

Computergrafik 2: Klassifikation

Prof. Dr. Michael Rohs, Dipl.-Inform. Sven Kratz

michael.rohs@ifi.lmu.de

MHCI Lab, LMU München

Folien teilweise von Andreas Butz, sowie von Klaus D. Tönnies
(Grundlagen der Bildverarbeitung. Pearson Studium, 2005.)

Vorlesungen

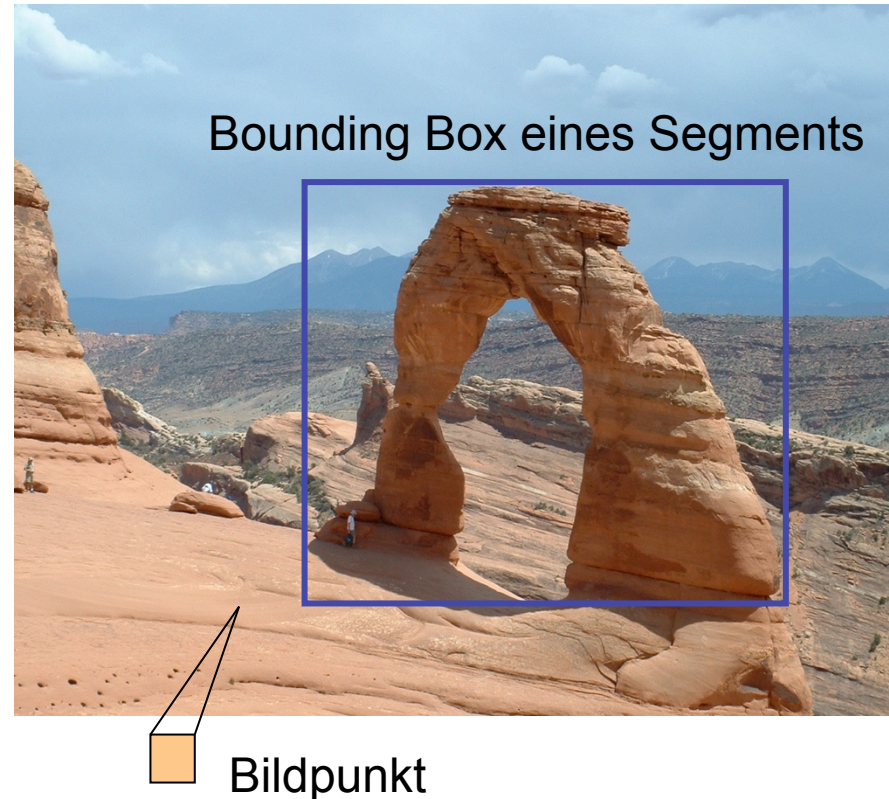
Datum	Thema
24.4.	Einführung, Organisatorisches (Übungen, Klausur)
1.5./8.5.	keine Vorlesungen (wegen 1. Mai und CHI-Konferenz)
15.5.	Abtastung von Bildern, Punktbasierte Verfahren der Bildverbesserung
22.5.	Licht, Farbe, Farbmanagement
30.5.	Konvolution, Filterung im Ortsraum (Verschiebung wegen Pfingstdienstag)
5.6.	Fouriertransformation: Grundlagen
12.6.	Filterung im Frequenzraum
19.6.	Kanten, Linien, Ecken
27.6.	Segmentierung
3.7.	Segmentierung, Morphologische Operationen
10.7.	Klassifikation
17.7.	Image Matching
24.7.	Klausur (Hörsaal M 018 im Hauptgebäude, 14-16 Uhr)

Themen heute

- Merkmale
- Klassifikation von Segmenten
- Texturmerkmale
- Formmerkmale
- Topologische Formmerkmale
- Bayes-sche Klassifikation
- kNN-Klassifikation
- Logistische Regression

Merkmale und Klassifikation

- Klassifikation: Zuordnung von Bedeutung zu
 - Segmenten
 - Segmentkombinationen
 - Pixeln
- Kriterium sind Merkmale, für die oft angenommen wird
 - durch Skalar repräsentierbar
 - voneinander unabhängig
- Merkmale in einem **Merkmalsvektor** zusammengefasst



Einfache Pixel-Merkmale

Grauwert $\mathbf{m}(p) = [f(p)]$



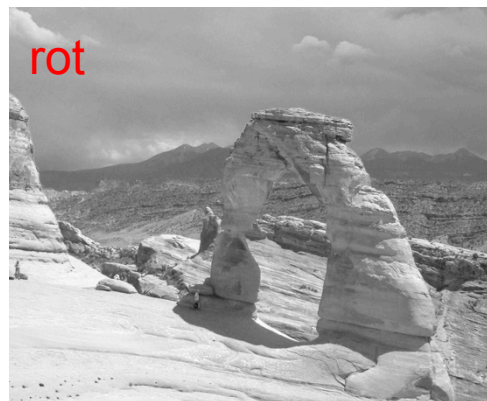
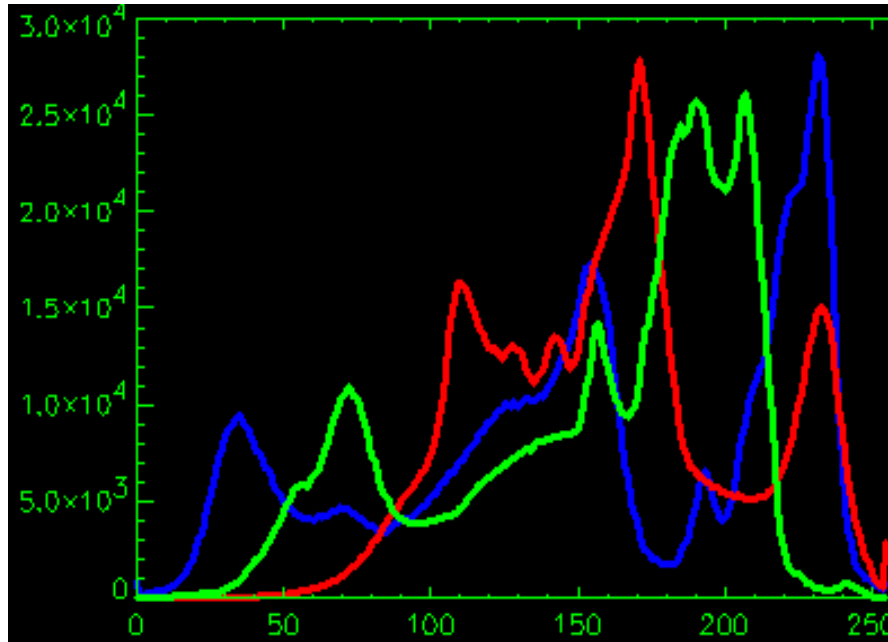
Farbwerte $\mathbf{m}(p) = [r(p) \ g(p) \ b(p)]$



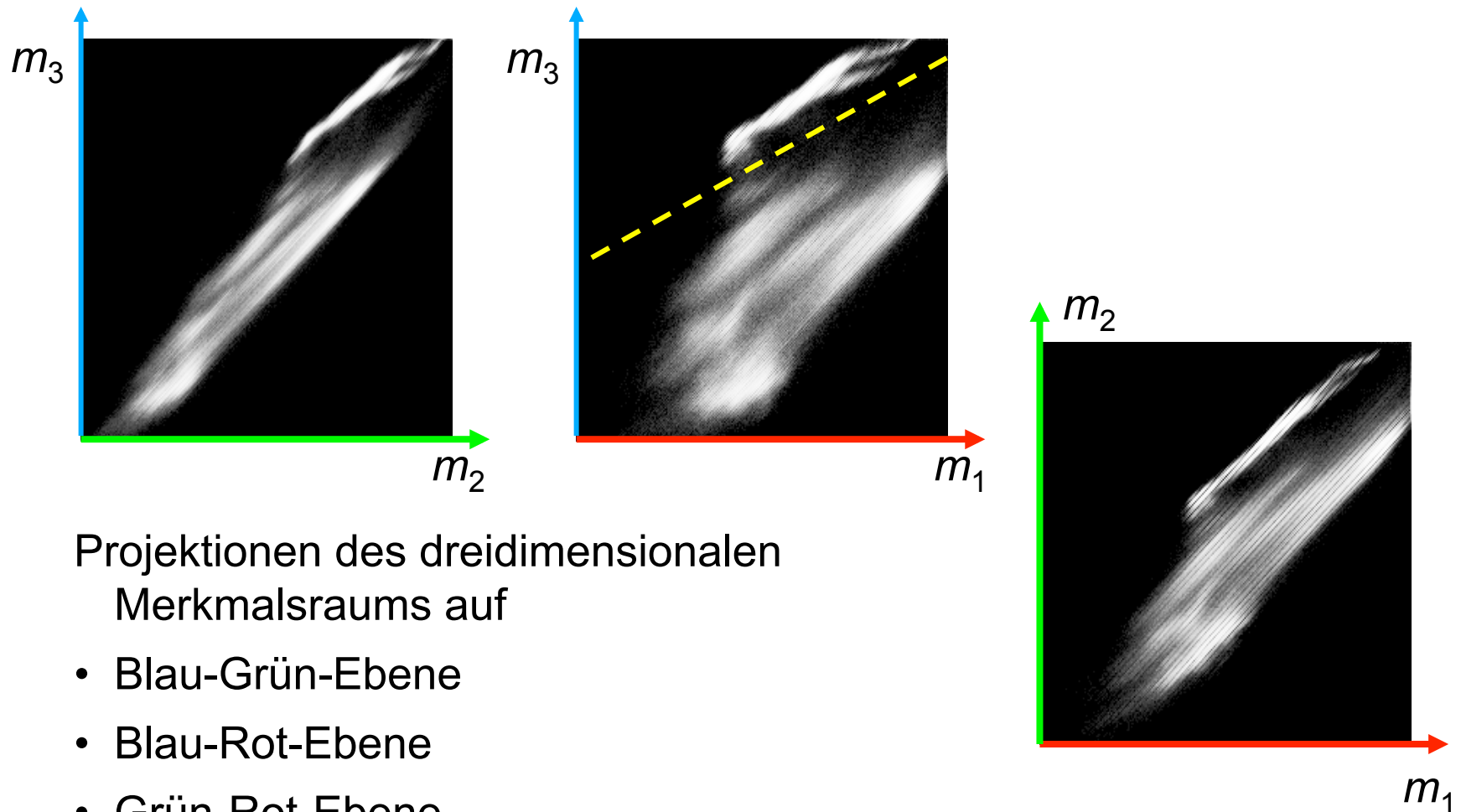
3D-Merkmalraum der Farbebenen



Histogramme



Mehrdimensionaler Merkmalsraum

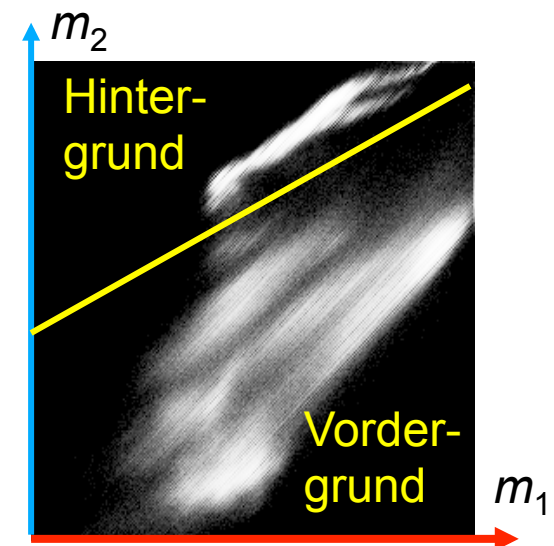


Projektionen des dreidimensionalen Merkmalsraums auf

- Blau-Grün-Ebene
- Blau-Rot-Ebene
- Grün-Rot-Ebene

Merkmalsabhängige Klassifikation

- Klassifikation:
 - Entscheidungsgrenze (decision boundary) im Merkmalsraum finden
 - Allen Merkmalsträgern (hier: Pixeln) eine Bedeutung zuordnen
- Beispiel:
 - Entscheidungsgrenze ist eine Gerade $a \cdot m_1 + b \cdot m_2 + c = 0$
 - Pixelklassifikation: $p(i,j) = \text{Vordergrund}$, falls $a \cdot m_1(p) + b \cdot m_2(p) + c < 0$, sonst Hintergrund



Klassifikations- ergebnis

$$a \cdot m_1(p) + b \cdot m_2(p) + c \geq 0$$



$$a \cdot m_1(p) + b \cdot m_2(p) + c < 0$$





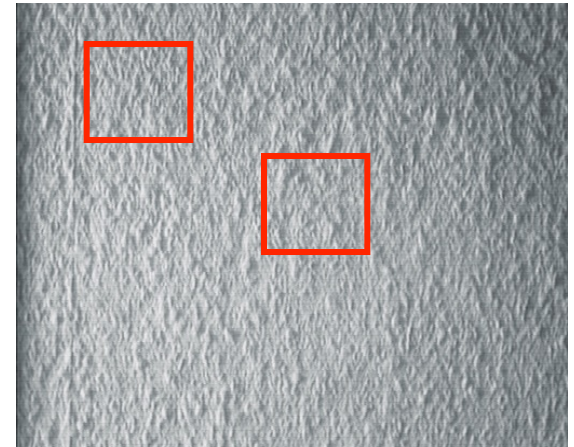
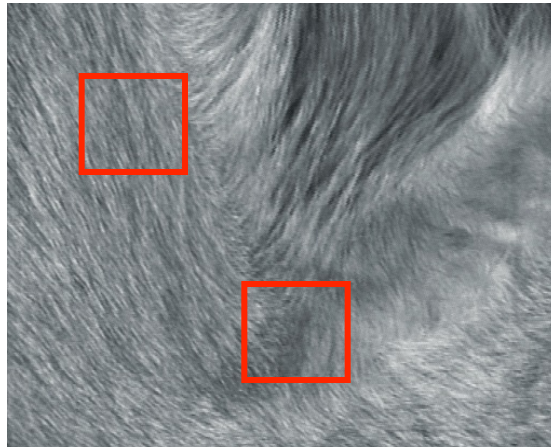
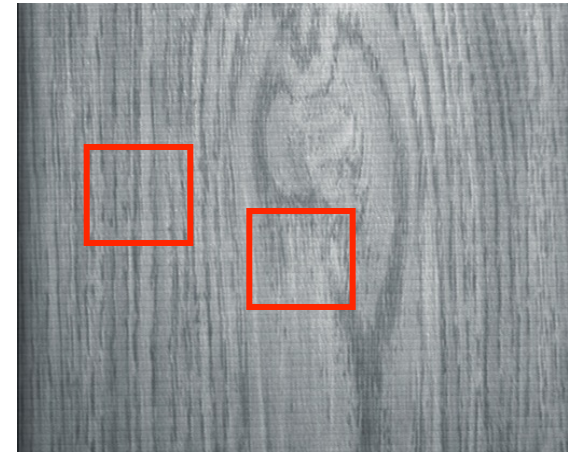
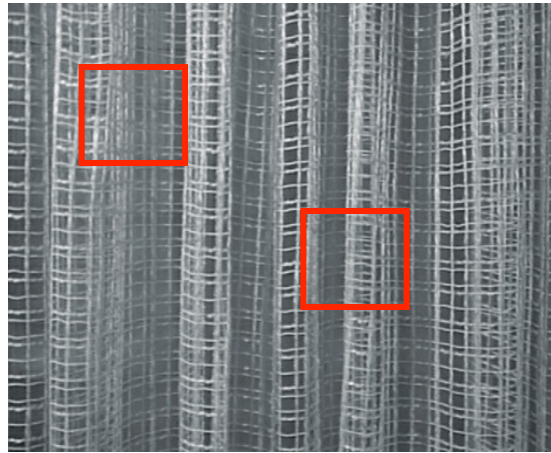
Texturmerkmale

Objekte zeichnen sich nicht nur durch Helligkeit und Farbe aus.

Textur: ein sich in einer bestimmten Region stochastisch oder deterministisch wiederholendes Muster (Invariante).

Texturmerkmal: messbare Charakterisierung der Invariante.

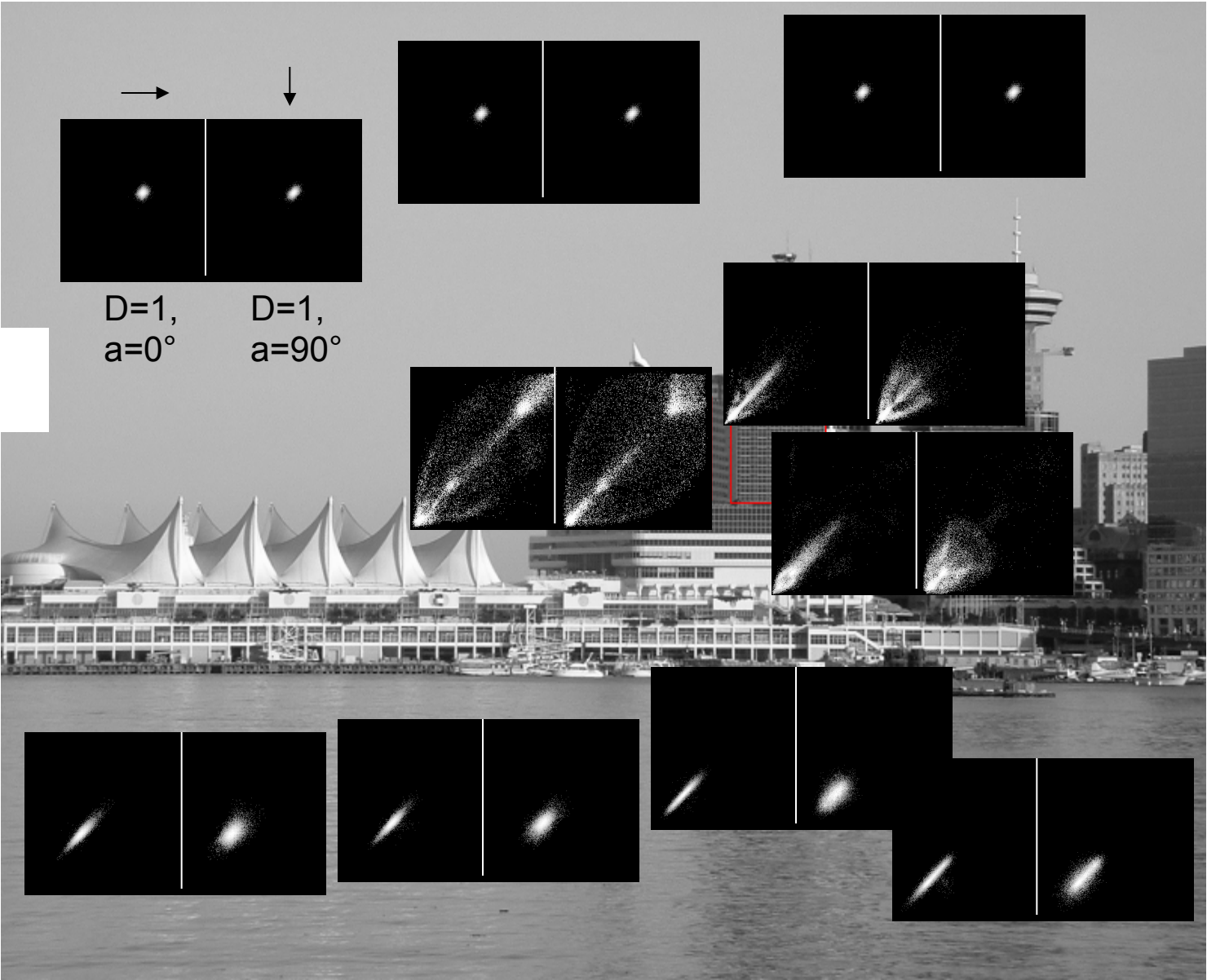
Textur beschreibt das **Objektinnere**.



Beispiel



Beispiel



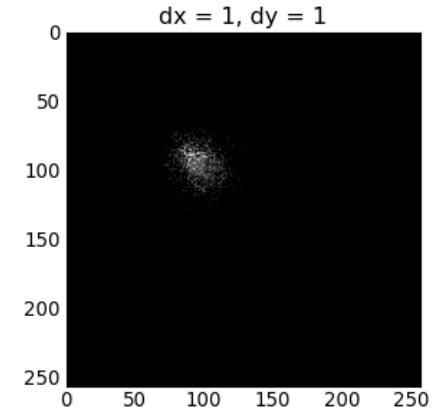
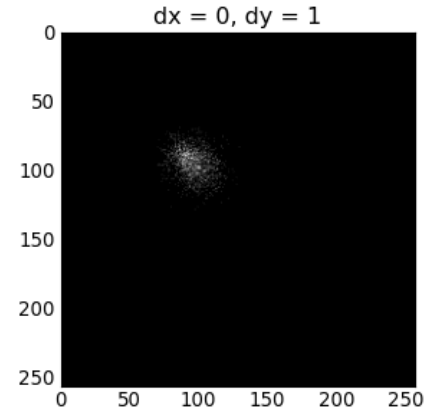
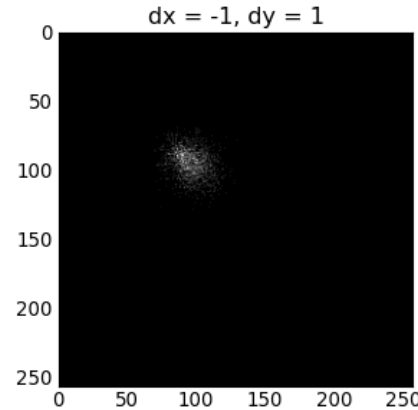
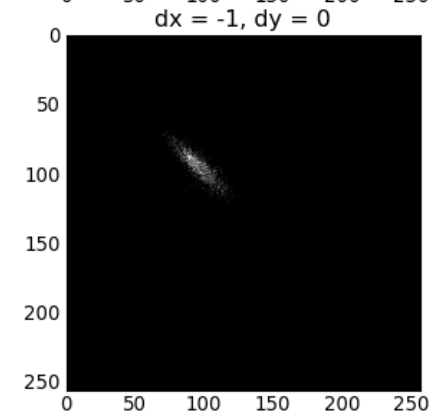
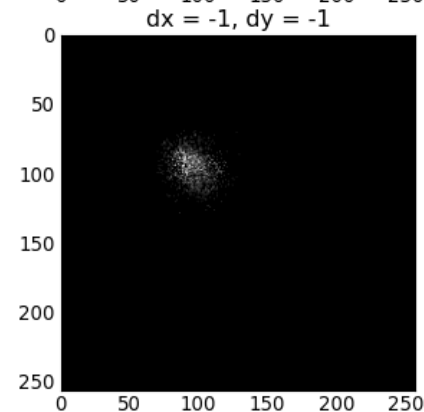
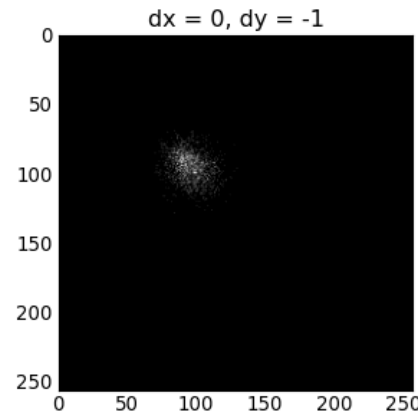
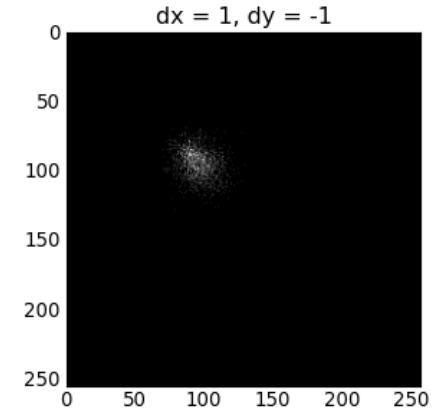
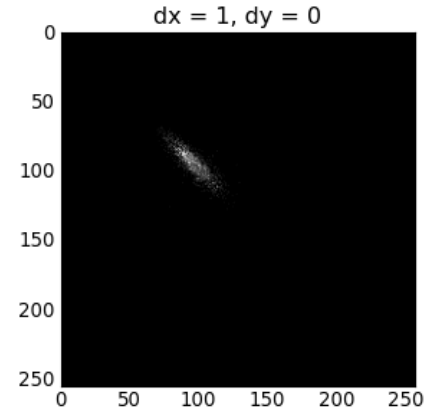
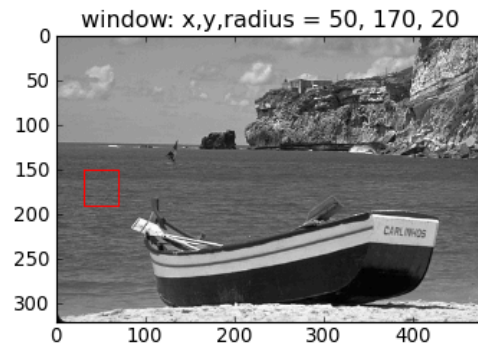
$D=1,$
 $a=0^\circ$

$D=1,$
 $a=90^\circ$

Co-Occurrence Matrix

für 8 Nachbarn

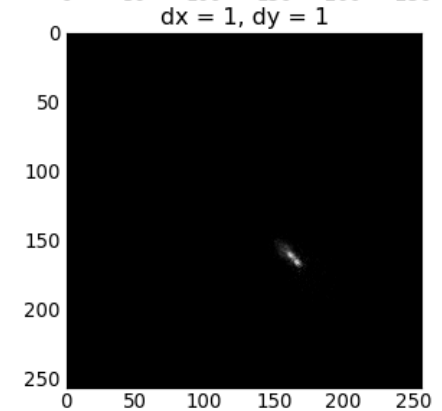
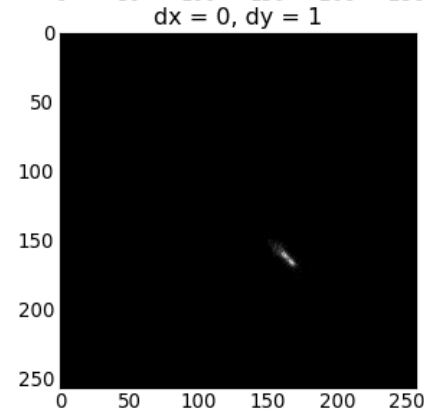
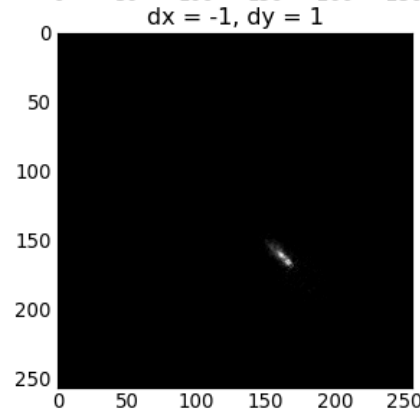
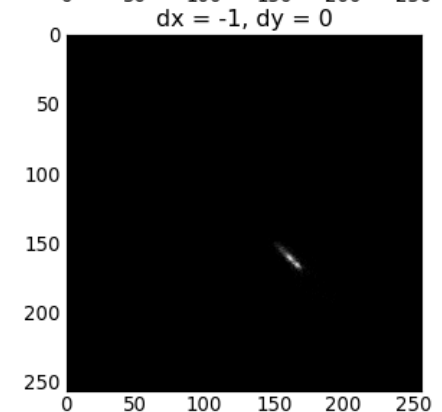
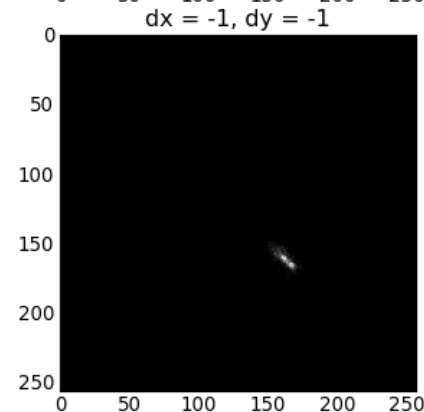
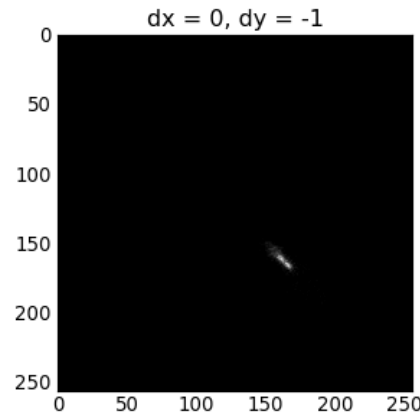
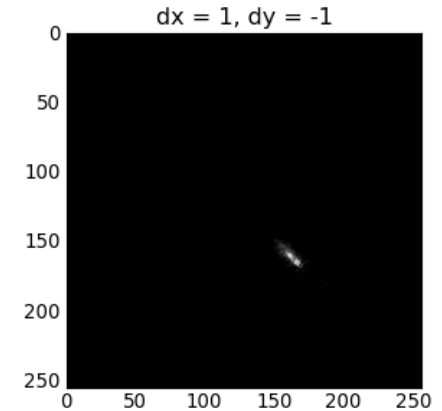
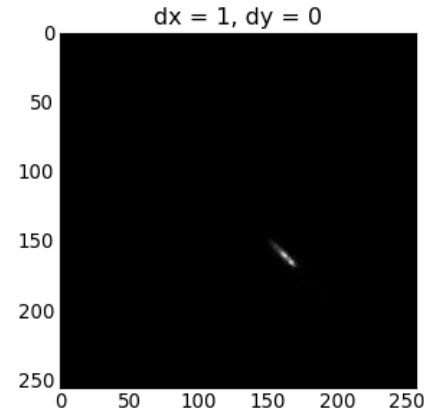
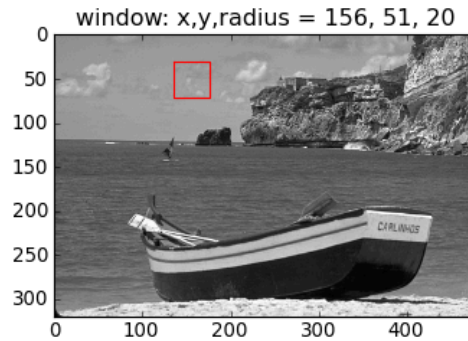
Matrix für (dx, dy)
=
Matrix für $(-dx, -dy)$



Co-Occurrence Matrix

für 8 Nachbarn

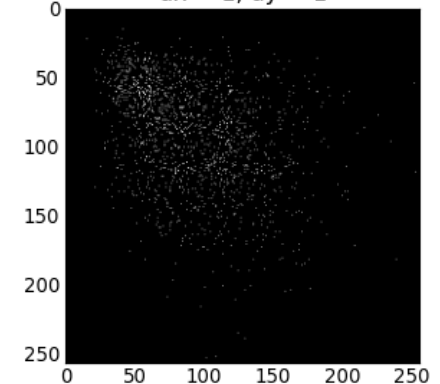
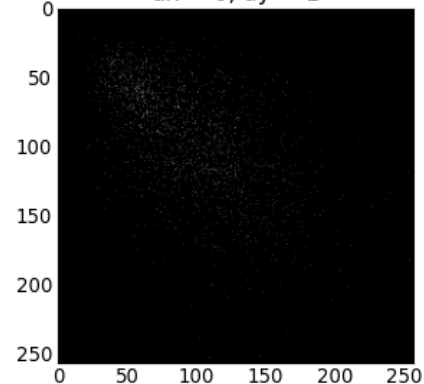
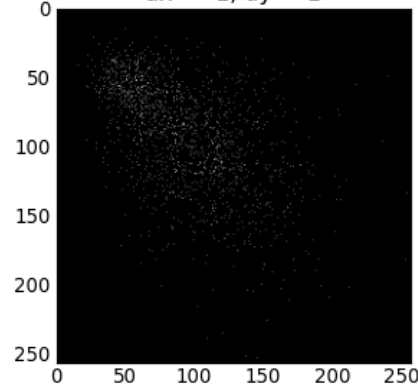
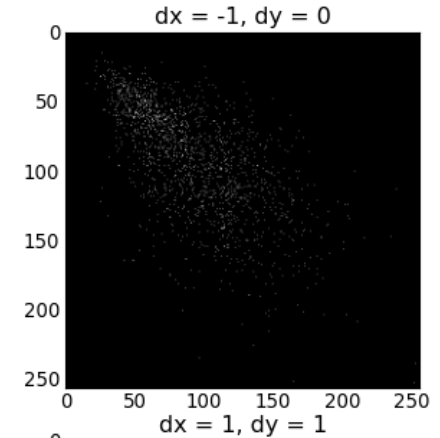
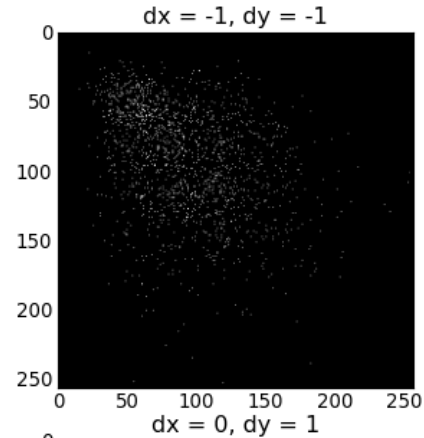
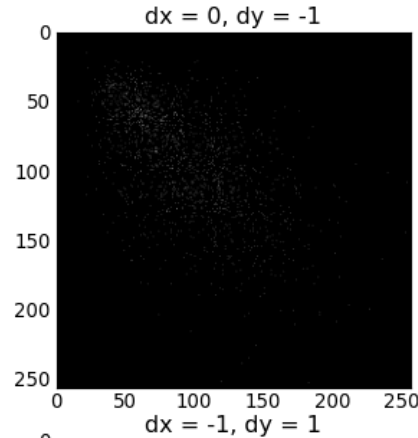
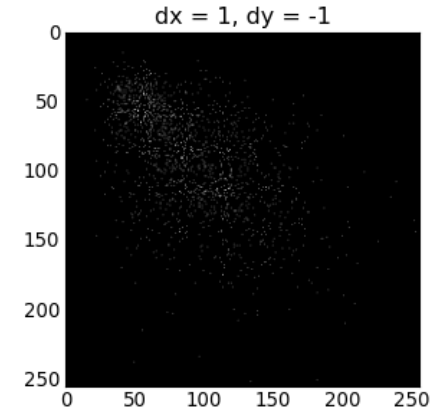
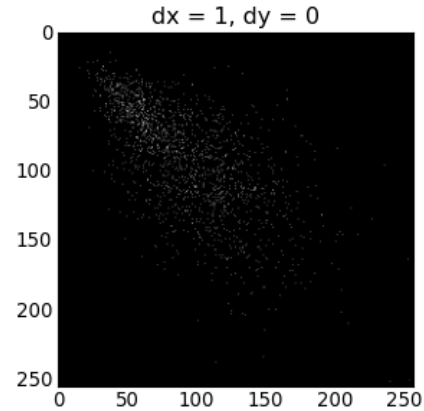
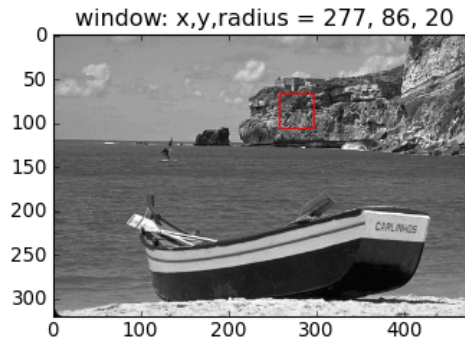
Matrix für (dx, dy)
=
Matrix für $(-dx, -dy)$



Co-Occurrence Matrix

für 8 Nachbarn

Matrix für (dx, dy)
=
Matrix für $(-dx, -dy)$



Haralick'sche Texturmaße

zunächst $P_{\Delta,\alpha}$ normieren: $P_{\Delta,\alpha} := \frac{1}{s} P_{\Delta,\alpha}$ mit $s = \sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2)$

Energie / Uniformität $\sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}^2(g_1, g_2)$

Kontrast $\sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} (g_1 - g_2)^2 \cdot P_{\Delta,\alpha}(g_1, g_2)$

Entropie $-\sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2) \cdot \log_2 [P_{\Delta,\alpha}(g_1, g_2)]$

Homogenität /
inverse Differenz $\sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} \frac{P_{\Delta,\alpha}(g_1, g_2)}{1 + |g_1 - g_2|}$

- liefern aussagekräftige Kennwerte für Texturen
- zur Segmentierung
 - Berechnung für $\Delta = 1$ und $\alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ$
 - Merkmalsvektor aus Texturmaßen
 - empfohlene Merkmale zur Texturklassifikation: Entropie, Kontrast, Korrelation

Haralick'sche Texturmaße (weitere)

zunächst $P_{\Delta,\alpha}$ normieren: $P_{\Delta,\alpha} := \frac{1}{s} P_{\Delta,\alpha}$ mit $s = \sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2)$

$$\text{Korrelation} \quad \sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} \frac{(g_1 - \mu_1) \cdot (g_2 - \mu_2) \cdot P_{\Delta,\alpha}(g_1, g_2)}{\sigma_1 \sigma_2}$$

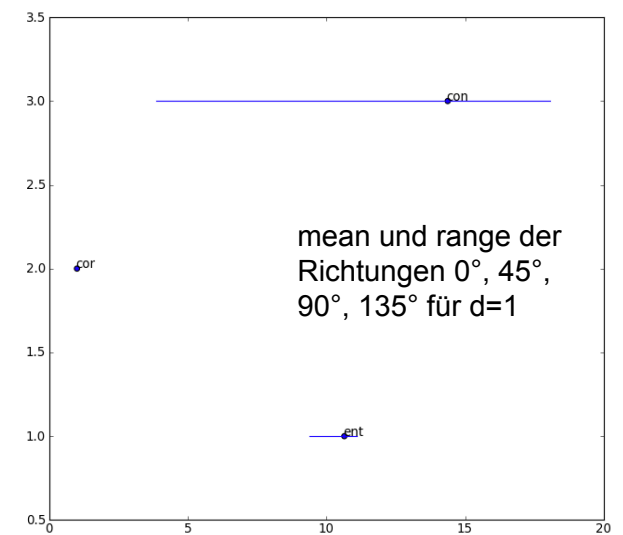
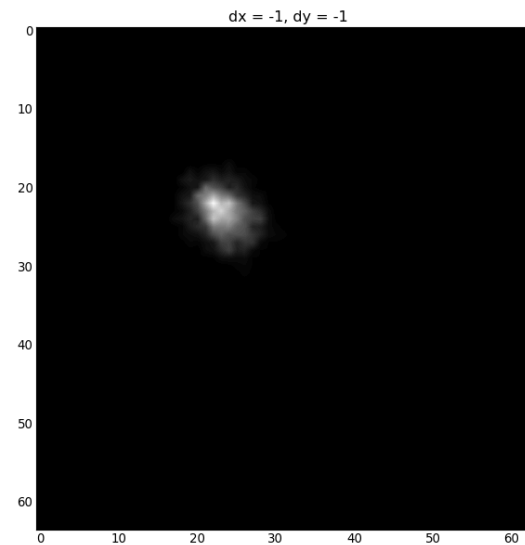
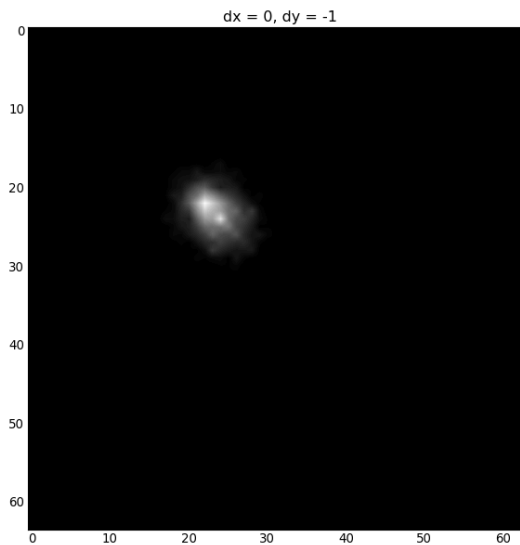
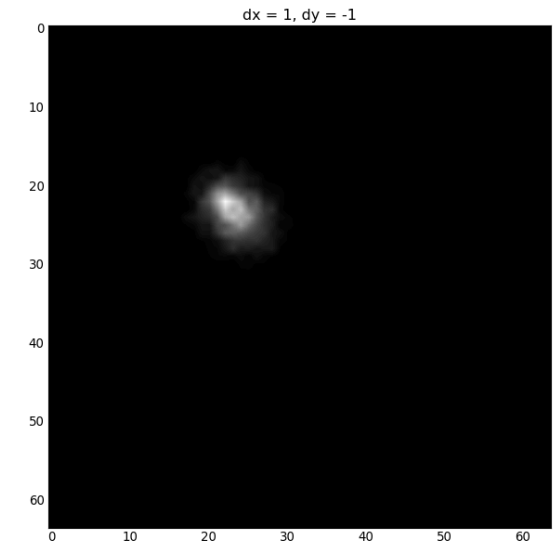
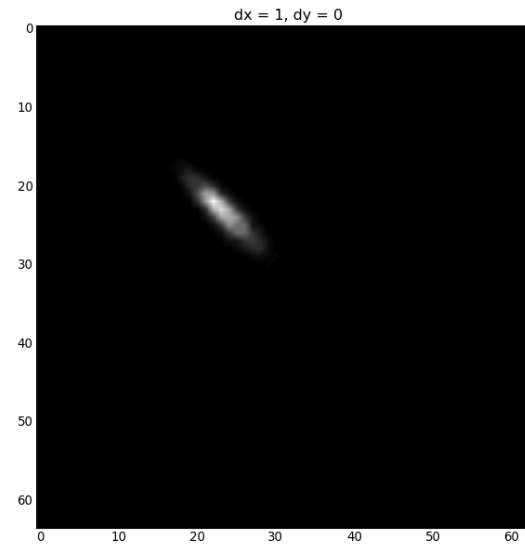
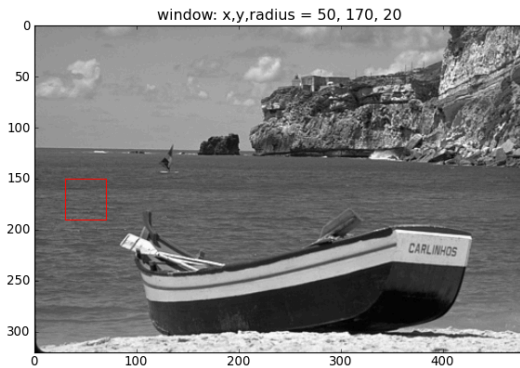
mit

$$\begin{aligned} \mu_1 &= \sum_{g_1=0}^{K-1} g_1 \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2) & \sigma_1 &= \sqrt{\sum_{g_1=0}^{K-1} (g_1 - \mu_1)^2 \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2)} \\ \mu_2 &= \sum_{g_2=0}^{K-1} g_2 \sum_{g_1=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2) & \sigma_2 &= \sqrt{\sum_{g_2=0}^{K-1} (g_2 - \mu_2)^2 \sum_{g_1=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2)} \end{aligned}$$

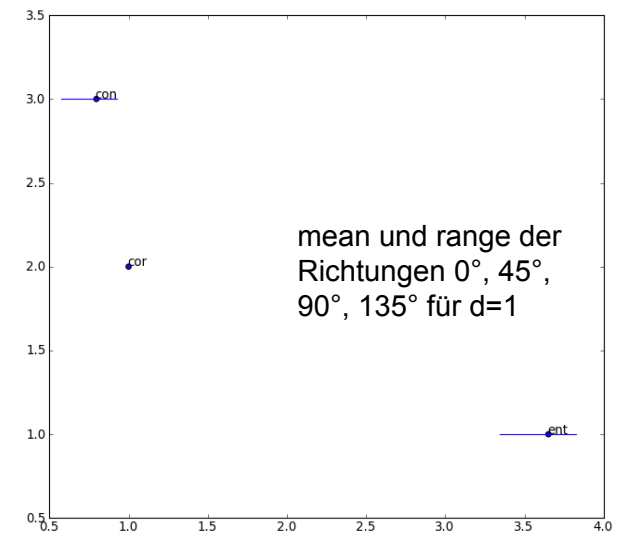
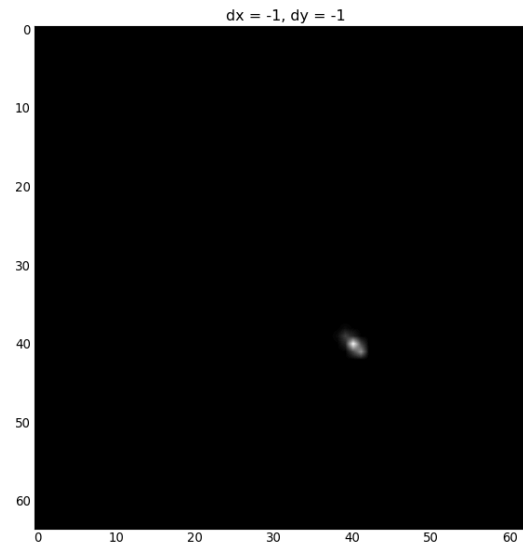
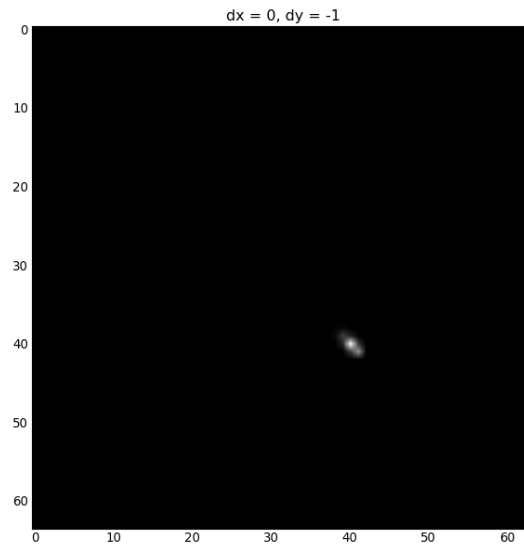
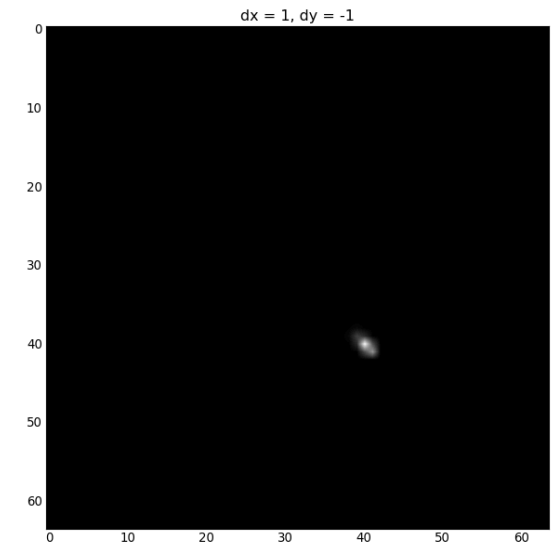
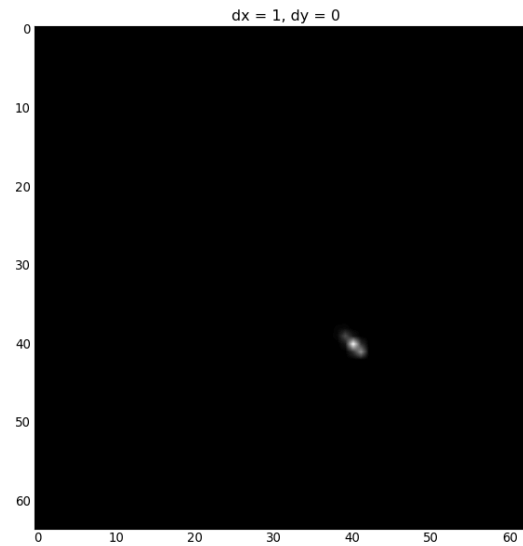
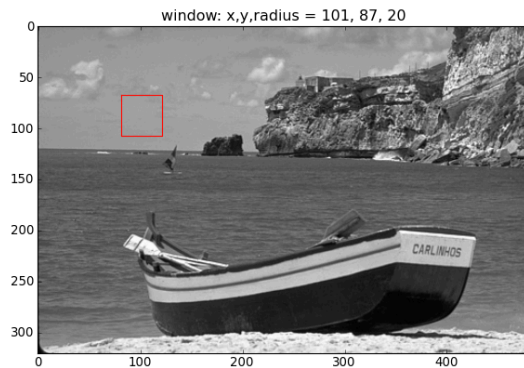
$$\text{inverse difference moment} \quad \sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} \frac{P_{\Delta,\alpha}(g_1, g_2)}{1 + (g_1 - g_2)^2}$$

$$\text{Unähnlichkeit} \quad \sum_{g_1=0}^{K-1} \sum_{g_2=0}^{K-1} P_{\Delta,\alpha}(g_1, g_2) \cdot |g_1 - g_2|$$

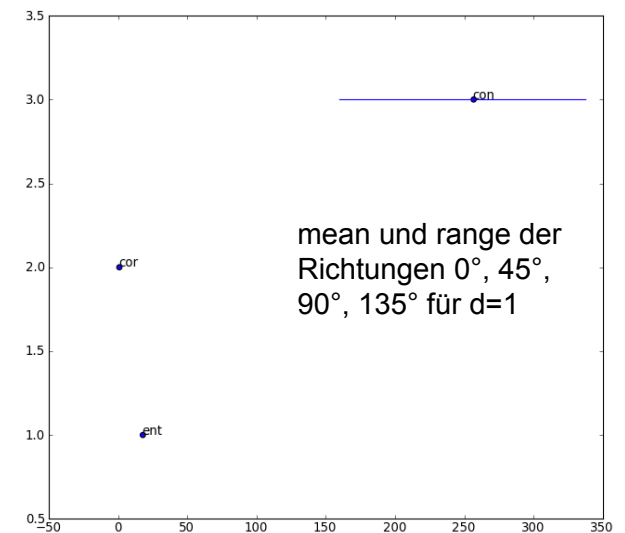
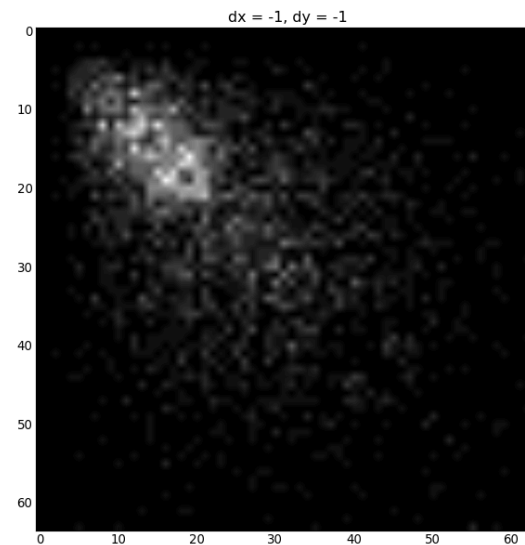
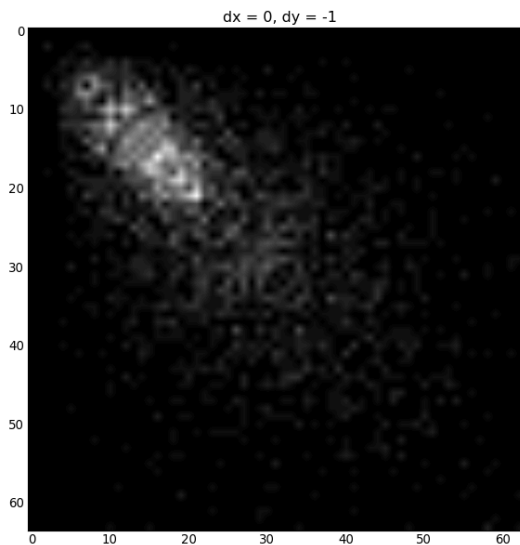
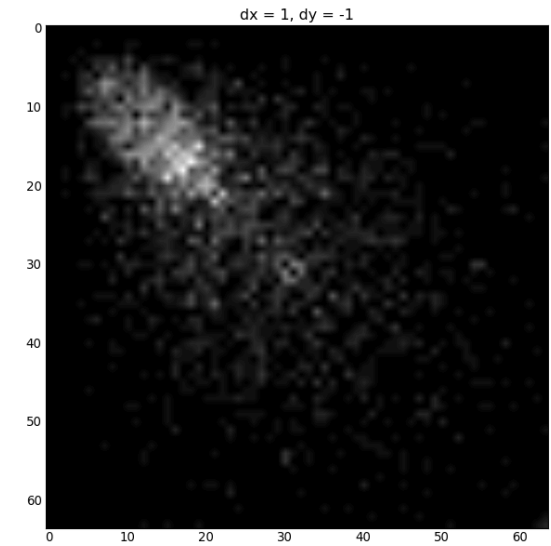
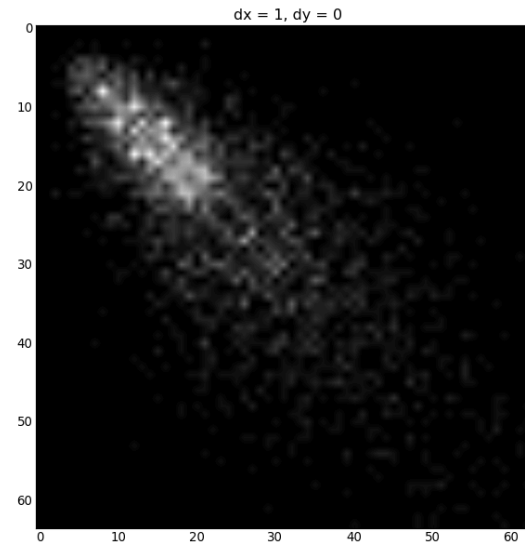
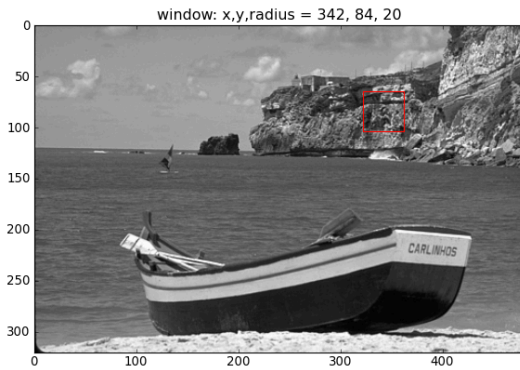
Co-Occurrence Matrix



Co-Occurrence Matrix

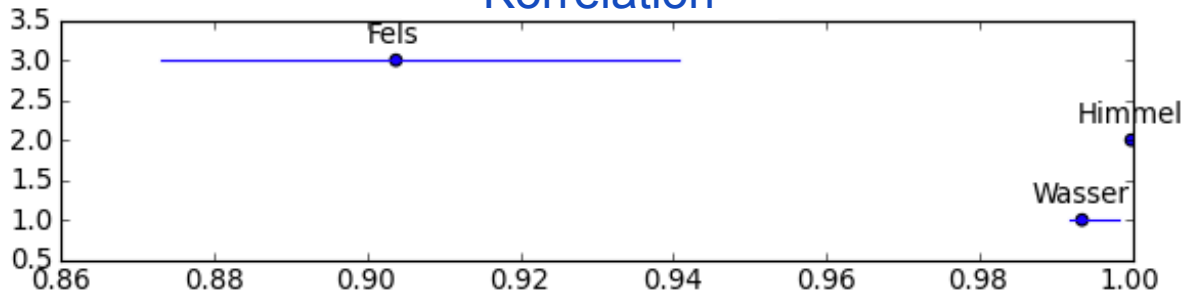


Co-Occurrence Matrix

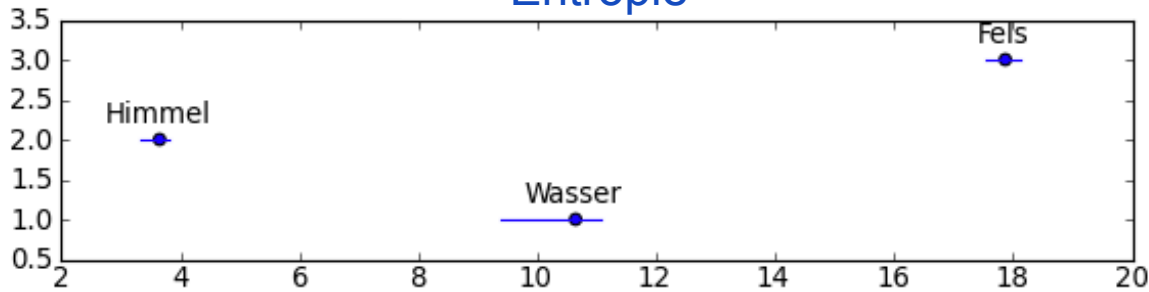


Korrelation, Entropie, Kontrast

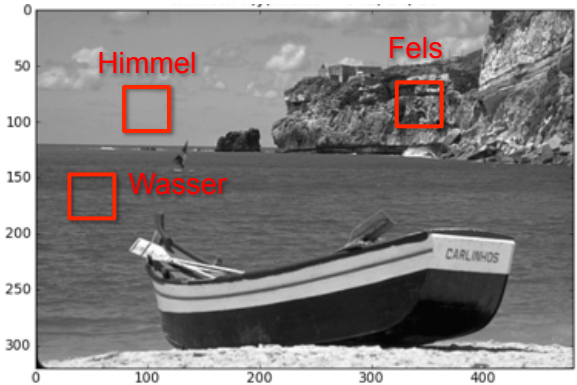
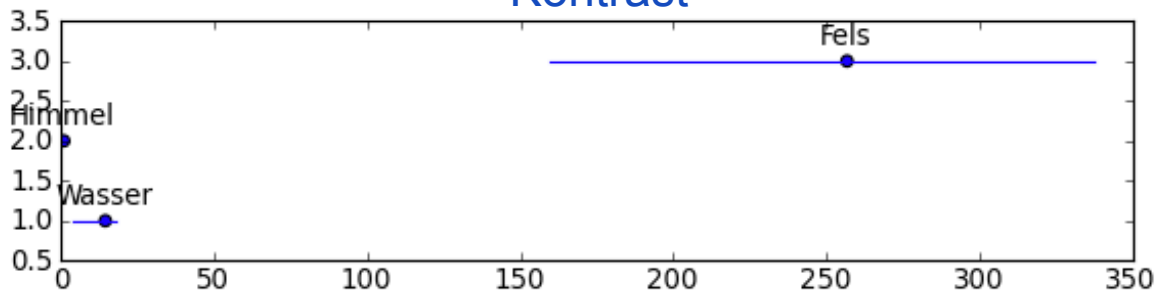
Korrelation



Entropie

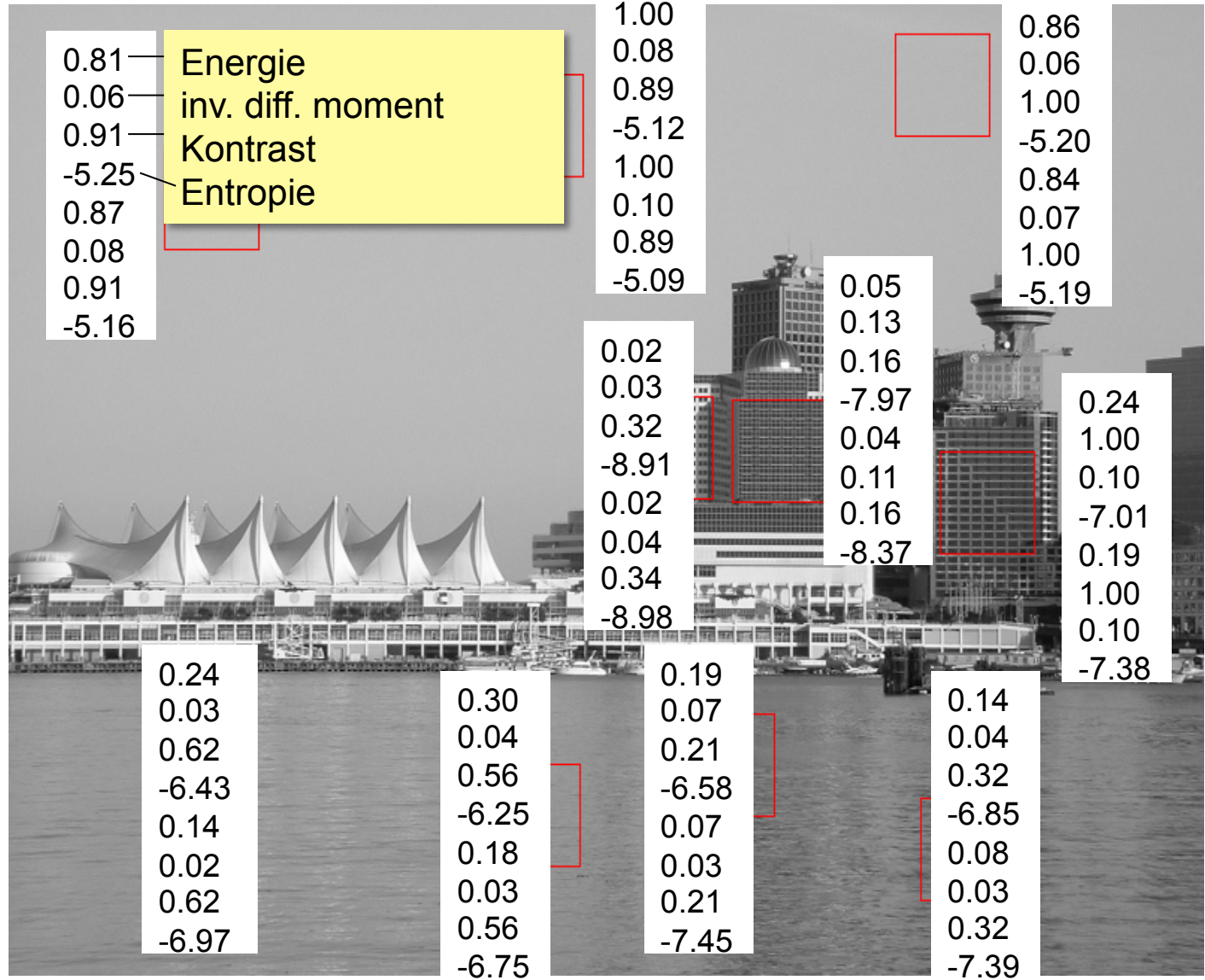


Kontrast



mean, min, max berechnet
über die Winkel 0° , 45° ,
 90° , 135°

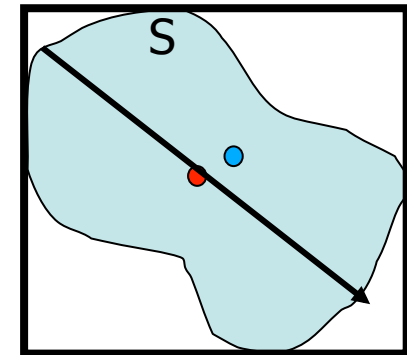
Merkmals- vektoren aus Hara- licksches Textur- maßen



Formmerkmale

- Formmerkmale beschreiben das Äußere von Segmenten
- Allgemeine Merkmale
 - Mittelpunkt der Bounding Box
 - Schwerpunkt m des Segments
- Translationsinvariante Merkmale
 - Bounding Box
 - Richtung der größten Ausdehnung und Ausdehnung in dieser Richtung

$$\vec{m}_g = \frac{1}{\sum_{p(i,j) \in S} p(i,j)} \sum_{p(i,j) \in S} \begin{pmatrix} i \\ j \end{pmatrix} \cdot p(i,j)$$



Formmerkmale

- Translations- und rotationsinvariante Merkmale
 - Eingeschlossene Fläche: F
 - Länge des Randes: R
 - Durchschnittliche Krümmungsänderung

- Translations-, rotations- und skalierungsunabhängige Merkmale
 - Kreisähnlichkeit, d.h. $4\pi F/R^2$
 - Hauptkomponentenanalyse
(Principal Component Analysis, PCA)

Hauptkomponentenanalyse

- Gegeben Form eines Objekts
 - Pixel des Objekts (bzw. dessen Rand) der Form $\mathbf{x} = (x_1, x_2)^T$
- Finde lokales Koordinatensystem, das Größe, Rotation, und Translation normalisiert
 - Berechne Schwerpunkt

$$\vec{m}_x = \frac{1}{K} \sum_{k=1}^K \vec{x}_k \quad \vec{x}' = \vec{x} - \vec{m}_x$$

- Berechne Kovarianzmatrix

$$C_x = \frac{1}{K} \sum_{k=1}^K (\vec{x}_k - \vec{m}_x) \cdot (\vec{x}_k - \vec{m}_x)^T = \frac{1}{K} \sum_{k=1}^K \vec{x}'_k \cdot \vec{x}'_k^T$$

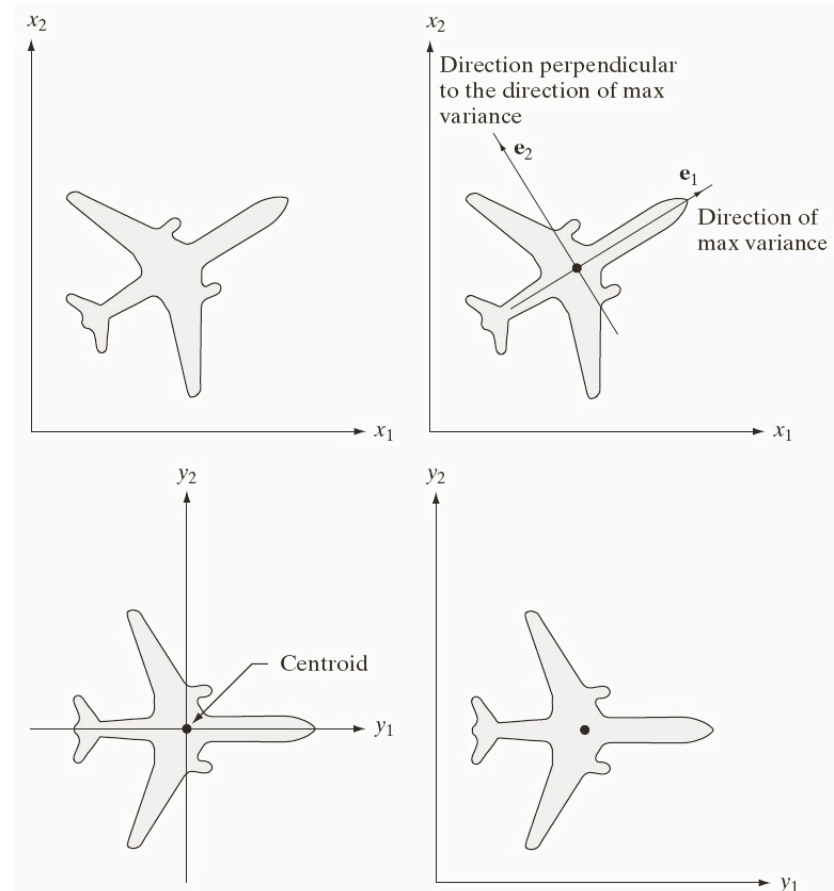


Abbildung: © R. C. Gonzalez & R. E. Woods, Digital Image Processing

Hauptkomponentenanalyse

- C_x ist real und symmetrisch \rightarrow C_x hat 2 orthogonale Eigenvektoren e_1, e_2 mit Eigenwerten λ_1, λ_2 ($\lambda_1 \geq \lambda_2$)

$\lambda_1 =$ Varianz der Hauptkomponente

- Sei A Matrix mit Zeilen e_1 und e_2
- Sei $y = A(x - m_x)$, dann ist $m_y = 0$ und die Elemente von y sind unkorreliert
- Wegen $A^{-1} = A^T$ gilt für die Rücktransformation $x = A^T y + m_x$
- Verhältnis λ_1 / λ_2 gut geeignet als Formmerkmal

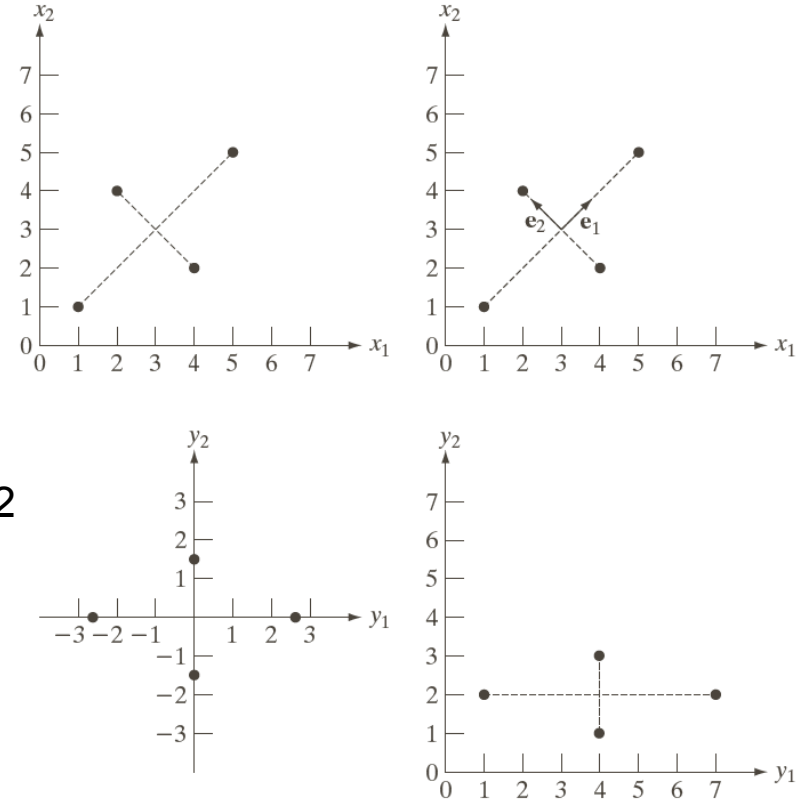
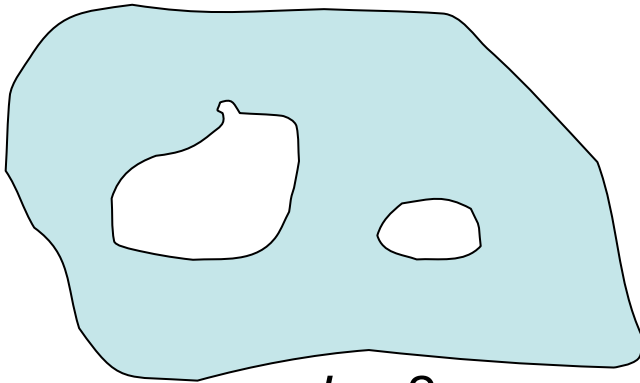
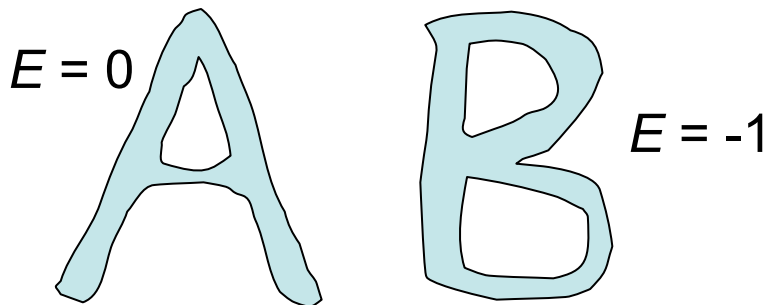


Abbildung: © R. C. Gonzalez & R. E. Woods, Digital Image Processing

Topologische Formmerkmale



$$\begin{aligned}L &= 2 \\C &= 1 \\E &= -1\end{aligned}$$



- Topologische Merkmale ändern sich auch dann nicht, wenn sich die Form des Objektes verändert, solange es nicht „zerrissen“ oder „geklebt“ wird.
- Beispiele
 - Anzahl der Löcher L
 - Anzahl C der verbundenen Strukturen eines Gebiets
 - Eulerzahl: $E=C-L$

Welche, wie viele Merkmale?

- Merkmale sollten **aus den vermuteten Objekteigenschaften abgeleitet** werden
- Merkmale sollten voneinander **unabhängig** sein
- Objekte der gleichen Klasse sollten im Merkmalsraum an ähnlichen Orten zu finden sei (**Häufungspunkte** der Klasse, **Cluster**)
- Segmente einer Klasse in eng begrenzter Region im Merkmalsraum; Segmente verschiedener Klassen gut voneinander separiert
- **Je weniger Merkmale** zur Unterscheidung notwendig sind, desto effektiver die Entscheidungsfindung

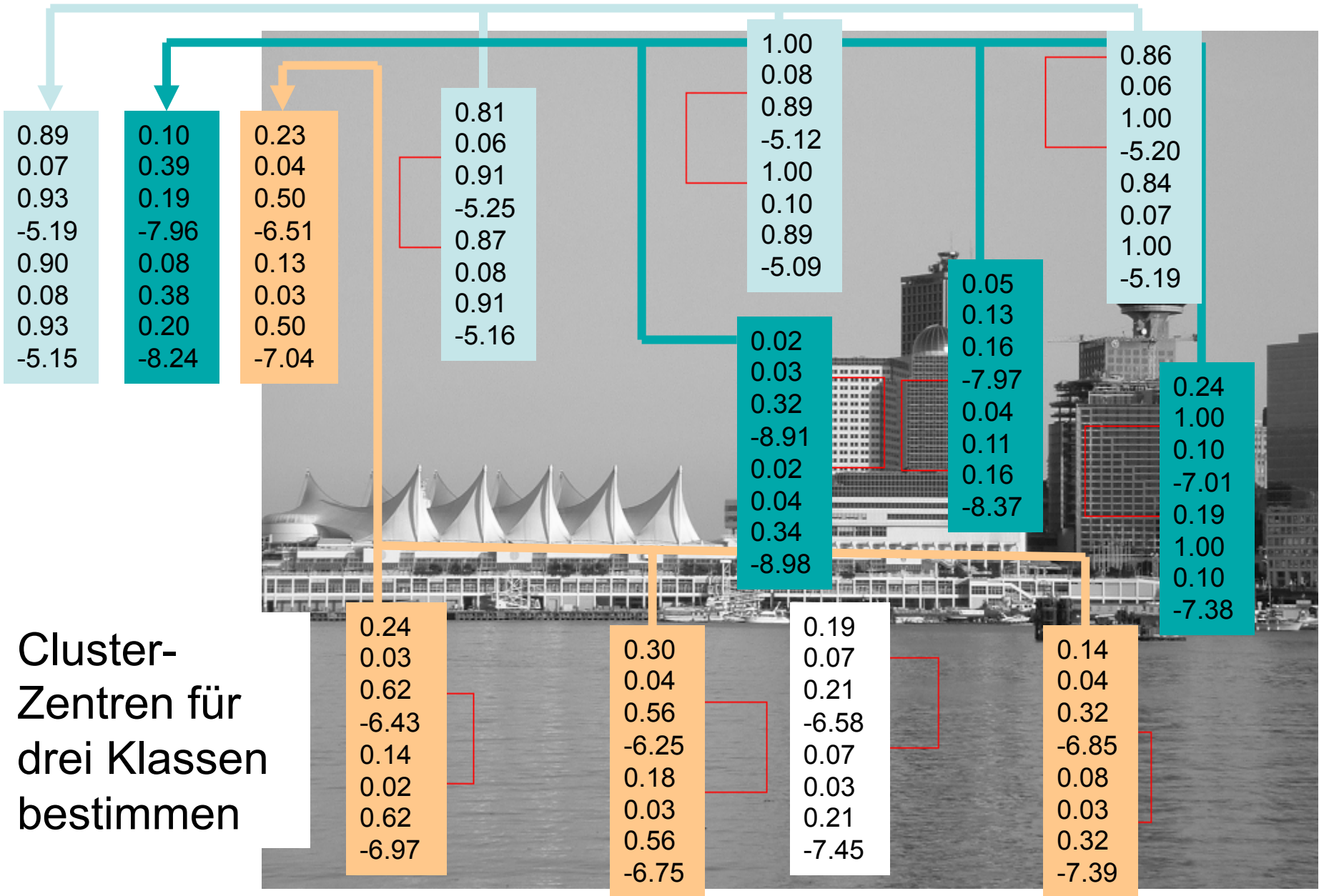
Klassifikation im n-dimensionalen Merkmalsraum

- Visuelle Suche nach Häufungspunkten oder Entscheidungsgrenzen ist nicht möglich
- Automatische Verfahren:
- **Optimale Methoden:** Minimierung der Wahrscheinlichkeit einer Fehlklassifikation
- **Klassifikation nach Training:** Entscheidung anhand von antrainierten Häufungspunkten oder **Entscheidungsgrenzen**
- **Clustering:** Automatische Suche nach Häufungsbereichen im Merkmalsraum

Methode des geringsten Abstandes

- Für jede der n Klassen existieren Stichproben
- Aus den Stichproben jeder Klasse wird ein durchschnittlicher Merkmalsvektor der Klasse berechnet (Clusterzentrum)
- Eine unbekannte Stichprobe wird derjenigen Klasse zugeordnet, zu der ihr Abstand am geringsten ist
- Häufige Probleme
 - Anzahl der Stichproben nicht ausreichend
 - Stichproben nicht repräsentativ
 - Unterschiedliche Skalierungen für die Merkmale

Cluster-Zentren für drei Klassen bestimmen

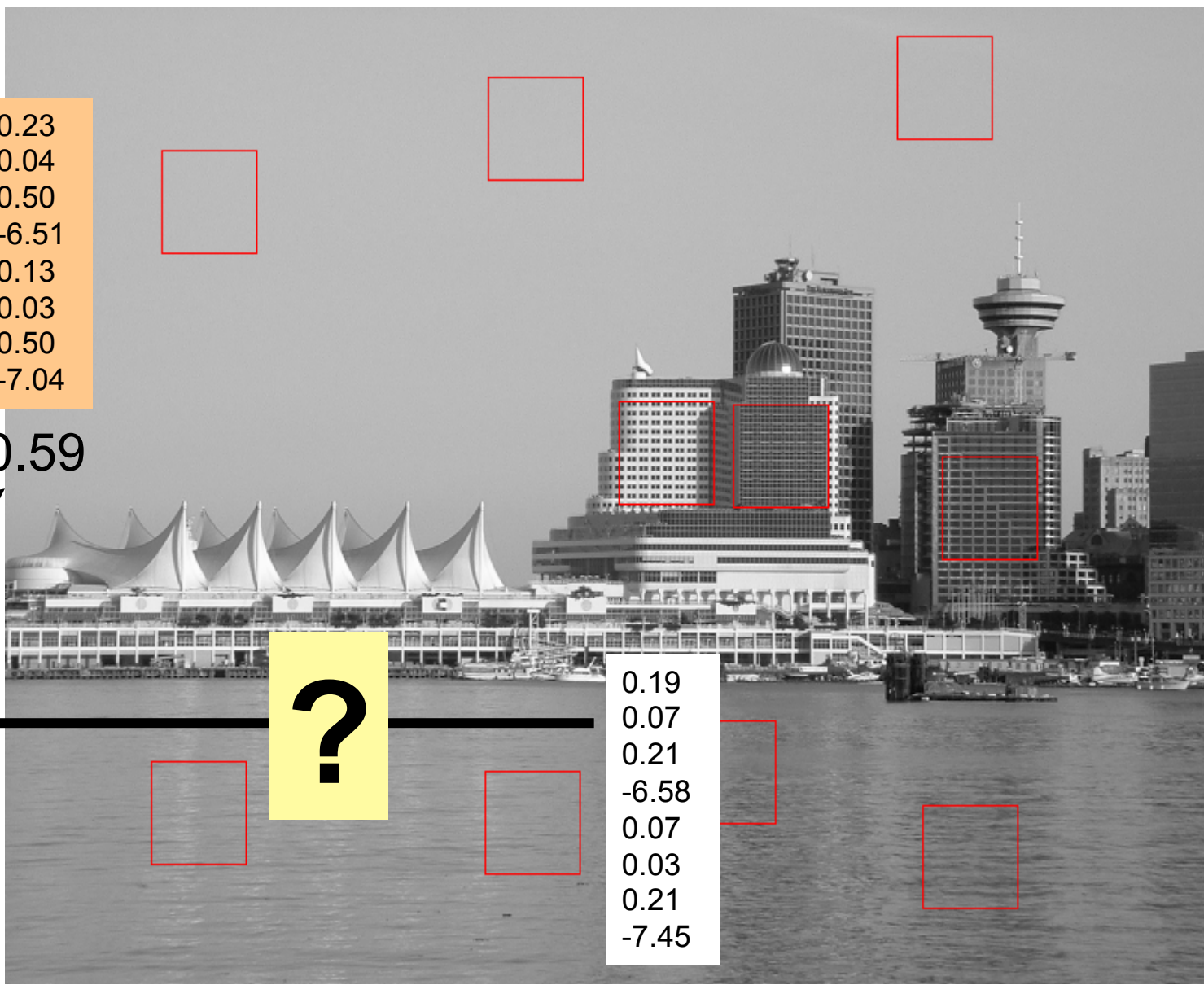


0.89	0.10	0.23
0.07	0.39	0.04
0.93	0.19	0.50
-5.19	-7.96	-6.51
0.90	0.08	0.13
0.08	0.38	0.03
0.93	0.20	0.50
-5.15	-8.24	-7.04

3.07 1.66 0.59

0.19
0.07
0.21
-6.58
0.07
0.03
0.21
-7.45

0.19
0.07
0.21
-6.58
0.07
0.03
0.21
-7.45



Bayes-sche Klassifikation

- Gibt es eine optimale Klassifikation?
- Optimal
 - Kriterium so gewählt, dass Anzahl der Fehlklassifikationen minimal
- Lösungsansatz
 - Merkmalsträger als Stichproben von statistischen Verteilungen der einzelnen Klassen im Merkmalsraum
 - Kriterium minimiert Wahrscheinlichkeit einer Fehlentscheidung

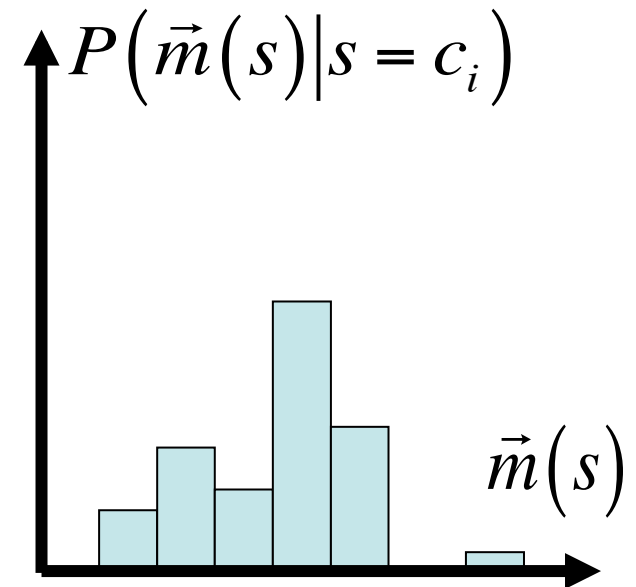


A-Posteriori-Wahrscheinlichkeit

- Wahrscheinlichkeit mit der ein Segment s mit Merkmalen m der Klasse c_i angehört:
$$P\left(s = c_i \mid \vec{m}(s)\right)$$
- **Klassifikation**: wähle diejenige Klasse, für die die Wahrscheinlichkeit am größten ist
- Die Anzahl der Fehlentscheidungen ist minimal, da die Lösung mit höchster Wahrscheinlichkeit gewählt wird
- Problem: A-Posteriori-Wahrscheinlichkeit schwer zu bestimmen
 - für jede Merkmalskombination müssen klassenabhängige Wahrscheinlichkeiten bestimmt werden

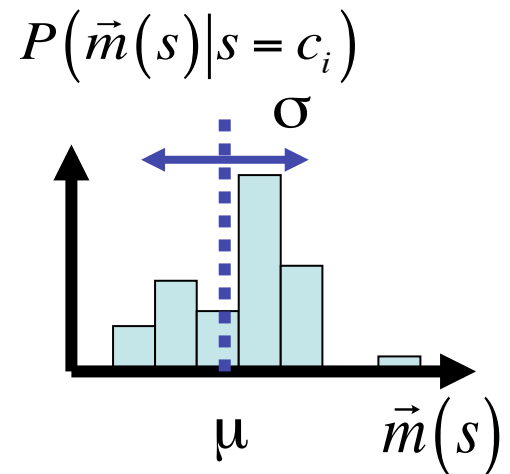
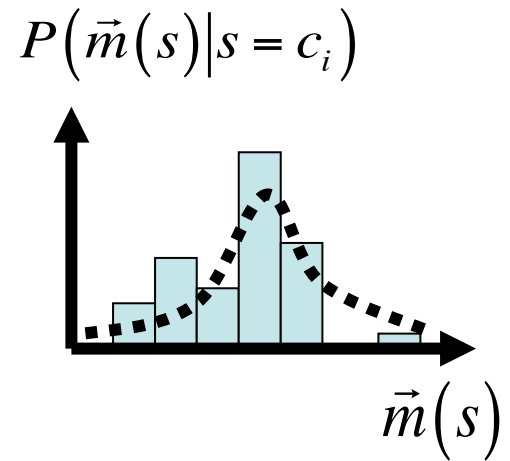
Merkmalsverteilungsfunktion

- Leichter ist zu bestimmen, wie sich für jede Klasse separat die Merkmalswerte verteilen: $P(\vec{m}(s) | s = c_i)$
- Bestimmung
 - Für jede Klasse c_i existiert eine Trainingsmenge T_i von bereits klassifizierten Segmenten
 - Trainingsmenge besteht aus Stichproben für die gesuchte Verteilungsfunktion
- Aus den Stichproben können die Verteilungsfunktionen oder deren Parameter geschätzt werden



Schätzung der Merkmalsverteilungsfunktion

- Schätzung einer beliebigen Funktion
 - Zerlegung des Merkmalsraums in Intervalle (Binning, Anzahl der Bins abhängig von der Anzahl der Stichproben)
 - Zuordnung der Stichproben zu Bins
 - Interpolation einer Funktion durch die Stichproben
- Schätzung von Parametern einer Funktion
 - Vorgabe einer Verteilungscharakteristik (z.B. Normalverteilung; muss begründbar sein!)
 - Schätzung der Parameter



A-Priori-Wahrscheinlichkeit

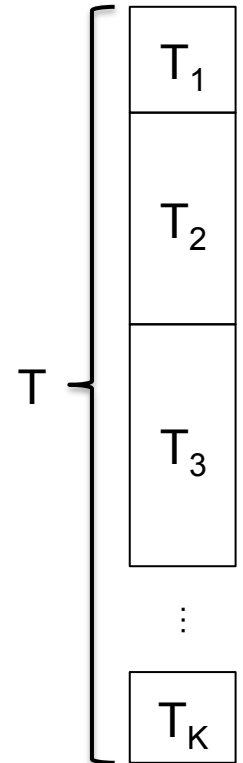
- A-Priori-Wahrscheinlichkeit: Klassenwahrscheinlichkeit ohne Kenntnis der Merkmale

$$P(s = c_i)$$

- Merkmalsverteilungsfunktion wird mit der A-Priori-Wahrscheinlichkeit gewichtet
- Beispiel:
 - Segmente von Personen allgemein:
 $P(s=\text{Mann})=0.5$, $P(s=\text{Frau})=0.5$
 - in Informatikfakultät:
 $P(s=\text{Mann})=0.8$, $P(s=\text{Frau})=0.2$

Schätzung der A-Priori-Wahrscheinlichkeit

- Bestimmung aus vorhandenem Wissen
 - Beispiel: „es gibt etwa gleich viele Männer und Frauen“
- Bestimmung aus der Trainingsmenge
 - Trainingsmenge $T = T_1 \cup T_2 \cup \dots \cup T_K$ ist repräsentativ für alle K Klassen
 - Anteil der Elemente von T_i an T entspricht der A-Priori-Wahrscheinlichkeit, dass ein beliebig ausgewähltes Element der Klasse c_j angehört



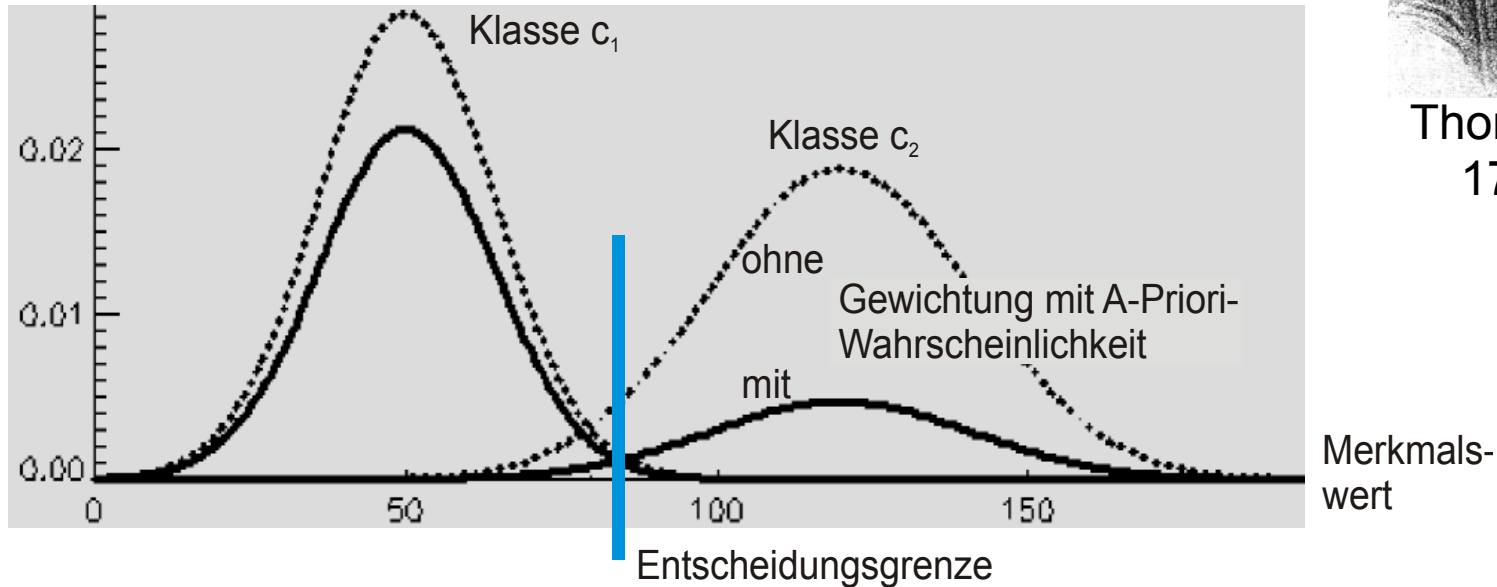
Satz von Bayes

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$



Thomas Bayes,
1700-1761

Wahrscheinlichkeit



$$P(s = c_i | \vec{m}(s)) = \frac{P(\vec{m}(s) | s = c_i) \cdot P(s = c_i)}{\sum_{k=1}^K P(\vec{m}(s) | s = c_k) \cdot P(s = c_k)}$$

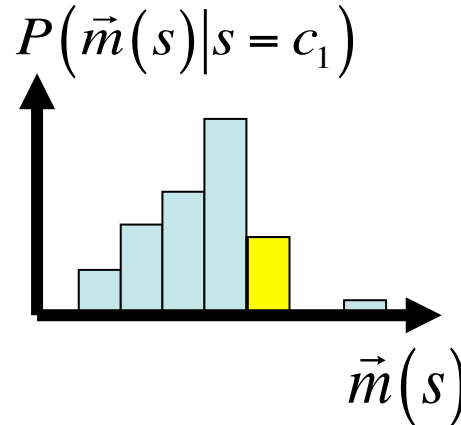
Satz von Bayes



Thomas Bayes,
1700-1761

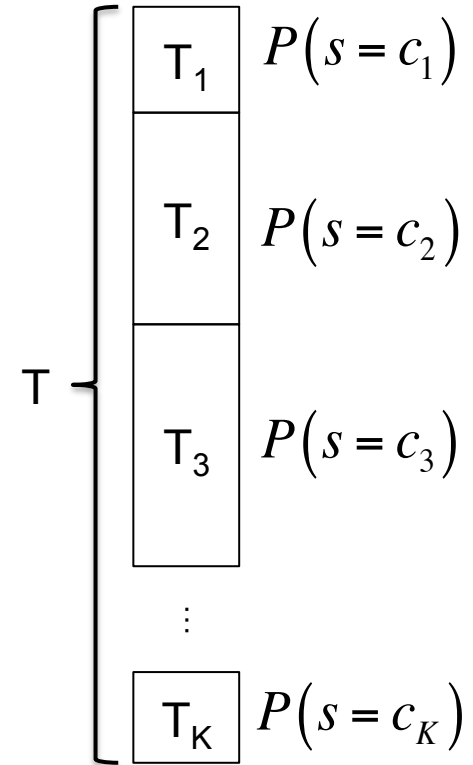
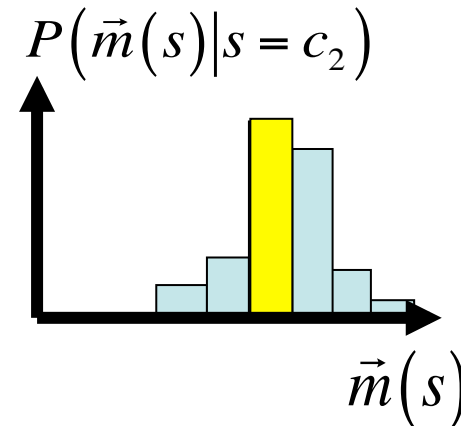
$$P(s = c_i | \vec{m}(s)) = \frac{P(\vec{m}(s) | s = c_i) \cdot P(s = c_i)}{\sum_{k=1}^K P(\vec{m}(s) | s = c_k) \cdot P(s = c_k)}$$

Klasse c_1 :



$$\sum_{k=1}^K P(\vec{m}(s) | s = c_k) \cdot P(s = c_k)$$

Klasse c_2 :



Beispiel



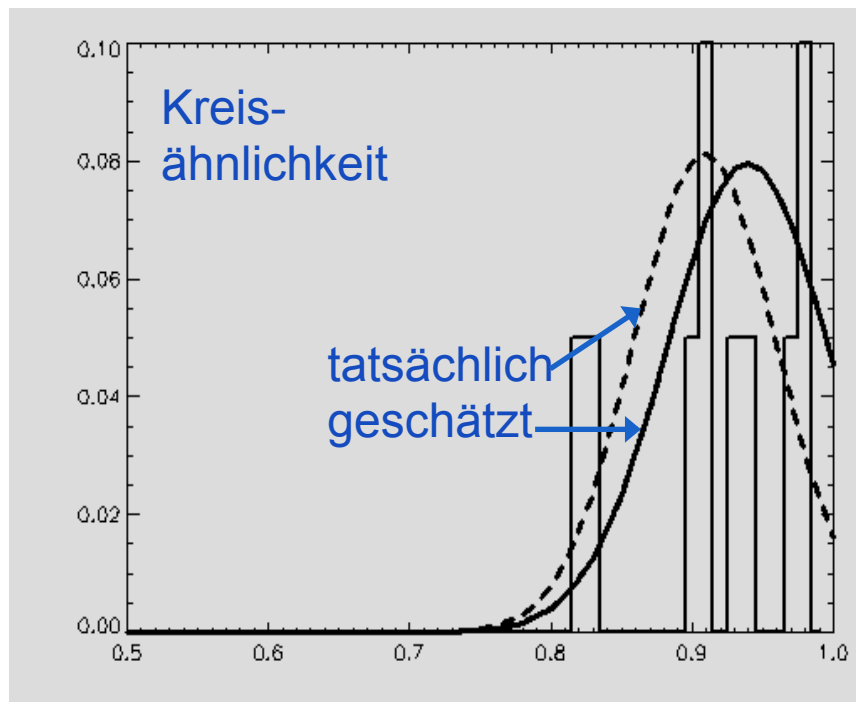
- Gegeben sei eine Menge von Bildern
 - jedes Bild enthält eine Frucht (a : Apfel oder b : Birne), die segmentiert wurde
- Merkmale m_1 und m_2 sind
 - m_1 = Kreisähnlichkeit (Äpfel sind runder als Birnen)
 - m_2 = Helligkeit (Äpfel sind dunkler als Birnen)
 - beide Merkmale sind unabhängig voneinander
- Es existiert für jede Klasse eine Trainingsmenge von 10 segmentierten Früchten
 - Die A-Priori-Wahrscheinlichkeit $P(s=a) = 0.67$ und $P(s=b) = 0.33$ (d.h. doppelt so viele Äpfel wie Birnen)

Merkmalsverteilungsfunktionen (Äpfel)

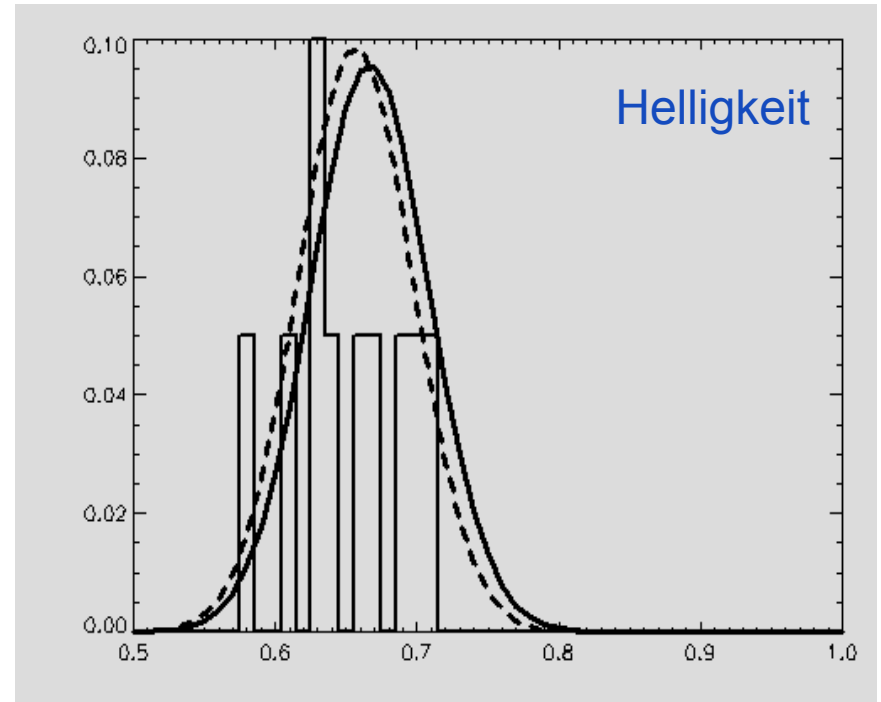


Merkmale in der Trainingsmenge

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} 0.95 & 0.92 & 0.84 & 0.99 & 0.91 & 0.92 & 0.94 & 0.98 & 0.99 & 0.83 \\ 0.62 & 0.64 & 0.64 & 0.72 & 0.70 & 0.65 & 0.59 & 0.67 & 0.68 & 0.71 \end{pmatrix}$$



$$E(m_1) = 0.93, \sigma^2(m_1) = 0.0032$$

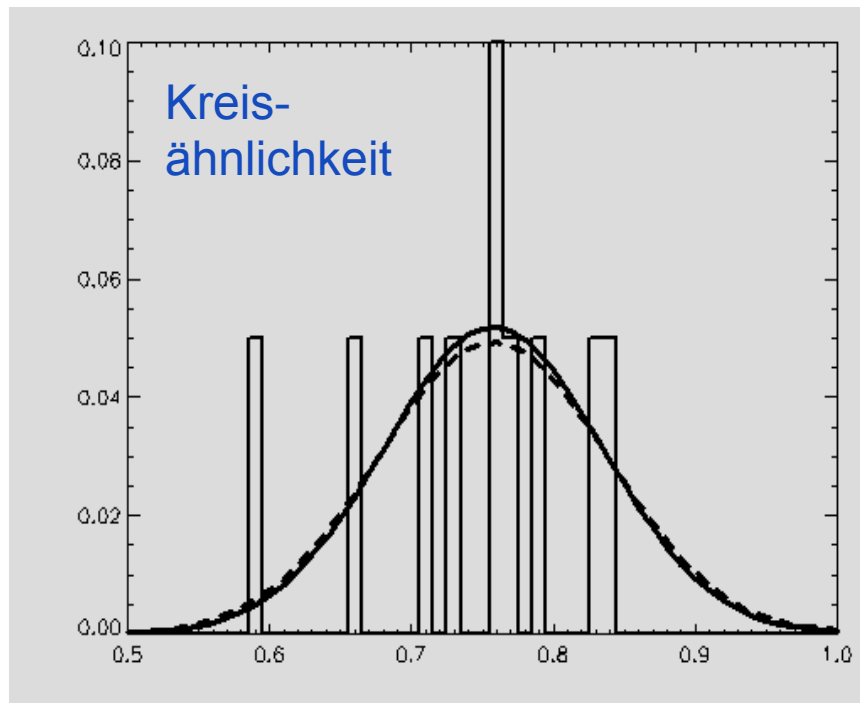


$$E(m_2) = 0.66, \sigma^2(m_2) = 0.0017$$

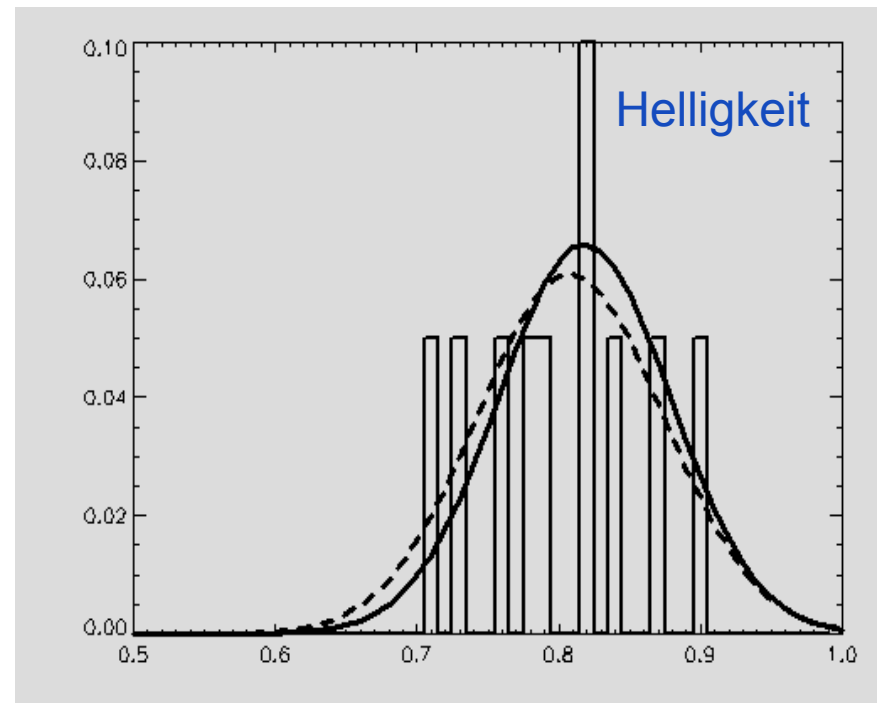
Merkmalsverteilungsfunktionen (Birnen)

Merkmale in der Trainingsmenge

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} 0.74 & 0.80 & 0.77 & 0.60 & 0.67 & 0.72 & 0.84 & 0.85 & 0.78 & 0.77 \\ 0.80 & 0.83 & 0.72 & 0.91 & 0.83 & 0.85 & 0.74 & 0.77 & 0.79 & 0.88 \end{pmatrix}$$

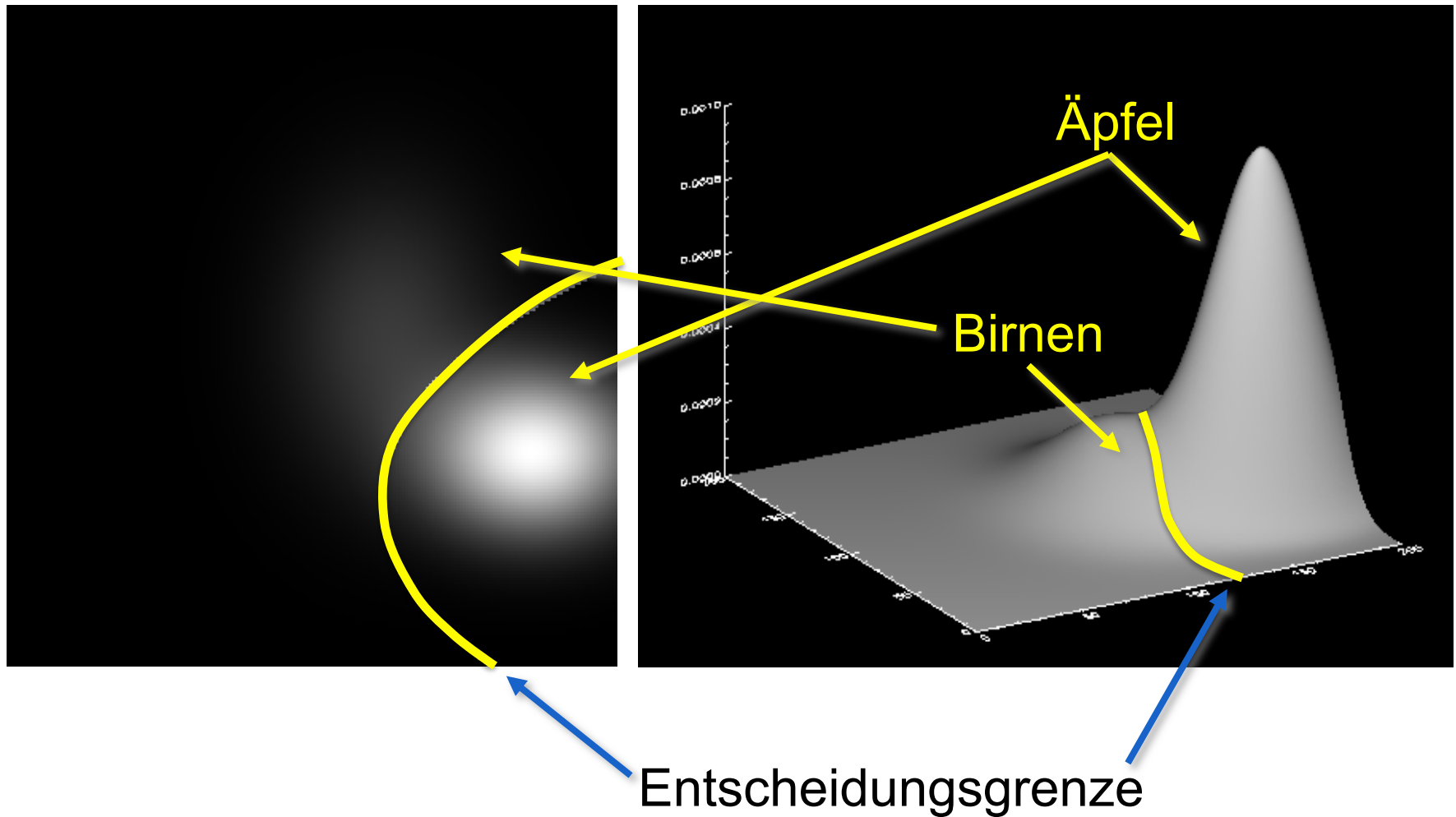


$$E(m_1) = 0.75, \sigma^2(m_1) = 0.0058$$



$$E(m_2) = 0.81, \sigma^2(m_2) = 0.0036$$

Überlagerte Klassenwahrscheinlichkeiten



Gesamtwahrscheinlichkeiten

Aus den Parametern der Testdatenmengen berechnete
2D-Verteilungsfunktionen

Annahme: unkorrelierte Merkmale! (Produkte der Einzelwahrsch.)

$m = (m_1, m_2) = (\text{Kreisähnlichkeit}, \text{Helligkeit})$

$$P(\bar{m}(s)|s = a) = \frac{1}{0.057\sqrt{2\pi}} \exp\left(-\frac{(m_1 - 0.93)^2}{0.0064}\right) \cdot \frac{1}{0.041\sqrt{2\pi}} \exp\left(-\frac{(m_2 - 0.66)^2}{0.0034}\right)$$

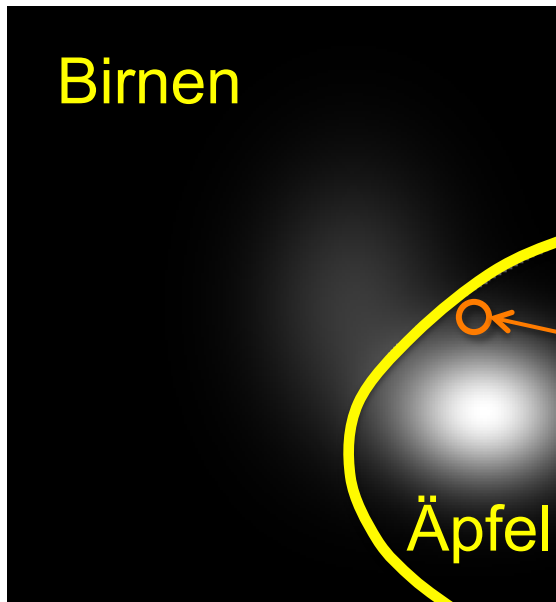
$$P(\bar{m}(s)|s = b) = \frac{1}{0.076\sqrt{2\pi}} \exp\left(-\frac{(m_1 - 0.75)^2}{0.0116}\right) \cdot \frac{1}{0.060\sqrt{2\pi}} \exp\left(-\frac{(m_2 - 0.81)^2}{0.0072}\right)$$

Klassifikation

Klassifikation eines unbekanntes Segments mit Werten (0.83, 0.72):

$$P\left(a \mid \begin{pmatrix} 0.83 & 0.72 \end{pmatrix}\right) = \frac{P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid a\right) \cdot 0.67}{P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid a\right) \cdot 0.67 + P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid b\right) \cdot 0.33} = 0.61$$

$$P\left(b \mid \begin{pmatrix} 0.83 & 0.72 \end{pmatrix}\right) = \frac{P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid b\right) \cdot 0.33}{P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid a\right) \cdot 0.67 + P\left(\begin{pmatrix} 0.83 & 0.72 \end{pmatrix} \mid b\right) \cdot 0.33} = 0.39$$

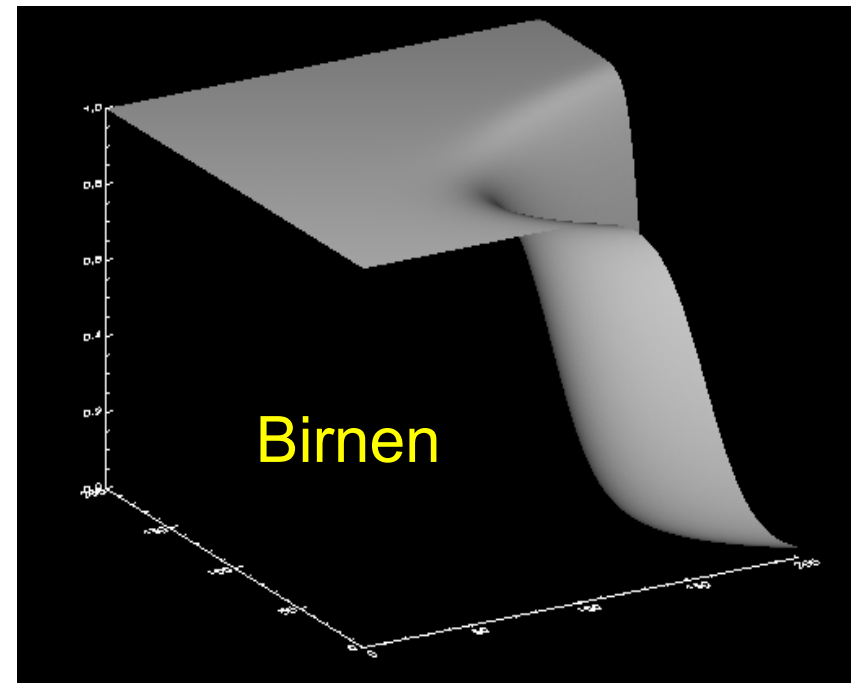
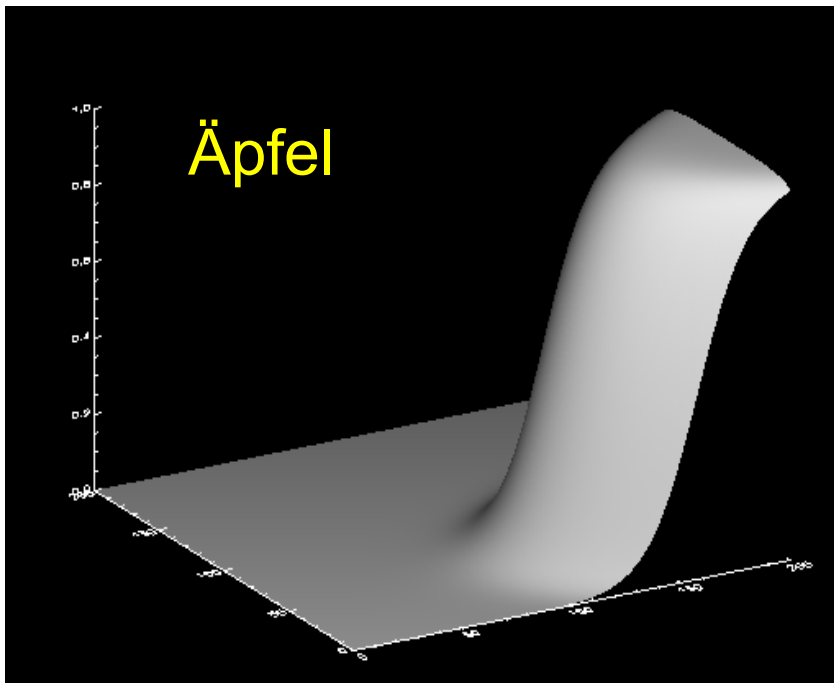


Tatsächliche Wahrscheinlichkeiten:

$P(a \mid (0.83 \ 0.72)) = 0.60$ und $P(b \mid (0.83 \ 0.72)) = 0.40$!

(0.83, 0.72) ... Apfel oder Birne?

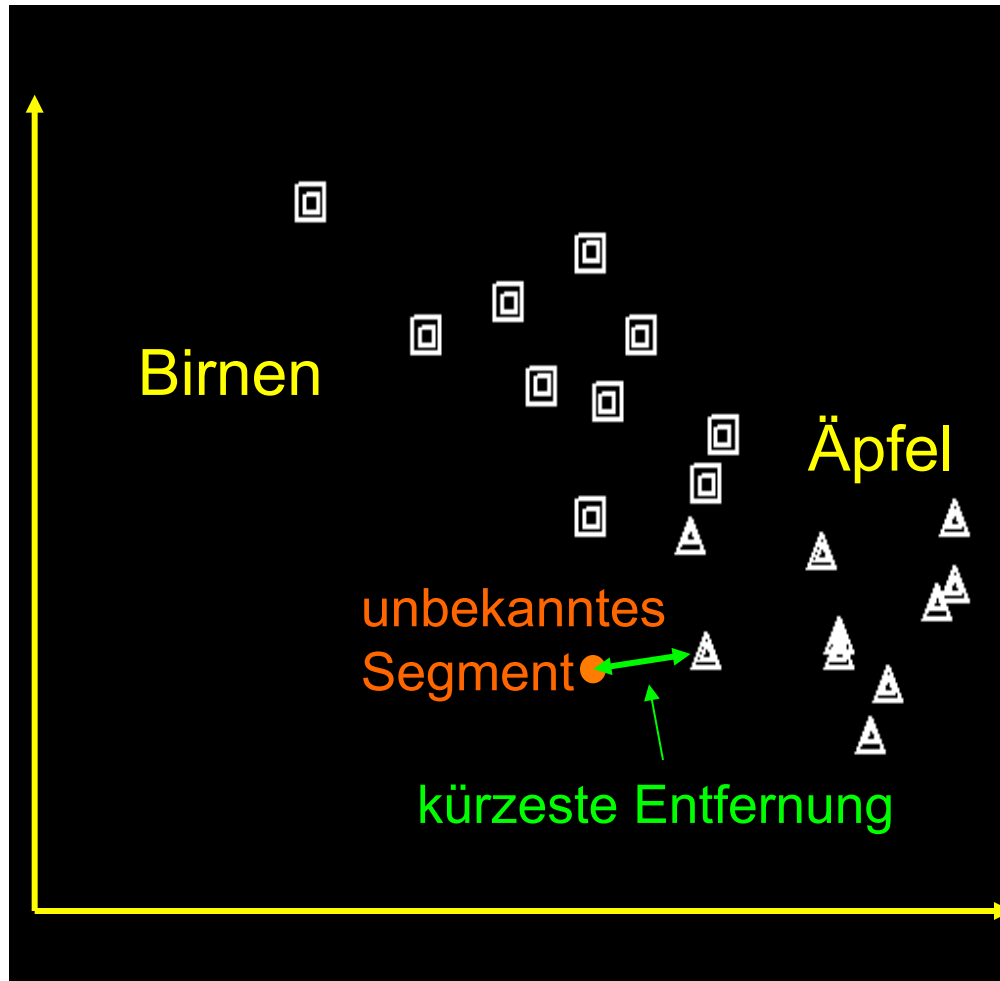
A-Posteriori-Wahrscheinlichkeiten



Direkte Schätzung der A-Posteriori-Wahrscheinlichkeiten

- Beobachtung
 - bei gut gewählten Merkmalen Unterschied zwischen den A-Posteriori-Wahrscheinlichkeiten für untersch. Klassen groß
 - selbst große Fehler bei der Schätzung der Merkmalsverteilungsfunktionen resultieren in geringen Fehlern bei der Schätzung der A-Posteriori-Wahrscheinlichkeit
- Konsequenz
 - direkte Schätzung der A-Posteriori-Wahrscheinlichkeit für eine gegebene Merkmalskombination

Single Nearest Neighbour



Für die zu bestimmende A-posteriori-Wahrscheinlichkeit wird die Umgebung **gerade so groß** gemacht, dass sie **eine klassifizierte Stichprobe** c_k umfasst.

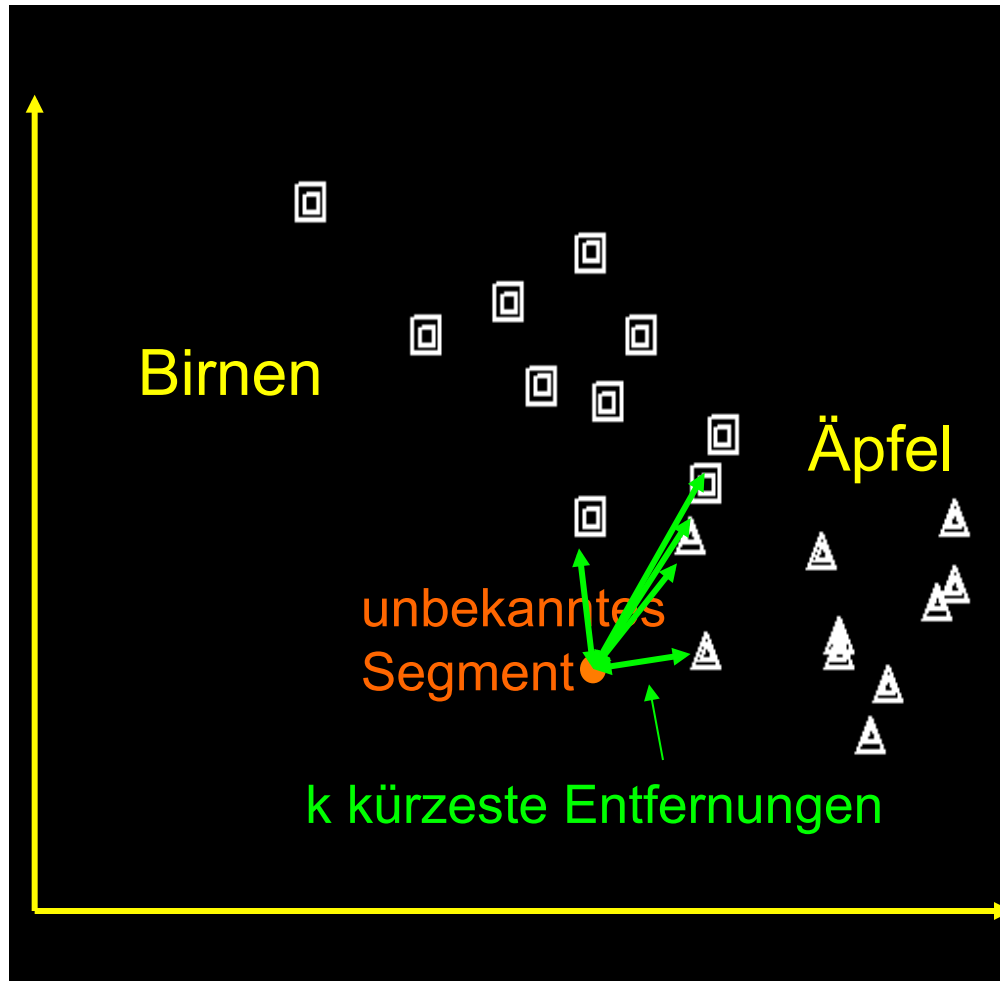
Schätzung von $P(s=c_i|m(s))$:

$$P = 1.0 \text{ für } c_i=c_k$$

$$P = 0.0 \text{ sonst}$$

anfällig gegenüber Ausreißern

k-Nearest-Neighbour (kNN)



Vergrößerung der Umgebung, so dass sie die k nächsten Nachbarn umfasst.

Schätzung von $P(s=c_i|m(s))$:

$$P = k_i/k$$

k_i : Anzahl der Stichproben, die der Klasse c_i angehören

kNN-Klassifikation

- je dichter die Stichproben liegen, desto kleiner ist Umgebung für gegebenes k
- falls k und die Stichprobendichte gegen unendlich gehen, nähert sich die Qualität der kNN-Klassifikation der der Bayes'schen Entscheidung an
- **Problem:** Für jede zu klassifizierende Stichprobe müssen die Entfernungen zu allen(!) Stichproben des Merkmalsraums berechnet werden
- **Lösung:** Vor der Klassifikation wird die Grenzfläche geschätzt, an der die Klasse wechselt

Lineare Entscheidungsgrenzen

- Klassifikation in zwei Klassen c_1 und c_2
- Klassen linear separierbar, falls es gibt lineare Funktion

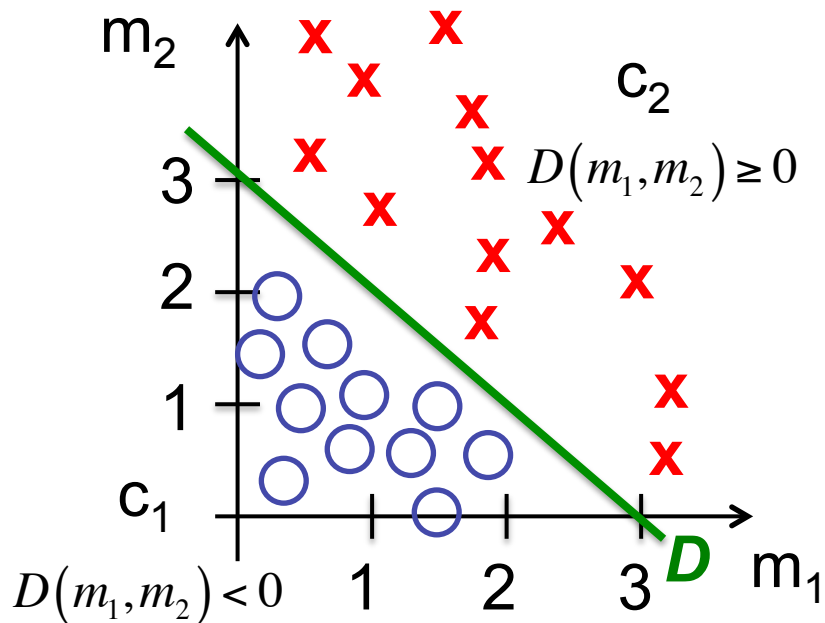
$$D(m_1, m_2, \dots, m_M) = w_0 + w_1 m_1 + \dots + w_M m_M$$

$$D(m) = w^T m \quad (m_0 = 1)$$

so dass $D(m) < 0$, falls Merkmale aus c_1
und $D(m) \geq 0$ falls Merkmale aus c_2

- **Training:** Finde die Parameter w_0, \dots, w_M
- **Klassifikation:** Einsetzen der Merkmale m eines unbekanntes Segments in Funktion D , Entscheidung nach Funktionswert

Lineare Entscheidungsgrenzen



Entscheidungsgrenze:

$$D(m_1, m_2) = 0$$

$$= w_0 + w_1 m_1 + w_2 m_2$$

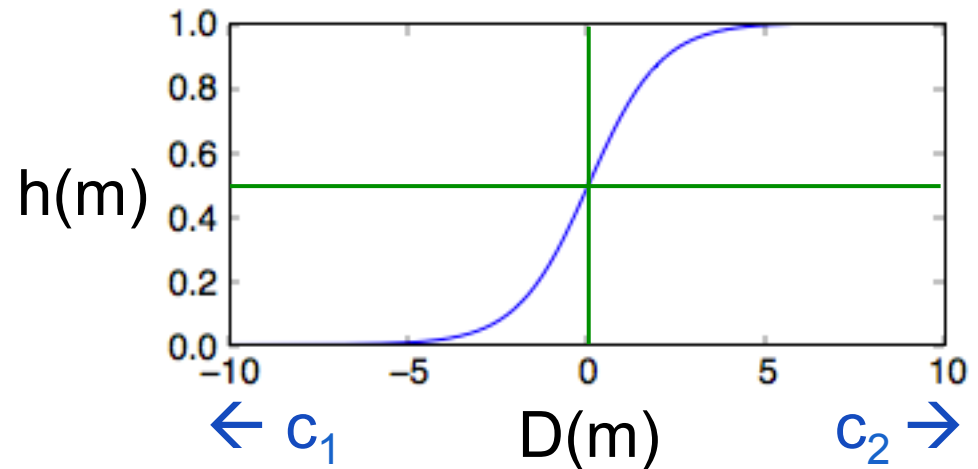
$$= -3 + 1 \cdot m_1 + 1 \cdot m_2$$

- Zuordnung zu Klasse c_1 , falls $D(m_1, m_2) < 0$
 - je kleiner $D(m_1, m_2)$, desto weiter von Grenze \rightarrow „sicherer“
- Zuordnung zu Klasse c_2 , falls $D(m_1, m_2) \geq 0$
 - je größer $D(m_1, m_2)$, desto weiter von Grenze \rightarrow „sicherer“

Logistische Regression

$$D(m) = w^T m \quad (m_0 = 1)$$

$$h(m) = \frac{1}{1 + e^{-D(m)}} = \frac{1}{1 + e^{-w^T m}}$$



- Hypothese $h(m)$ gibt Vertrauen in Klasse c_2 an:
 $h(m) = P(s=c_2 | m)$
- Zuordnung zu Klasse c_1 , falls $h(m) < 0.5$
- Zuordnung zu Klasse c_2 , falls $h(m) \geq 0.5$

Finden der Parameter w ?

- Trainingsmenge $\{ (m^{(1)}, y^{(1)}), (m^{(2)}, y^{(2)}), \dots, (m^{(M)}, y^{(M)}) \}$
- Kostenfunktion $Cost(h_w(m), y) = \begin{cases} -\log(h_w(m)) & , y = 1 \\ -\log(1 - h_w(m)) & , y = 0 \end{cases}$
 - falls $y = 1$ und $h_w(m)=1$: $Cost=0$
 - falls $y = 0$ und $h_w(m)=0$: $Cost=0$
 - falls $y = 1$ und $h_w(m) \rightarrow 0$: $Cost \rightarrow \infty$
 - falls $y = 0$ und $h_w(m) \rightarrow 1$: $Cost \rightarrow \infty$
- Gesamtkosten über Trainingsmenge

hohe Kosten, falls
Hypothese h_w
falsch liegt

$$J(w) = \frac{1}{M} \sum_{i=1}^M Cost(h_w(m^{(i)}), y^{(i)})$$



Gesamtkosten $J(w)$
minimieren!

Finden der Parameter w ?

- Kostenfunktion $Cost(h_w(m), y) = \begin{cases} -\log(h_w(m)) & , y = 1 \\ -\log(1 - h_w(m)) & , y = 0 \end{cases}$
 $= -y \log(h_w(m)) - (1 - y) \log(1 - h_w(m))$

- Gesamtkosten über Trainingsmenge

$$J(w) = -\frac{1}{M} \sum_{i=1}^M \left[y^{(i)} \log(h_w(m^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(m^{(i)})) \right]$$

- $-\text{grad}(J(w))$: Richtung des stärksten Abstiegs

$$\frac{\partial}{\partial w_j} J(w) = \dots = \frac{1}{M} \sum_{i=1}^M \left(h_w(m^{(i)}) - y^{(i)} \right) \cdot m_w^{(i)}$$

Gradientenabstieg zur Minimierung von J

$$J(w) = -\frac{1}{M} \sum_{i=1}^M \left[y^{(i)} \log(h_w(m^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(m^{(i)})) \right]$$

while (J(w) zu groß) {

 für alle j simultan {

$$w_j := w_j - \alpha \frac{\partial}{\partial w_j} J(w)$$

 }

}