Col}\{b(x, y)\} \in \mathbb{R}^{n^2}$, $d = n^2$

Datenmatrix: $\mathbf{M} := [\mathbf{m}_1 - \bar{\mathbf{m}}, \dots, \mathbf{m}_N - \bar{\mathbf{m}}]$

Streumatrix: $\mathbf{S} = \mathbf{M}\mathbf{M}^T$

$n^2 \times n^2$ -Matrix! Sehr groß! Hat nur $N-1$ nichtverschwindende Eigenvektoren.

Rechentrick: $\mathbf{L} := \mathbf{M}^T\mathbf{M}$

$N \times N$ -Matrix! Eigenvektoren leichter berechenbar.

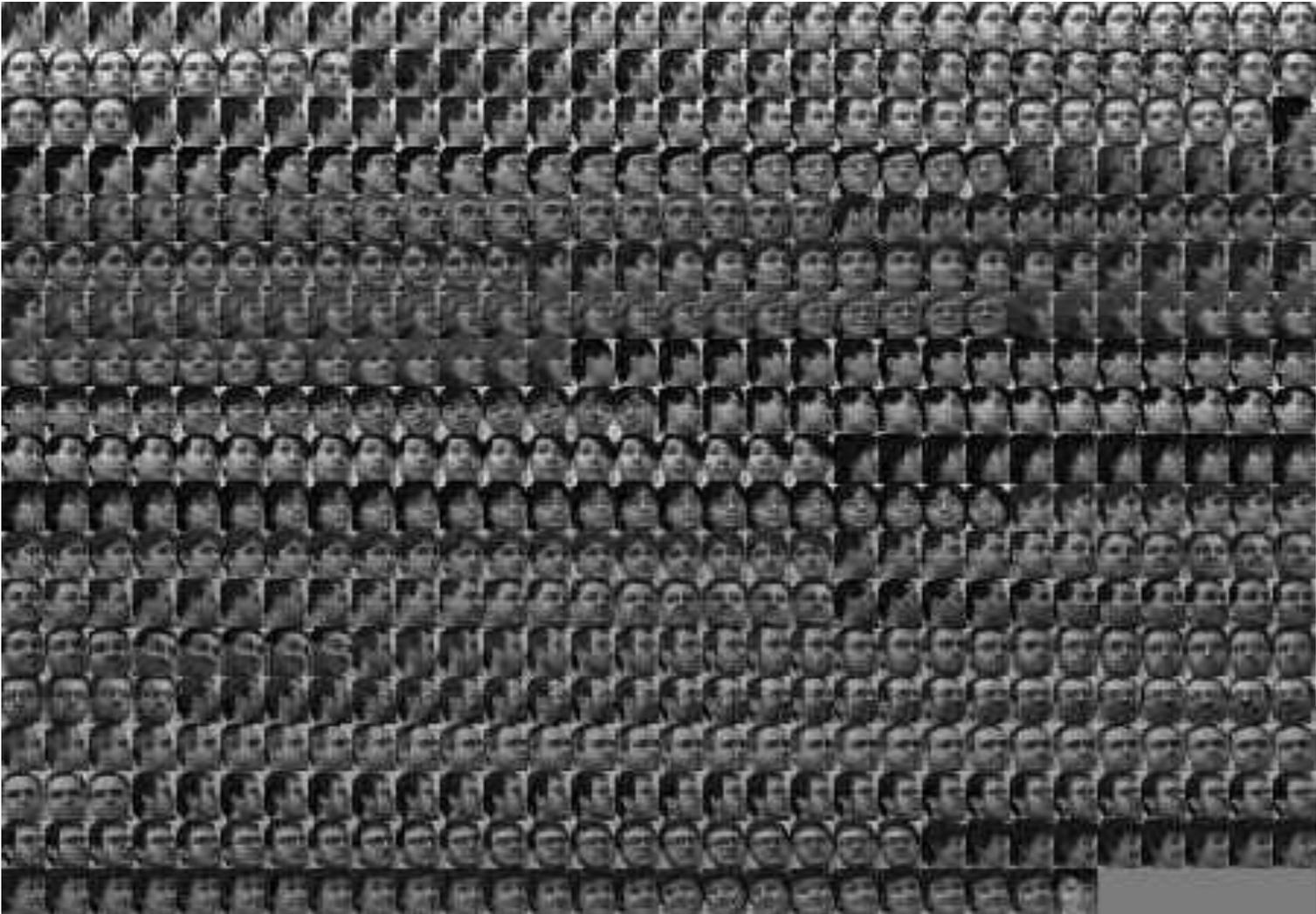
Es gilt: Die ersten $N-1$ ($N < n^2$) zu den größten Eigenwerten gehörenden Eigenvektoren von \mathbf{S} lassen sich aus den fallend sortierten Eigenvektoren $\boldsymbol{\eta}_i$ von \mathbf{L} berechnen mittels:

$$\mathbf{e}_i = \frac{\mathbf{M}\boldsymbol{\eta}_i}{\|\mathbf{M}\boldsymbol{\eta}_i\|} \quad i = 1, \dots, N$$

Eigengesichter: \mathbf{e}_i

2.7.1. Hauptkomponentenanalyse

Beispiel: Gesichtserkennung mit Eigengesichtern (Eigenfaces)



Mittleres Gesicht \bar{m}



Bildgröße:
30x24 Pixel

Quelle: <http://pubweb.northwestern.edu/~acb206/ece432/FaceRecReport.html>

© 2018 Lehrstuhl für Interaktive Echtzeitsysteme, KIT, alle Rechte einschließlich Kopier- und Weitergaberechte bei uns.

2.7.1. Hauptkomponentenanalyse

Beispiel: Gesichtserkennung mit Eigengesichtern (Eigenfaces)

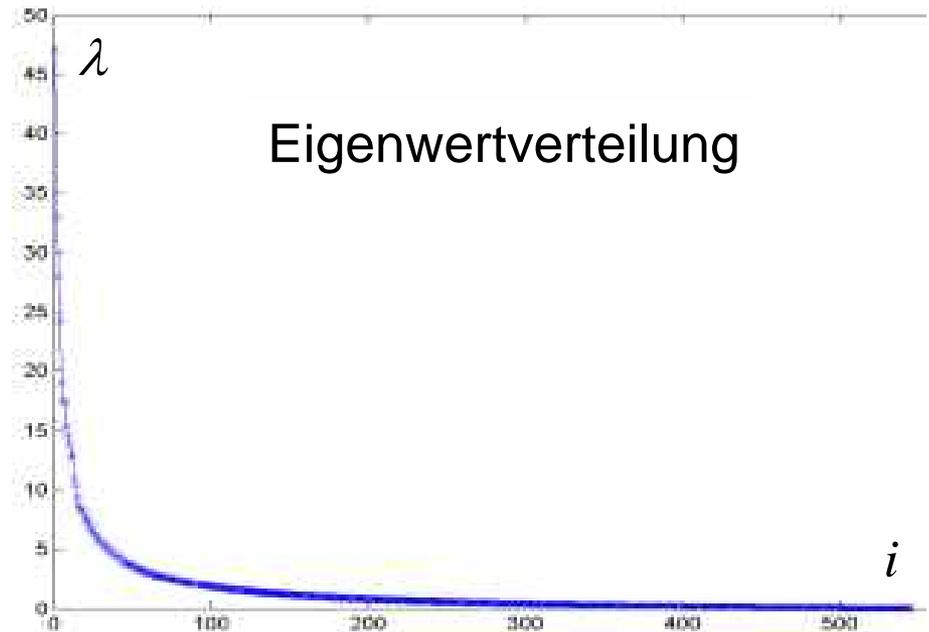
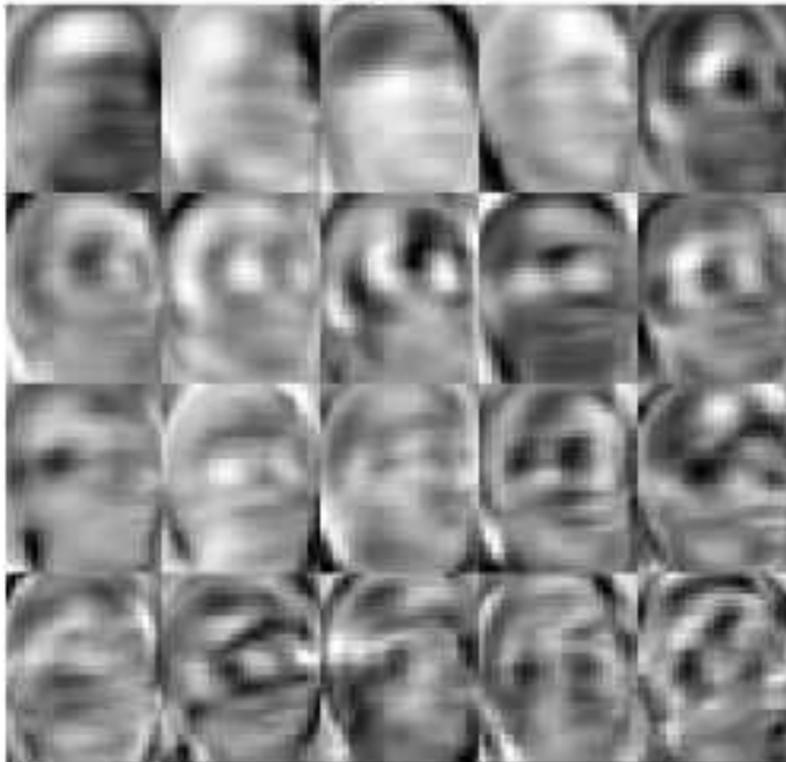


Mittleres Gesicht

Merkmale für die Gesichtserkennung:

Die ersten 20 Eigengesichter

$$\mathbf{m}' = [\mathbf{e}_1, \dots, \mathbf{e}_{d'}]^T (\mathbf{m} - \bar{\mathbf{m}})$$

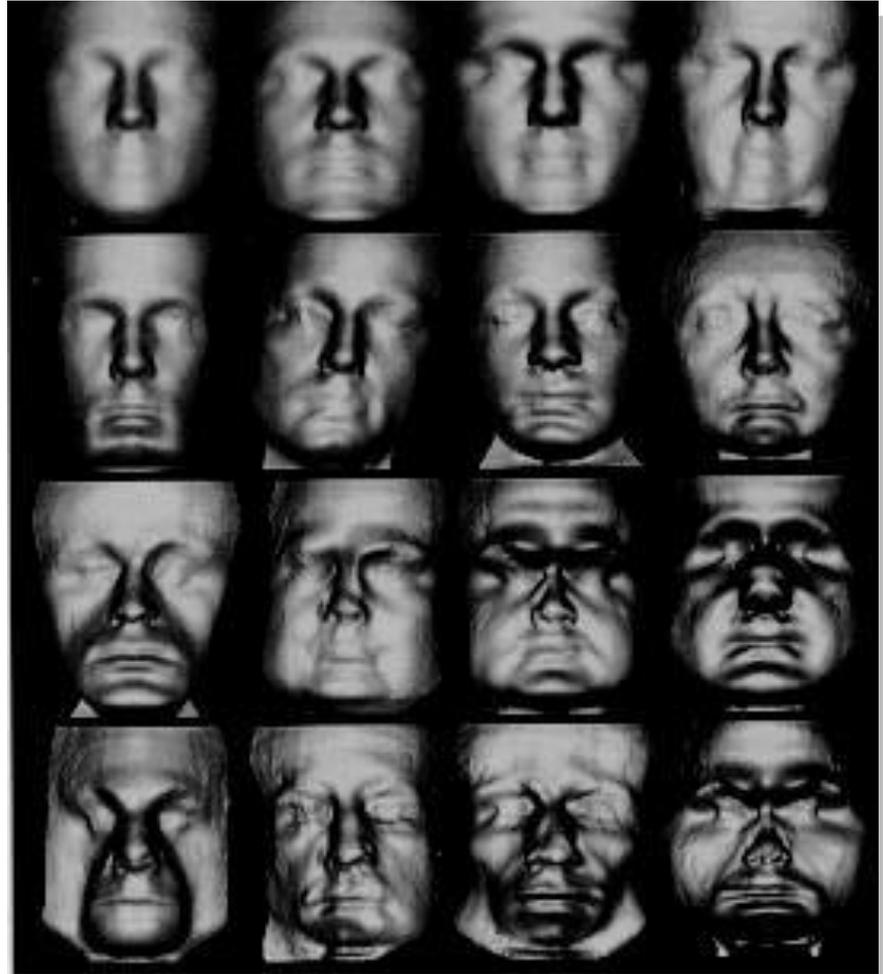


<http://pubweb.northwestern.edu/~acb206/ece432/FaceRecReport.html>

2.7.1. Hauptkomponentenanalyse

Beispiel: Gesichtserkennung mit Eigengesichtern (Eigenfaces), 3D-Daten

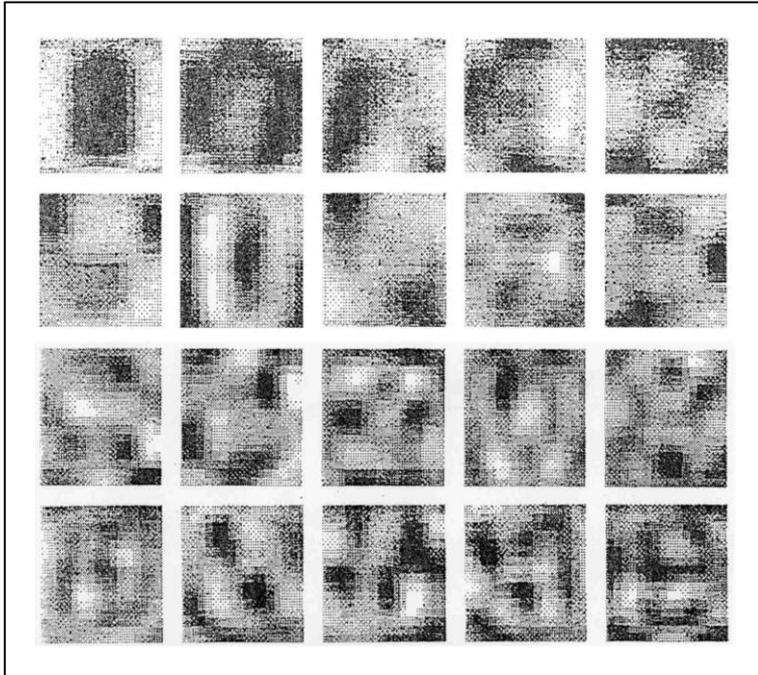
Erste 16 Eigengesichter
von 3D-Gesichtsdaten



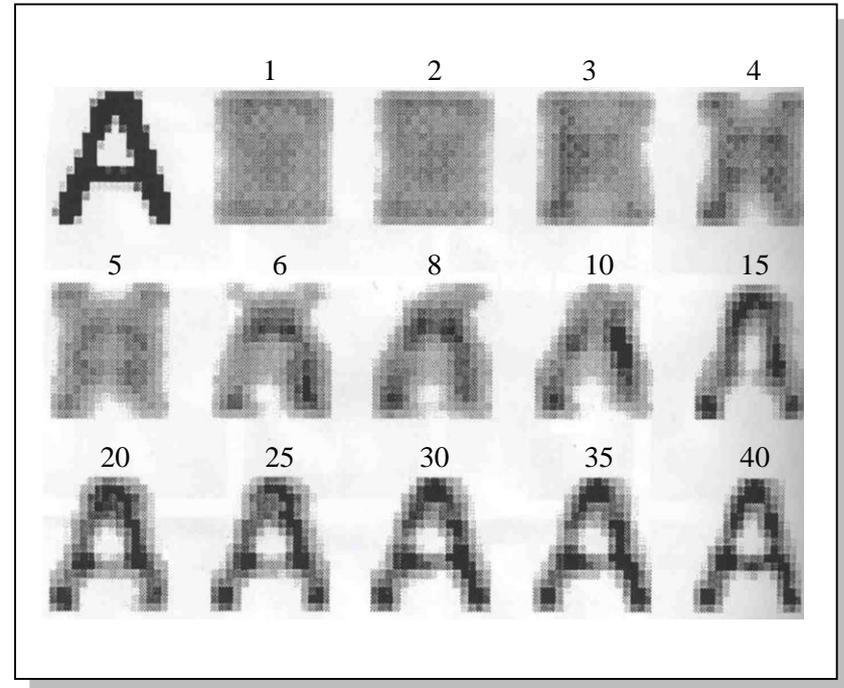
<http://www.ausbcomp.com/~bbott/wik/eigenface.htm>

2.7.1. Hauptkomponentenanalyse

Beispiel: Merkmale für die Zeichenerkennung



Die ersten 20 von 256 Eigenvektoren der Buchstaben A-Z.

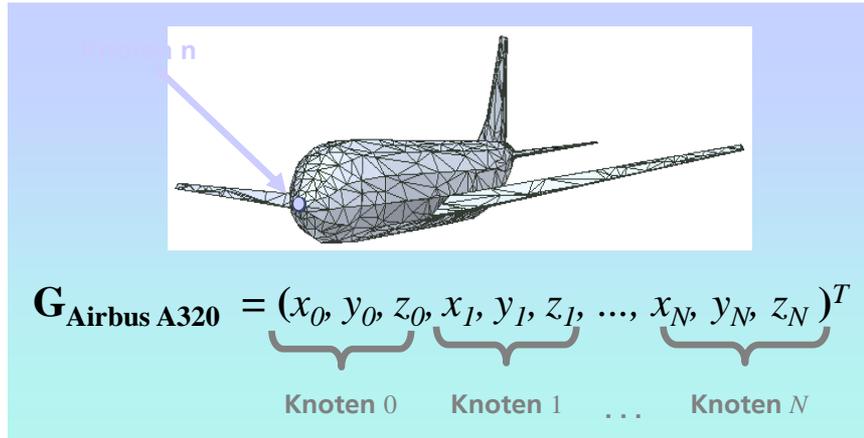


Originalbild (16×16 Pixel) eines Zeichens A. Schrittweise Rekonstruktion durch Berücksichtigung einer steigenden Zahl von Eigenkomponenten (1, 2, 3, 4, 5, 6, 8, 10, 15, 20, 25, 30, 35, 40)

Quelle: J. Schürmann, Pattern Classification, Wiley-Interscience 1996

2.7.1. Hauptkomponentenanalyse

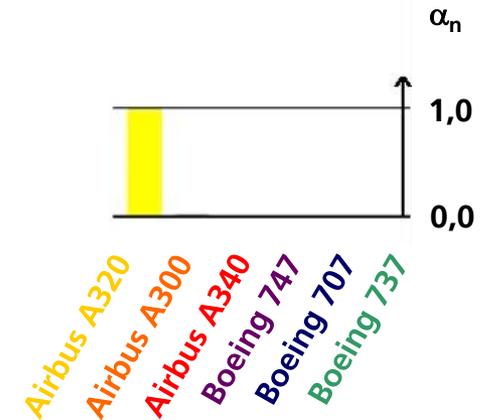
Beispiel: Modellgestützte Flugzeugtyperkennung



Quelle: Fraunhofer-IOSB Karlsruhe



$$\begin{aligned} \mathbf{G}_{\text{aktuell}} &= m_1 \mathbf{G}_{\text{Airbus A320}} + m_2 \mathbf{G}_{\text{Airbus A300}} \\ &+ m_3 \mathbf{G}_{\text{Airbus A340}} + m_4 \mathbf{G}_{\text{Boeing 747}} \\ &+ m_5 \mathbf{G}_{\text{Boeing 707}} + m_6 \mathbf{G}_{\text{Boeing 737}} \end{aligned}$$



Datengetriebenes adaptives Flugzeugmodell

2.7.1. Hauptkomponentenanalyse

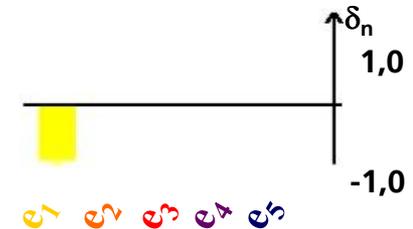
Beispiel: Modellgestützte Flugzeugtyperkennung

Hauptkomponentenanalyse

$$\mathbf{G} = \mathbf{G}_{\text{mean}} + m'_1 \mathbf{e}_1 + m'_2 \mathbf{e}_2 + m'_3 \mathbf{e}_3 + m'_4 \mathbf{e}_4 + \dots$$

↑ ↑
Transformierte Merkmale Eigenvektoren

Quelle: Fraunhofer-IOSB Karlsruhe



Relevanz der transformierten Merkmale ist stark fallend.

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Andere Bezeichnungen: Kernelized Principal Component Analysis (Kernel-PCA)

Idee: HKA in einem transformierten Merkmalsraum. Die nichtlineare Transformation der Merkmale $\varphi(\mathbf{m}_k)$, $k = 1, \dots, N$ kann die Repräsentation der Daten verbessern.

Quelle: B. Schölkopf, A.J. Smola, K.-R. Müller: *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*. Technical Report 44, Max-Planck-Institut für biologische Kybernetik, 1996.

Nichtlineare Transformation: $\varphi: \mathbb{R}^d \supseteq \mathbb{M} \rightarrow F$, \mathbf{m}_k a $\varphi(\mathbf{m}_k)$

Zunächst gelte: $\frac{1}{N} \sum_{k=1}^N \varphi(\mathbf{m}_k) = \mathbf{0}$

Streumatrix: $\mathbf{C} := \frac{1}{N} \sum_{k=1}^N \varphi(\mathbf{m}_k) \varphi(\mathbf{m}_k)^T \approx \mathbb{E} \{ \varphi(\mathbf{m}) \varphi(\mathbf{m})^T \}$

Ansatz Kernel-PCA: Lösen der Eigenwertgleichung im transformierten Merkmalsraum.

$$\lambda \mathbf{v} = \mathbf{C} \mathbf{v}$$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Aufgabenstellung: Gesucht sind die Eigenvektoren \mathbf{v} und Eigenwerte λ von \mathbf{C} .

Herleitung: $\lambda \mathbf{v} = \mathbf{C} \mathbf{v}$ mit $\mathbf{C} := \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}(\mathbf{m}_k) \boldsymbol{\varphi}(\mathbf{m}_k)^T$

Die Lösungen \mathbf{v} liegen in dem von den $\{\boldsymbol{\varphi}(\mathbf{m}_k), k = 1, \dots, N\}$ aufgespannten Raum.

$$\begin{aligned} \lambda \mathbf{v} &= \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}(\mathbf{m}_k) (\boldsymbol{\varphi}(\mathbf{m}_k)^T \mathbf{v}) \Rightarrow \mathbf{v} \in \text{span} \{ \boldsymbol{\varphi}(\mathbf{m}_k), k = 1, \dots, N \} \\ \Rightarrow \mathbf{v} &= \sum_{k=1}^N \alpha_k \boldsymbol{\varphi}(\mathbf{m}_k) \end{aligned} \tag{1}$$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Gleichungssystem äquivalent zur Eigenwertgleichung:

$$\lambda \boldsymbol{\varphi}(\mathbf{m}_k)^T \mathbf{v} = \boldsymbol{\varphi}(\mathbf{m}_k)^T \mathbf{C} \mathbf{v} \quad \text{mit } k = 1, \dots, N$$

Mit (1) folgt daraus:

$$\lambda \boldsymbol{\varphi}(\mathbf{m}_k)^T \sum_{i=1}^N \alpha_i \boldsymbol{\varphi}(\mathbf{m}_i) = \boldsymbol{\varphi}(\mathbf{m}_k)^T \frac{1}{N} \sum_{j=1}^N \boldsymbol{\varphi}(\mathbf{m}_j) \boldsymbol{\varphi}(\mathbf{m}_j)^T \sum_{i=1}^N \alpha_i \boldsymbol{\varphi}(\mathbf{m}_i)$$

und dann:

$$\lambda \sum_{i=1}^N \boldsymbol{\varphi}(\mathbf{m}_k)^T \boldsymbol{\varphi}(\mathbf{m}_i) \alpha_i = \frac{1}{N} \sum_{i=1}^N \left(\boldsymbol{\varphi}(\mathbf{m}_k)^T \sum_{j=1}^N \boldsymbol{\varphi}(\mathbf{m}_j) \right) \left(\boldsymbol{\varphi}(\mathbf{m}_j)^T \boldsymbol{\varphi}(\mathbf{m}_i) \right) \alpha_i \quad (2)$$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Definition der $N \times N$ Matrix \mathbf{K} , $(\mathbf{K})_{ij} = K_{ij} := \boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m}_j)$, $i, j = 1, \dots, N$

Mit (2) folgt: $N\lambda\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$ mit $\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_N)^\top$

Zu dessen Lösung löst man:

$$N\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

Eigenwertgleichung

Eigenwerte $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ und Eigenvektoren $\boldsymbol{\alpha}^1, \dots, \boldsymbol{\alpha}^N$ mit einem p ,
so dass gilt: $\lambda_1, \dots, \lambda_{p-1} = 0; \lambda_p > 0$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Normalisierung der $\mathbf{v}^1, \dots, \mathbf{v}^N \rightarrow (\mathbf{u}^k)^T \mathbf{u}^k = 1, \mathbf{u}^k \propto \mathbf{v}^k, k = 1, \dots, N$

Mit (1) folgt:

$$A_k = (\mathbf{v}^k)^T \mathbf{v}^k = \sum_{i=1}^N \sum_{j=1}^N \alpha_i^k \alpha_j^k \boldsymbol{\varphi}(\mathbf{m}_i)^T \boldsymbol{\varphi}(\mathbf{m}_j) = \sum_{i=1}^N \sum_{j=1}^N \alpha_i^k \alpha_j^k K_{ij} = (\boldsymbol{\alpha}^k)^T \mathbf{K} \boldsymbol{\alpha}^k = \lambda_k (\boldsymbol{\alpha}^k)^T \boldsymbol{\alpha}^k$$

und mit

$$\boldsymbol{\beta}^k = (\beta_1^k, \dots, \beta_N^k)^T := \frac{1}{\sqrt{A_k}} (\alpha_1^k, \dots, \alpha_N^k)^T$$

wird die Normalisierung erreicht.

$$1 = (\mathbf{u}^k)^T \mathbf{u}^k = \sum_{i=1}^N \sum_{j=1}^N \beta_i^k \beta_j^k \boldsymbol{\varphi}(\mathbf{m}_i)^T \boldsymbol{\varphi}(\mathbf{m}_j) = \sum_{i=1}^N \sum_{j=1}^N \beta_i^k \beta_j^k K_{ij} = (\boldsymbol{\beta}^k)^T \mathbf{K} \boldsymbol{\beta}^k = \lambda_k (\boldsymbol{\beta}^k)^T \boldsymbol{\beta}^k$$

Die $\boldsymbol{\beta}^k$ sind Lösungen von:

$$N \lambda_k \boldsymbol{\beta}^k = \mathbf{K} \boldsymbol{\beta}^k$$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Projektion der transformierten Merkmale $\varphi(\mathbf{m})$ auf den Eigenvektor \mathbf{u}^k .

$$(\mathbf{u}^k)^T \varphi(\mathbf{m}) = \sum_{i=1}^N \beta_i^k (\varphi(\mathbf{m}_i)^T \varphi(\mathbf{m}))$$

Kernel-Trick: Implizite Nutzung des u.U. ∞ -dimensionalen Eigenfunktionsraumes einer so genannten Kernfunktion K als transformierten Merkmalsraum. Dabei müssen die transformierten Merkmale nicht explizit berechnet werden. (Details Kapitel 7.5. Support Vektor Maschine)

Kernfunktion: $K(\mathbf{m}, \mathbf{s}) = \varphi(\mathbf{m})^T \varphi(\mathbf{s}) = \langle \varphi(\mathbf{m}), \varphi(\mathbf{s}) \rangle$

z.B.:

Polynomiale Kernfunktion: $K(\mathbf{m}, \mathbf{s}) = (\mathbf{m}^T \mathbf{s} + c)^d$

Radiale Basis Kernfunktion: $K(\mathbf{m}, \mathbf{s}) = \exp(-\|\mathbf{m} - \mathbf{s}\|^2 / (2\sigma^2))$

Sigmoid Kernfunktion: $K(\mathbf{m}, \mathbf{s}) = \tanh(\kappa(\mathbf{m}^T \mathbf{s}) + \theta)$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Vorgehensweise: Berechnung der nichtlinearen HKA.

1. Berechnen der Matrix \mathbf{K} mittels Kernfunktion.

$$K_{ij} = \boldsymbol{\varphi}(\mathbf{m}_i)^T \boldsymbol{\varphi}(\mathbf{m}_j) = K(\mathbf{m}_i, \mathbf{m}_j), \quad i, j = 1, \dots, N$$

2. Diagonalisieren von \mathbf{K} , d.h. bestimmen der Eigenvektoren $\boldsymbol{\alpha}^1, \dots, \boldsymbol{\alpha}^N$ und Eigenwerte $\lambda_1, \dots, \lambda_N$ der Eigenwertgleichung:

$$N \lambda \boldsymbol{\alpha} = \mathbf{K} \boldsymbol{\alpha}$$

3. Normalisieren der Eigenvektoren $\mathbf{v}^1, \dots, \mathbf{v}^N$.

$$(\mathbf{u}^k)^T \mathbf{u}^k = 1 \quad \Rightarrow \quad \lambda_k (\boldsymbol{\beta}^k)^T \boldsymbol{\beta}^k = 1$$

4. Projektion der Merkmale auf die Eigenvektoren \mathbf{u}^k .

$$m'_k = \sum_{i=1}^N \beta_i^k K(\mathbf{m}_i, \mathbf{m})$$

$$m'_k = (\mathbf{u}^k)^T \boldsymbol{\varphi}(\mathbf{m}) = \sum_{i=1}^N \beta_i^k K(\mathbf{m}_i, \mathbf{m})$$

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Vorgehensweise für $\frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}(\mathbf{m}_k) \neq \mathbf{0}$

Bemerkung: Da $\boldsymbol{\varphi}$ nicht explizit berechnet wird, können die transformierten Merkmalsvektoren nicht direkt zentriert werden. Die obige Herleitung muss dann für

$$\boldsymbol{\psi}(\mathbf{m}_i) := \boldsymbol{\varphi}(\mathbf{m}_i) - \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}(\mathbf{m}_k)$$

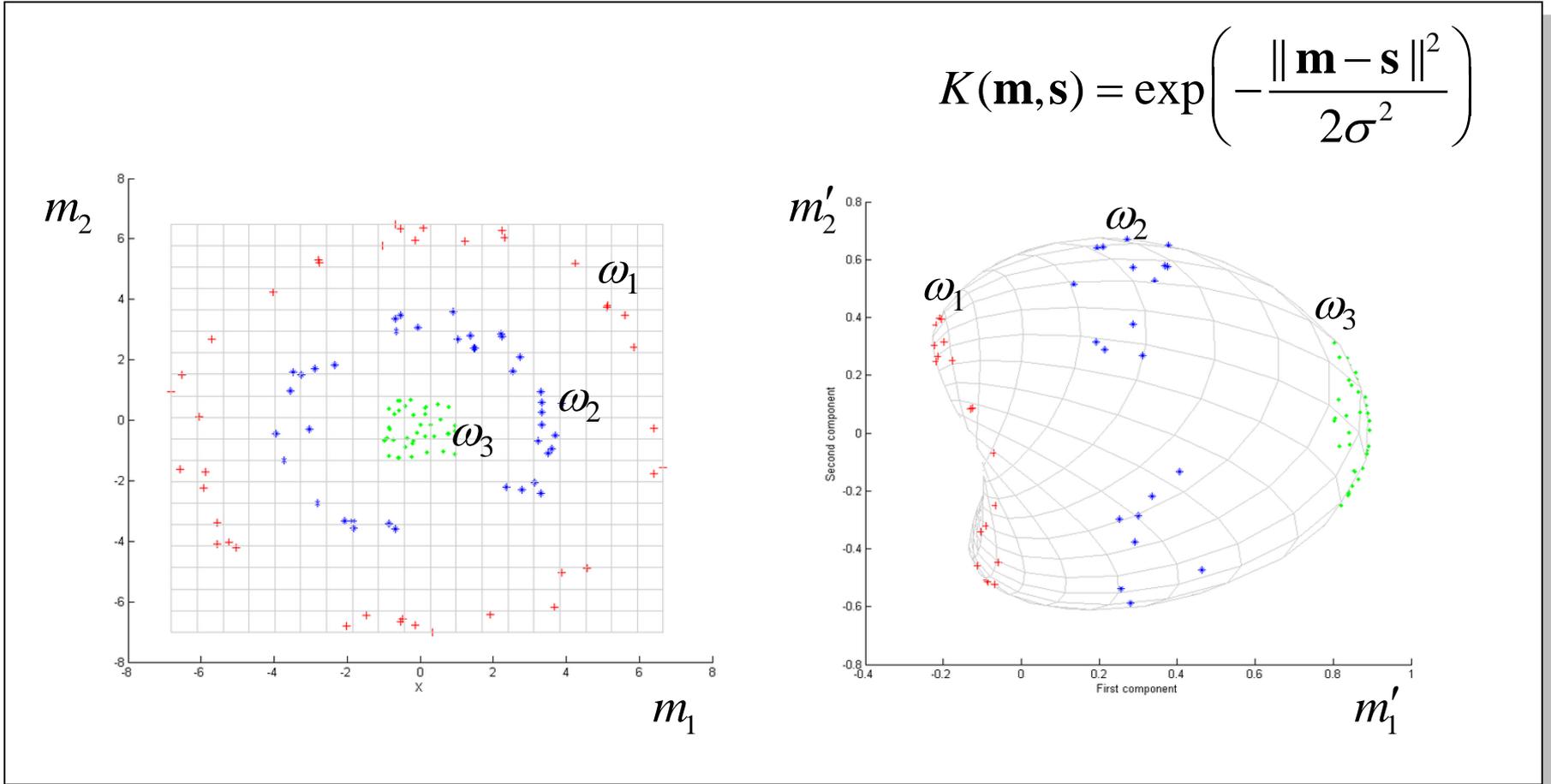
durchgerechnet werden. → Resultat: Der Algorithmus zur Berechnung der nichtlinearen HKA wird mit der Matrix $\boldsymbol{\Gamma}$ anstatt \mathbf{K} durchgeführt.

$$\boldsymbol{\Gamma} := \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K} \mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N \quad \text{mit} \quad (\mathbf{1}_N)_{ij} := \frac{1}{N}, \quad i, j = 1, \dots, N$$

Quelle: B. Schölkopf, A.J. Smola, K.-R. Müller: *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*. Technical Report 44, Max-Planck-Institut für biologische Kybernetik, 1996.

2.7.2. Nichtlineare Hauptkomponentenanalyse (Kernelized PCA)

Beispiel:



http://en.wikipedia.org/wiki/Kernel_principal_component_analysis

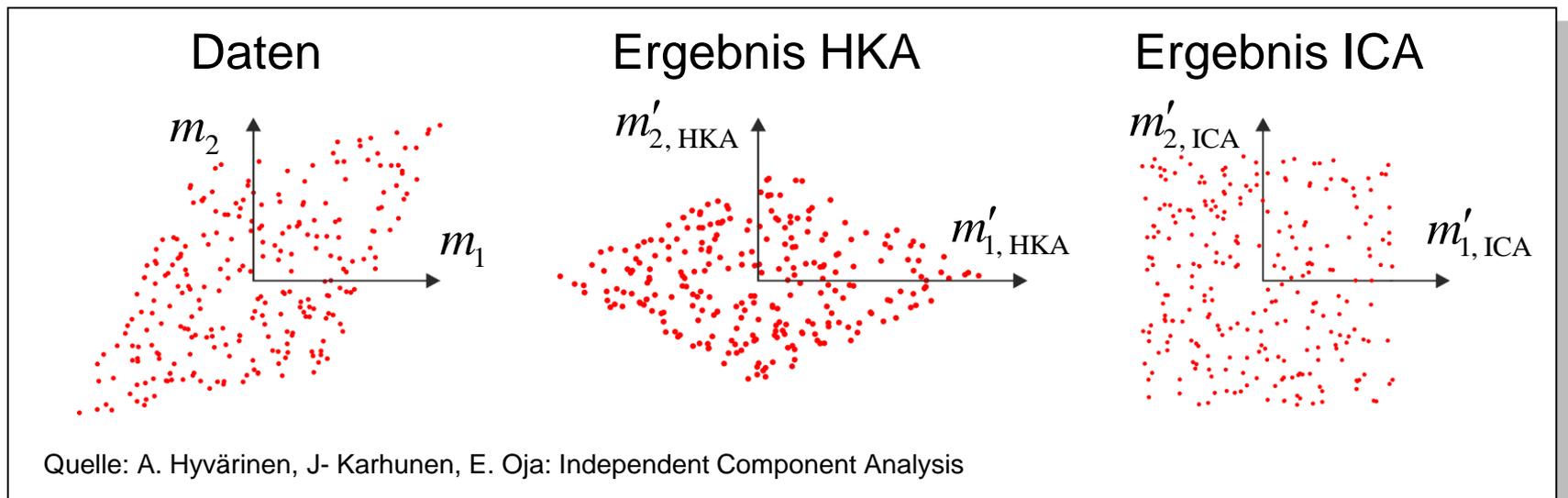
Erste Komponente nach der Transformation reicht zur Unterscheidung der Klassen aus.

2.7.3. Independent Component Analysis (ICA)

Ziel: Finden eines linear transformierten Merkmalvektors $\mathbf{m}' = \mathbf{U}(\mathbf{m} - \bar{\mathbf{m}})$ dessen Komponenten $p(\mathbf{m}') = p(m'_1, m'_2, \dots, m'_d) = p(m'_1) \cdot p(m'_2) \cdot \dots \cdot p(m'_d)$ **stochastisch unabhängig** sind.

Aufgabenstellung: Bestimmen der Koeffizienten $\mathbf{U} = \begin{pmatrix} u_{11} & \mathbf{K} & u_{1d} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ u_{d1} & \mathbf{L} & u_{dd} \end{pmatrix} := \mathbf{YZ}$ der Transformation $\mathbf{m}' = \mathbf{U}(\mathbf{m} - \bar{\mathbf{m}})$

Vergleich: Transformierter Merkmalvektor der Hauptkomponentenanalyse liefert **nur paarweise unkorrelierte** Komponenten.



2.7.3. Independent Component Analysis (ICA)

Exkurs Korrelationskoeffizient

Seien a und b Zufallsvariablen mit: $a, b : p(a, b)$

Definition **Korrelationskoeffizient**:
$$\rho_{ab} := \frac{\mathbb{E}\{(a - \mathbb{E}\{a\})(b - \mathbb{E}\{b\})\}}{\sqrt{\text{Var}\{a\} \text{Var}\{b\}}} \in [-1, 1]$$

Der Korrelationskoeffizient ρ_{ab} quantifiziert die **lineare stochastische Abhängigkeit** von a und b .

Definition: Zwei Zufallsvariablen heißen **stochastisch unabhängig**, falls gilt:

$$p(a, b) = p(a)p(b) \Leftrightarrow p(a | b) = p(a) \Leftrightarrow p(b | a) = p(b)$$

Satz: Stochastische Unabhängigkeit \Rightarrow Unkorreliertheit.

Die Umkehrung gilt nur bei normalverteilten Zufallsvariablen.

Bsp.: Sei $a : N(a; \mathbb{E}\{a\}, \text{Var}\{a\}) = N(a; 0, 1)$ und $b := a^2$

a und b sind stochastisch abhängig wegen: $p(b | a) = \delta(b - a^2) \neq p(b)$

a und b sind jedoch unkorreliert:
$$\rho_{ab} = \mathbb{E}\{ab\} = \mathbb{E}\{a^3\} = 0$$

2.7.3. Independent Component Analysis (ICA)

Vorgehensweise ICA:

(1) Vorbereiten der Daten: Zentrieren und *Whitening*

- **Zentrieren** der Daten $D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$ bzgl. **empirischem Mittelwert**:

$$\mathbf{m} \rightarrow \mathbf{m} - \bar{\mathbf{m}} \quad \text{mit} \quad \bar{\mathbf{m}} := \frac{1}{N} \sum_{k=1}^N \mathbf{m}_k$$

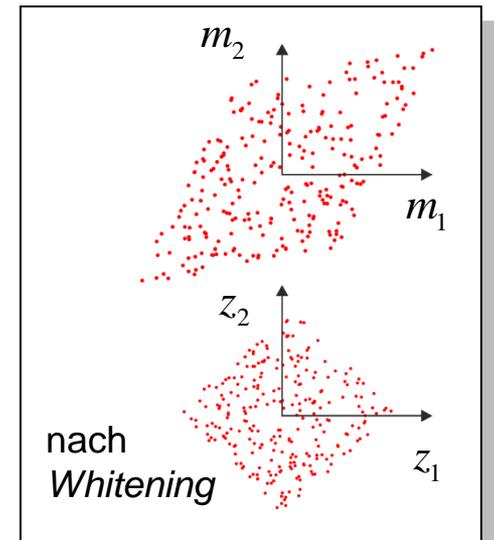
- **Whitening** der Daten, d.h. Transformation $\mathbf{z} := \mathbf{Z}(\mathbf{m} - \bar{\mathbf{m}})$, so dass gilt: **paarweise Unkorreliertheit** und $\text{Var}\{z_i\} = 1$

$$\text{Cov}\{\mathbf{z}\} = \mathbf{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$$

Z.B. mittels HKA und Normierung:

[Wdh. HKA: $\mathbf{m}' = \mathbf{A}^T(\mathbf{m} - \bar{\mathbf{m}})$]

$$\mathbf{Z} := \begin{bmatrix} \frac{1}{\sqrt{\hat{\kappa}_1}} & \mathbf{L} & \mathbf{0} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{0} & \mathbf{L} & \frac{1}{\sqrt{\hat{\kappa}_d}} \end{bmatrix} \mathbf{A}^T = \begin{bmatrix} \sqrt{\frac{N-1}{\lambda_1}} & \mathbf{L} & \mathbf{0} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{0} & \mathbf{L} & \sqrt{\frac{N-1}{\lambda_d}} \end{bmatrix} \mathbf{A}^T$$



Zur Definition \mathbf{A} , κ , λ siehe Abschn. 2.7.1. Beachte hier: $d' = d$

2.7.3. Independent Component Analysis (ICA)

Vorteile: *Whitening* vermindert Freiheitsgrade der Matrix \mathbf{Y} .

Anm.: Die *Independent Components* sind durch das *whitening* bis auf eine orthogonale Transformation (Drehung, Spiegelung) bestimmt.

Beweis: Die Varianzen der *Independent Components* sind **mehrdeutig**.

$$\mathbf{m}' = \mathbf{U}(\mathbf{m} - \bar{\mathbf{m}}) = \sum_{i=1}^d \mathbf{u}_i (m_i - \bar{m}_i), \text{ mit } \mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_d) \in \mathbb{R}^{d \times d}$$

$$\mathbf{m}' = \sum_{i=1}^d \begin{pmatrix} 1 \\ \alpha_i \end{pmatrix} \mathbf{u}_i (\alpha_i (m_i - \bar{m}_i))$$

D.h. Forderung $\text{Cov}\{\mathbf{m}'\} = \mathbf{I}$ ist ohne Einschränkung zulässig.

Unkorreliertheit ist notwendige Bedingung für **stochastische Unabhängigkeit**.

Forderung: *Whitening* soll bei Transformation $\mathbf{m}' = \mathbf{U}(\mathbf{m} - \bar{\mathbf{m}}) = \mathbf{Y}\mathbf{Z}(\mathbf{m} - \bar{\mathbf{m}})$ erhalten bleiben.

$$\mathbf{E}\{\mathbf{m}'\mathbf{m}'^T\} = \mathbf{Y}\mathbf{E}\{(\mathbf{Z}(\mathbf{m} - \bar{\mathbf{m}})(\mathbf{Z}(\mathbf{m} - \bar{\mathbf{m}}))^T)\}\mathbf{Y}^T = \mathbf{Y}\mathbf{Y}^T = \mathbf{I}$$

Vorteil: Eine orthogonale Matrix $\mathbf{Y} \in \mathbb{R}^{d \times d}$ besitzt nur $d(d-1)/2$ Freiheitsgrade (unabhängige Werte, Einträge).

2.7.3. Independent Component Analysis (ICA)

Exkurs Transinformation (Mutual Information)

Seien a und b Zufallsvariablen (ZV) mit: $a, b : p(a, b)$

Es gilt: $p(a, b) = p(a | b)p(b) = p(b | a)p(a)$

Definition: Transinformation $I(a; b)$ ist die Information im Shannon'schen Sinne einer ZV b bezüglich einer ZV a :

$$I(a; b) := h(a) - h(a | b) = h(b) - h(b | a) \quad h(.) \text{ ist die differentielle Entropie.}$$

$$I(a; b) = \underbrace{h(a)}_{\substack{\text{Unsicherheit} \\ \text{des Wertes } a}} - \underbrace{h(a | b)}_{\substack{\text{Unsicherheit} \\ \text{des Wertes } a \\ \text{gegeben } b}}$$

Die Transinformation $I(a; b)$ ist die mittlere Verminderung der Unsicherheit der ZV a nach Beobachtung des Wertes der ZV b .

$$= - \int_{\text{supp } p(a)} p(a) \log(p(a)) da + \iint_{\text{supp}\{p(a, b)\}} p(a, b) \log(p(a | b)) da db$$

$$= - \int_{\text{supp } p(a, b)} p(a, b) \log(p(a)) da db + \iint_{\text{supp}\{p(a, b)\}} p(a, b) \log(p(a | b)) da db$$

2.7.3. Independent Component Analysis (ICA)

Exkurs Transinformation (Mutual Information)

$$\begin{aligned} I(a;b) &= \iint_{\text{supp}\{p(a,b)\}} p(a,b) \log\left(\frac{p(a|b)}{p(a)}\right) da db \\ &= \iint_{\text{supp}\{p(a,b)\}} p(a,b) \log\left(\frac{p(a,b)}{p(a)p(b)}\right) da db = D(p(a,b) \| p(a)p(b)) \end{aligned}$$

$$I(a;b) = D(p(a,b) \| p(a)p(b))$$

Marginalverteilungsdichten: $p(a) = \int p(a,b) db$ und $p(b) = \int p(a,b) da$

Die Transinformation $I(a;b)$ ist gleich dem Kullback-Leibler-Abstand der Verbundverteilungsdichte vom Produkt der Marginalverteilungsdichten. Dieser Abstand nimmt für den Fall stochastischer Unabhängigkeit seinen Minimalwert 0 an.

2.7.3. Independent Component Analysis (ICA)

Vorgehensweise ICA:

(2) Optimierung der stochastischen Unabhängigkeit mittels *Transinformation*.

Transinformation: Spezialfall der *Kullback-Leibler Distanz*.

$$J(\mathbf{m}) = D(p(\mathbf{m}) \parallel p(m_1)p(m_2)\dots p(m_d)) = \int p(\mathbf{m}) \log \left(\frac{p(\mathbf{m})}{p(m_1)p(m_2)\dots p(m_d)} \right) d\mathbf{m}$$

Wdh.: $D(p \parallel p') \geq 0$ und

$D(p(m_1, m_2) \parallel p(m_1) \cdot p(m_2)) = 0$, wenn gilt: $p(m_1, m_2) = p(m_1) \cdot p(m_2)$.

Optimierungsfunktional: (numerische Optimierung)

$$J^*(\mathbf{m}') = \min_{\mathbf{Y}} \{D(p(\mathbf{m}') \parallel p(m_1')p(m_2')\dots p(m_d'))\} \rightarrow \text{optimales } \mathbf{Y}$$

mit $\mathbf{m}' = \mathbf{YZ}(\mathbf{m} - \bar{\mathbf{m}})$, $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_d)$ und Nebenbedingung: $\mathbf{YY}^T = \mathbf{I}$.

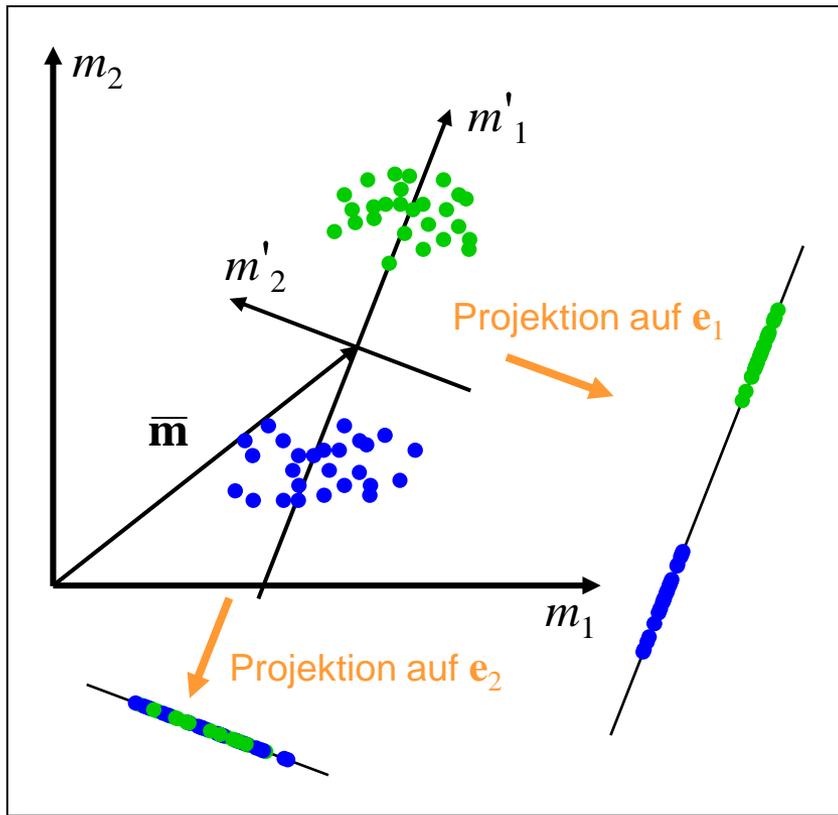
Anm.: $J^*(\mathbf{m}') = 0$ kann nicht generell garantiert werden.

2.7.4. Multiple Discriminant Analysis

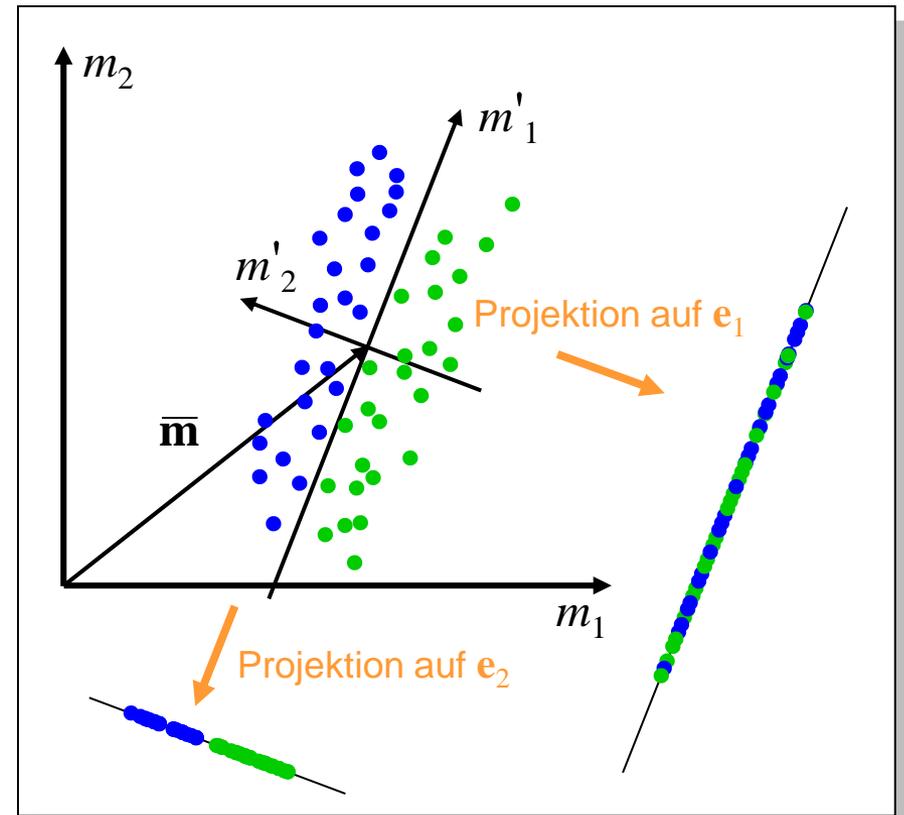
Nachteile der HKA (PCA) und ICA :

Sie zielen nicht auf optimale Klassentrennbarkeit.

HKA z.B. minimiert die Summe der quadratischen Fehler im Hinblick auf Beschreibung der Gesamtstichprobe D . Beispiele:



Trennung mittels m'_1 gut.



Keine Trennung mittels m'_1 .

2.7.4. Multiple Discriminant Analysis

Ziel: Merkmale für hohe Klassenunterscheidbarkeit

Einfachster Fall:

Datensatz: $D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$

2 Klassen: $c = 2$

$$D_1 := \{\mathbf{m}_i \mid \omega = \omega_1\}, \quad D_2 := \{\mathbf{m}_i \mid \omega = \omega_2\}$$

$$|D_1| = N_1, \quad |D_2| = N_2, \quad N_1 + N_2 = N$$

Projektion auf 1-dimensionalen Unterraum: $\text{span}(\mathbf{w})$

$$m' = \mathbf{w}^T \mathbf{m} \quad \Rightarrow \quad m'_1, \dots, m'_N \text{ in } D'_1, D'_2 \text{ unterteilt}$$

Gesucht: \mathbf{w} , sodass die 2 Klassen optimal getrennt werden.

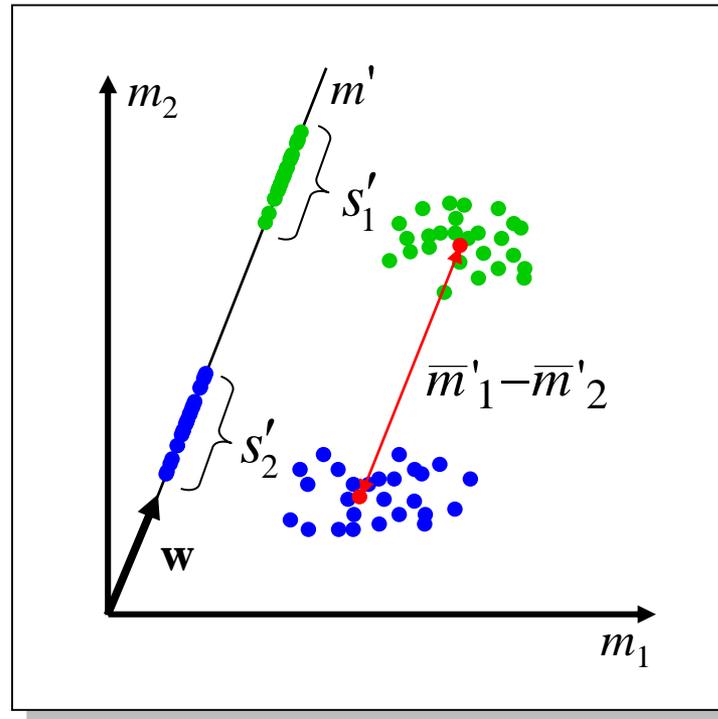
2.7.4. Multiple Discriminant Analysis

Fisher Linear Discriminant:

$$J(\mathbf{w}) = \frac{|\bar{m}'_1 - \bar{m}'_2|^2}{s_1'^2 + s_2'^2}$$

$$\bar{m}'_i = \frac{1}{N_i} \sum_{m' \in D'_i} m' = \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} \mathbf{w}^T \mathbf{m} = \mathbf{w}^T \bar{\mathbf{m}}_i \quad s_i'^2 := \frac{1}{N_i} \sum_{m' \in D'_i} (m' - \bar{m}'_i)^2 \quad i = 1, 2$$

Veranschaulichung:



2.7.4. Multiple Discriminant Analysis

Streumatrizen: $\mathbf{S}_i := \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} (\mathbf{m} - \bar{\mathbf{m}}_i)(\mathbf{m} - \bar{\mathbf{m}}_i)^T$, $\mathbf{S}_w := \mathbf{S}_1 + \mathbf{S}_2$

$$s_i'^2 = \frac{1}{N_i} \sum_{m' \in D'_i} (m' - \bar{m}'_i)^2 = \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} (\mathbf{w}^T \mathbf{m} - \mathbf{w}^T \bar{\mathbf{m}}_i)^2$$

$$= \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} \mathbf{w}^T (\mathbf{m} - \bar{\mathbf{m}}_i)(\mathbf{m} - \bar{\mathbf{m}}_i)^T \mathbf{w} = \mathbf{w}^T \mathbf{S}_i \mathbf{w}$$

$$s_1'^2 + s_2'^2 = \mathbf{w}^T \mathbf{S}_w \mathbf{w}$$

$$(\bar{m}'_1 - \bar{m}'_2)^2 = (\mathbf{w}^T \bar{\mathbf{m}}_1 - \mathbf{w}^T \bar{\mathbf{m}}_2)^2 = \mathbf{w}^T \underbrace{\begin{pmatrix} \bar{\mathbf{m}}_1 & -\bar{\mathbf{m}}_2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{\mathbf{m}}_1 & \bar{\mathbf{m}}_2 \\ 1 & 1 \end{pmatrix}^T}_{=: \mathbf{S}_B} \mathbf{w} = \mathbf{w}^T \mathbf{S}_B \mathbf{w}$$

$$\Rightarrow \boxed{J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}} \quad : \text{Rayleigh Koeffizient}$$

Bemerkung zu den Indizes: *B*: *Between*, *w*: *Within*

2.7.4. Multiple Discriminant Analysis

Gesucht: \mathbf{w} , so dass $J(\mathbf{w})$ maximal wird.

$$\text{Maximierung von } \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \Leftrightarrow \text{Maximierung von } \mathbf{w}^T \mathbf{S}_B \mathbf{w} \Big|_{\mathbf{w}^T \mathbf{S}_w \mathbf{w} = 1}$$

$$f(\mathbf{w}, \lambda) = \mathbf{w}^T \mathbf{S}_B \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{S}_w \mathbf{w} - 1) \quad \lambda: \text{Lagrange Multiplikator}$$

$$\nabla_{\mathbf{w}} f = 2\mathbf{S}_B \mathbf{w} - 2\lambda \mathbf{S}_w \mathbf{w} = \mathbf{0}$$

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w} \quad \text{verallgemeinertes Eigenwertproblem}$$

Falls \mathbf{S}_w invertierbar:

$$\mathbf{S}_w^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w}$$

$\mathbf{S}_B \mathbf{w}$ hat immer die gleiche Richtung wie $\bar{\mathbf{m}}_1 - \bar{\mathbf{m}}_2$

$$\mathbf{w} \propto \mathbf{S}_w^{-1} (\bar{\mathbf{m}}_1 - \bar{\mathbf{m}}_2)$$

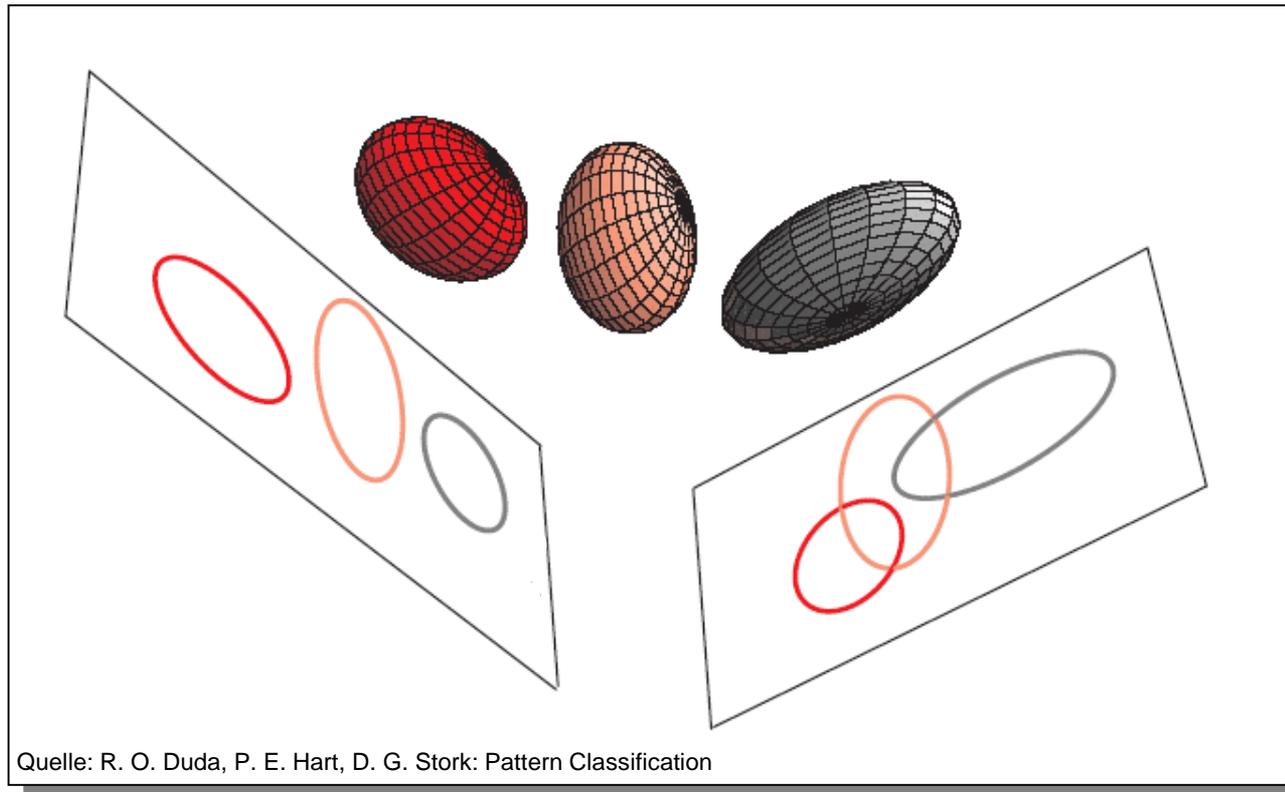
2.7.4. Multiple Discriminant Analysis

Allgemeiner Fall:

Zahl der Klassen: c

⇒ Die Projektion findet auf einen $c-1$ - dimensionalen Unterraum statt.

Bsp.: Projektion von 3 Clustern auf 2-dimensionale Unterräume



2.7.2. Multiple Discriminant Analysis

Allgemeiner Fall: Zahl der Klassen: c

Streumatrizen: $\mathbf{S}_w := \sum_{i=1}^c \mathbf{S}_i$

Intraklassenstreuung: $\mathbf{S}_i := \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} (\mathbf{m} - \bar{\mathbf{m}}_i)(\mathbf{m} - \bar{\mathbf{m}}_i)^T$

Mittelwertvektor der Klasse i : $\bar{\mathbf{m}}_i := \frac{1}{N_i} \sum_{\mathbf{m} \in D_i} \mathbf{m}$

Mittelwertvektor: $\bar{\mathbf{m}} := \frac{1}{N} \sum_{\mathbf{m}} \mathbf{m} = \frac{1}{N} \sum_{i=1}^c N_i \bar{\mathbf{m}}_i$

Interklassenstreuung \triangleq Streuung der Klassenzentren :

$$\mathbf{S}_B := \sum_{i=1}^c N_i (\bar{\mathbf{m}}_i - \bar{\mathbf{m}})(\bar{\mathbf{m}}_i - \bar{\mathbf{m}})^T$$

(Achtung: etwas andere Def. als oben für $c = 2$)

2.7.2. Multiple Discriminant Analysis

$$\mathbf{m}'_j = \mathbf{w}_j^T \mathbf{m}, \quad j = 1, \dots, c-1, \quad \mathbf{m}' = \mathbf{W}^T \mathbf{m} \quad \mathbf{W} := (\mathbf{w}_1, \dots, \mathbf{w}_{c-1})$$

$$\bar{\mathbf{m}}'_i = \frac{1}{N_i} \sum_{\mathbf{m}' \in D'_i} \mathbf{m}', \quad \bar{\mathbf{m}}' = \frac{1}{N} \sum_{i=1}^c N_i \bar{\mathbf{m}}'_i$$

$$\mathbf{S}'_w = \sum_{i=1}^c \frac{1}{N_i} \sum_{\mathbf{m}' \in D'_i} (\mathbf{m}' - \bar{\mathbf{m}}'_i)(\mathbf{m}' - \bar{\mathbf{m}}'_i)^T$$

$$\mathbf{S}'_B = \sum_{i=1}^c N_i (\bar{\mathbf{m}}'_i - \bar{\mathbf{m}}')(\bar{\mathbf{m}}'_i - \bar{\mathbf{m}}')^T$$

$$\Rightarrow \begin{cases} \mathbf{S}'_w = \mathbf{W}^T \mathbf{S}_w \mathbf{W} \\ \mathbf{S}'_B = \mathbf{W}^T \mathbf{S}_B \mathbf{W} \end{cases}$$

$$J(\mathbf{W}) := \frac{|\mathbf{S}'_B|}{|\mathbf{S}'_w|} = \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}$$

Rayleigh Koeffizient

2.7.2. Multiple Discriminant Analysis

$$J(\mathbf{W}) = \frac{|\mathbf{S}'_B|}{|\mathbf{S}'_w|} = \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}$$

Gesucht: \mathbf{w}_j so, dass $J(\mathbf{W})$ maximal wird.

$$\Leftrightarrow \mathbf{S}_B \mathbf{w}_j = \lambda_j \mathbf{S}_w \mathbf{w}_j$$

$$\Leftrightarrow (\mathbf{S}_B - \lambda_j \mathbf{S}_w) \mathbf{w}_j = 0$$

\mathbf{W} hat als Spaltenvektoren die zu den größten Eigenwerten gehörenden verallgemeinerten Eigenvektoren.

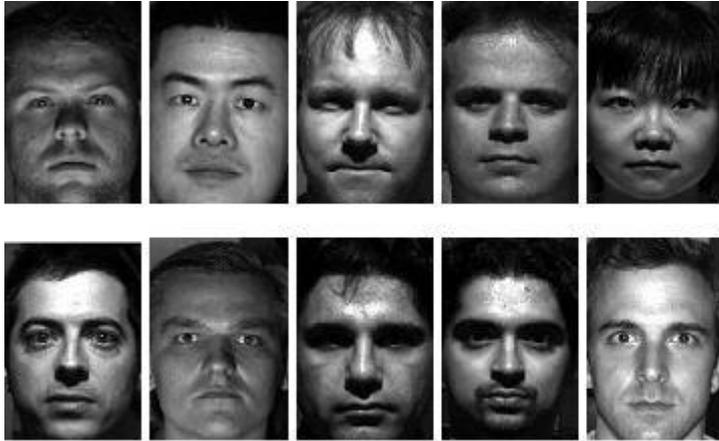
Vorgehensweise:

- Nullstellen des charakteristischen Polynoms $|\mathbf{S}_B - \lambda_j \mathbf{S}_w| = 0$ bestimmen.
- Gleichungen $(\mathbf{S}_B - \lambda_j \mathbf{S}_w) \mathbf{w}_j = 0$ lösen.

2.7.4. Multiple Discriminant Analysis

Bsp.: Fisherfaces

„Yale“-Datenbasis:



Variation der Beleuchtung



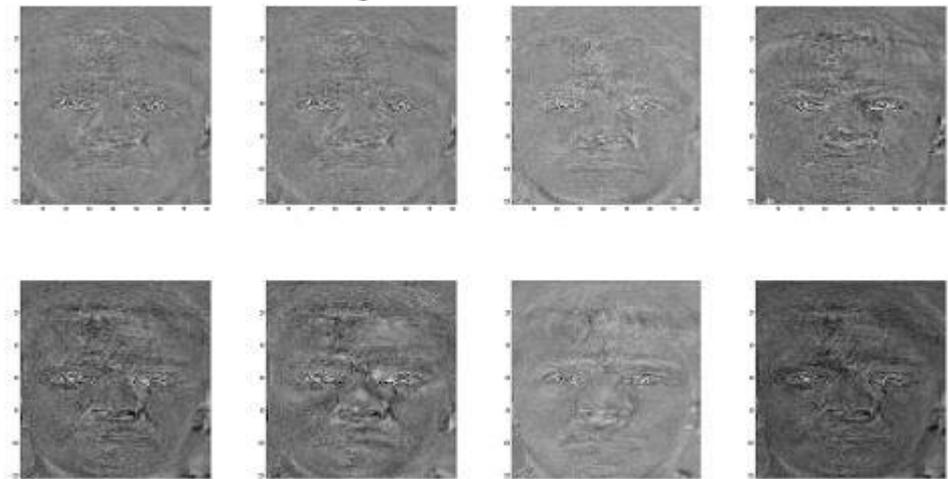
<http://www.cs.mcgill.ca/~arajwa/766/results/>

Erste 10 Eigengesichter



Erzielte Erkennungsrate bei disjunkter Lern- und Teststichprobe: ca. 68%

Erste 8 Fishergesichter



Erzielte Erkennungsrate ca. 90%

2.7.5. Dimensionsreduktion durch Merkmalsauswahl

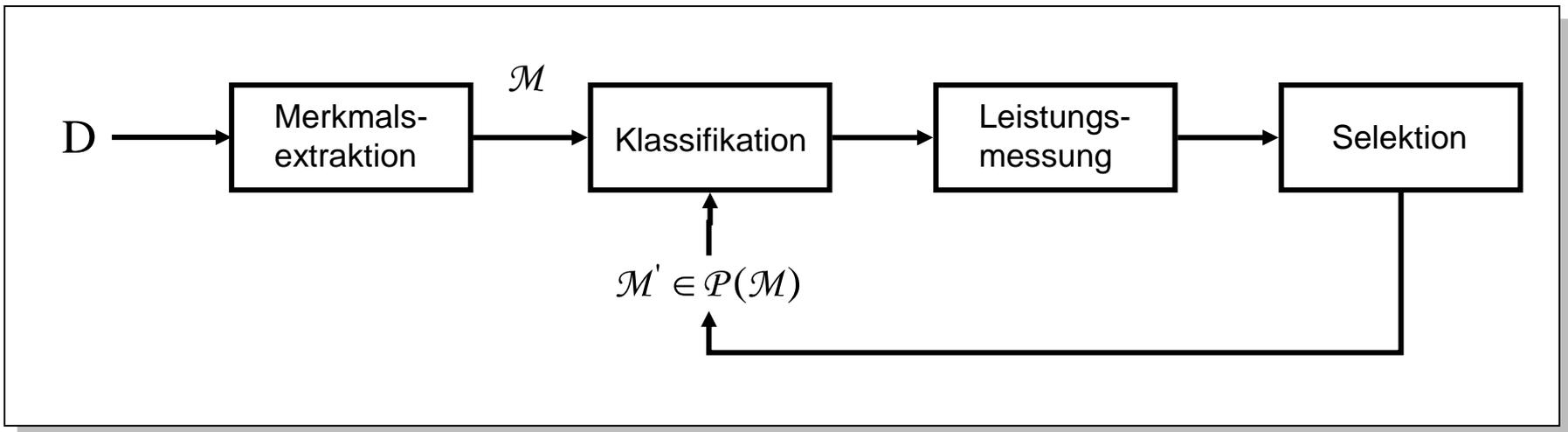
Ausgangslage: Menge $\mathcal{M} = \{m_1, \dots, m_d\}$ von für ein konkretes Problem in Betracht gezogenen Merkmalen.

D: **endliche Stichprobe** von Mustern bekannter Klassenzugehörigkeit

Problem: d zu hoch in Relation zu $|D|$

$\mathcal{P}(\mathcal{M})$: **Potenzmenge** von \mathcal{M} , $|\mathcal{P}(\mathcal{M})| = 2^d$

Gesucht: $\mathcal{M}^* \in \mathcal{P}(\mathcal{M})$ mit bester Klassifikationsleistung



2.7.5. Dimensionsreduktion durch Merkmalsauswahl

Bei großen d ist *Brute Force* Auswahl (vollständige Suche) aussichtslos.

Bemerkung: Auswahl und maximale Klassifikationsleistung hängt vom ausgewählten Klassifikator ab.

Suboptimales iteratives Verfahren:

- Aus \mathcal{M} wird das Merkmal mit bester Klassifikationsleistung ausgewählt.
- Zweites Merkmal aus \mathcal{M} hinzunehmen, sodass Klassifikationsleistung maximal wird.
- Drittesusw.
-
- Abbruch, sobald Klassifikationsleistung ausreichend oder keine Verbesserung mehr erzielt wird.

Bsp: $d = 50 \rightarrow$ maximal 1275 untersuchte Teilmengen
 $2^d \approx 10^{15}$ zu untersuchende Teilmengen

Suboptimales Verfahren liefert i.d.R. brauchbare Ergebnisse.