2D Graphik: Klassifikation

Vorlesung „2D Graphik”

Andreas Butz, Otmar Hilliges

Freitag, 20. Januar 2006

Merkmale und Klassifikation

Klassifikation:

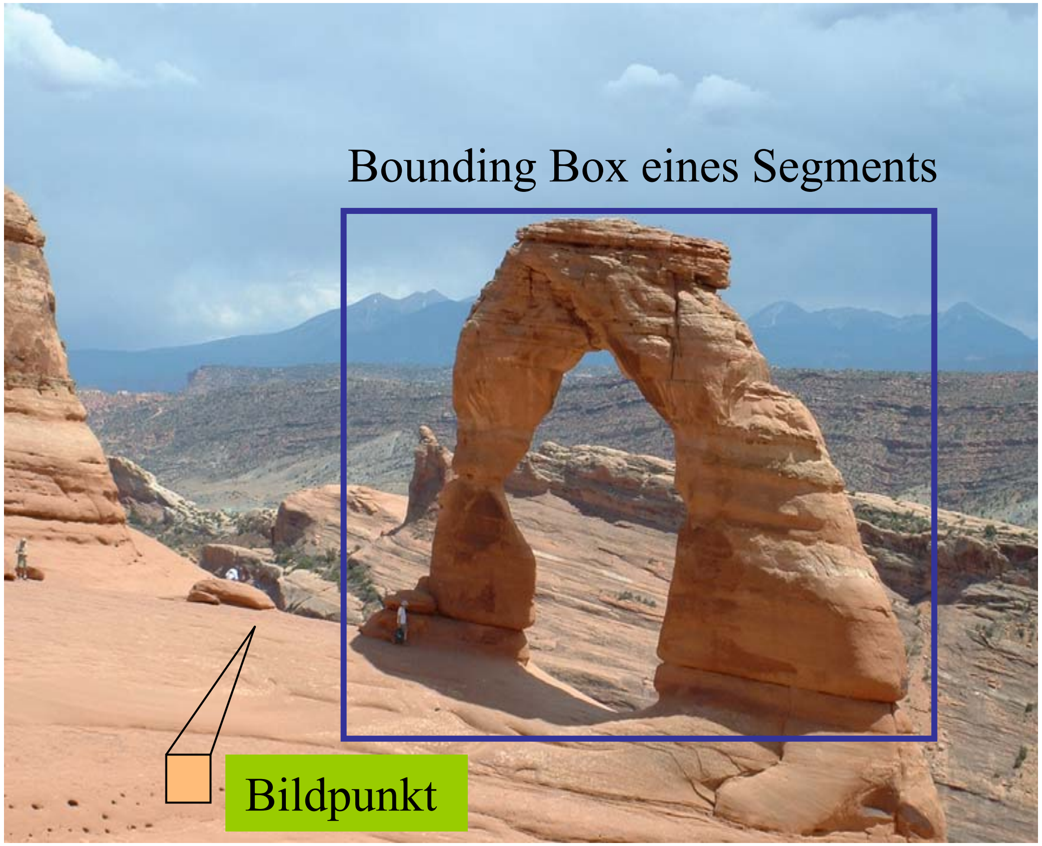
Zuordnung von Bedeutung zu

- Segmenten
- Segmentkombinationen
- Pixeln

Kriterium sind Merkmale, für die oft angenommen wird:

- durch Skalar repräsentierbar
- voneinander unabhängig

Merkmale können dann in einem Merkmalsvektor zusammengefasst werden.



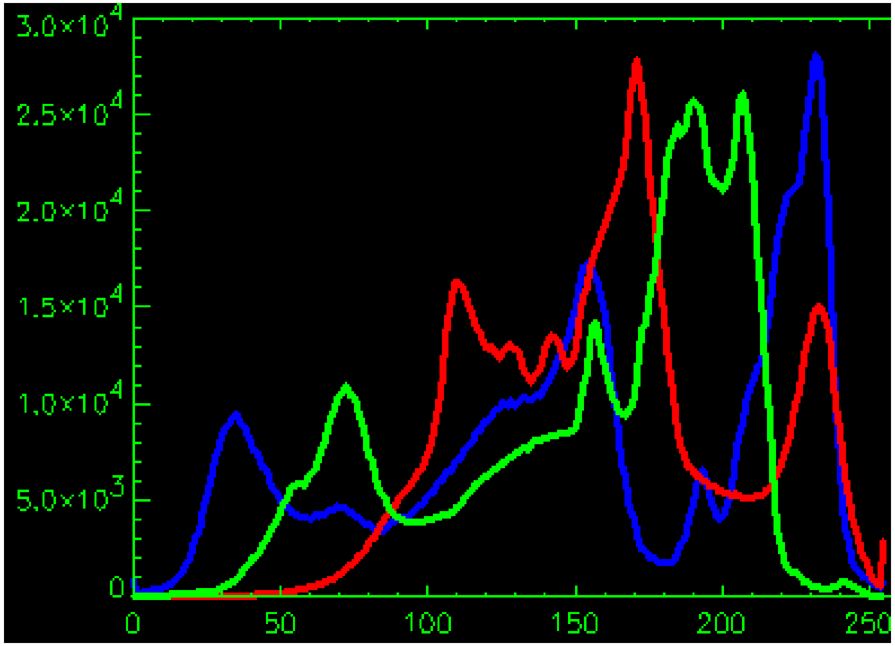
Einfache Pixel-Merkmale

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

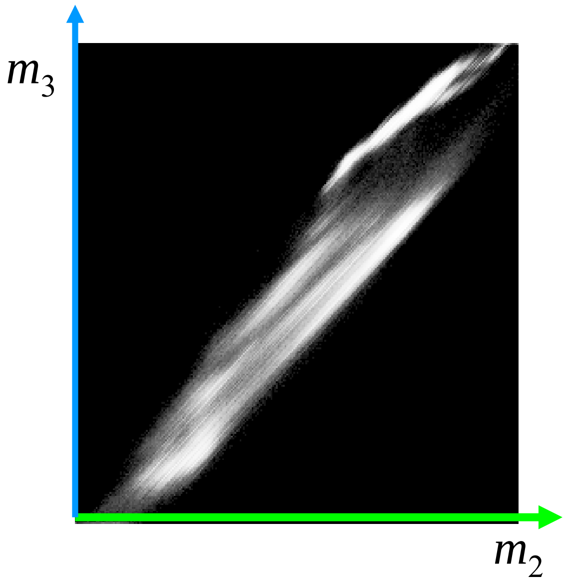
Farbwerte $\mathbf{m}(p) = [r(p)\ g(p)\ b(p)]$

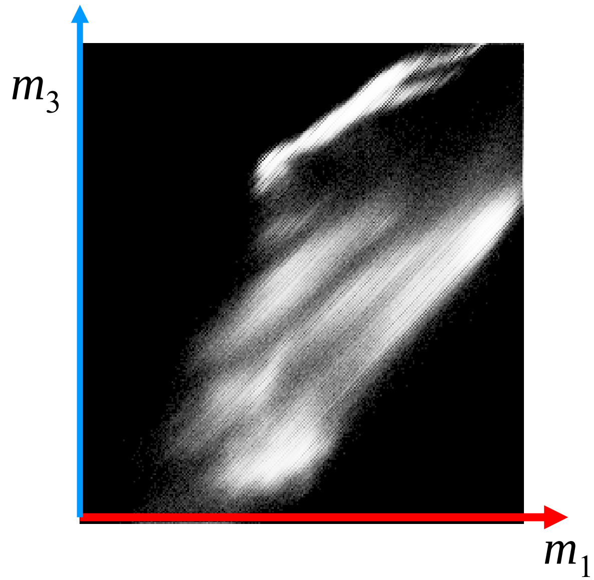
3-d Merkmalsraum der Farbauszüge

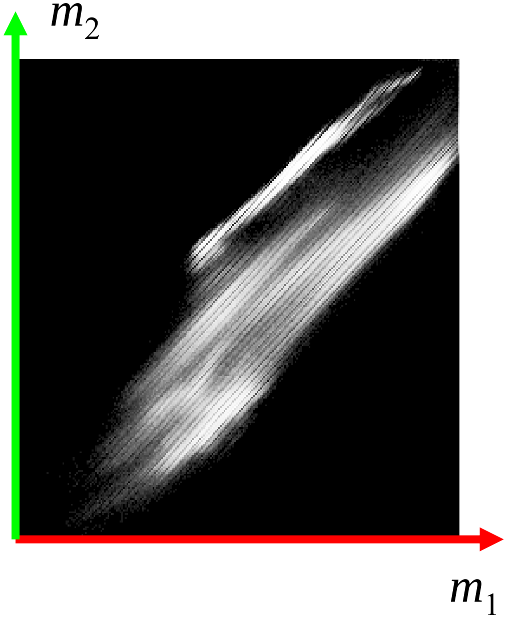
Histogramme



Mehrdimensionaler Merkmalsraum



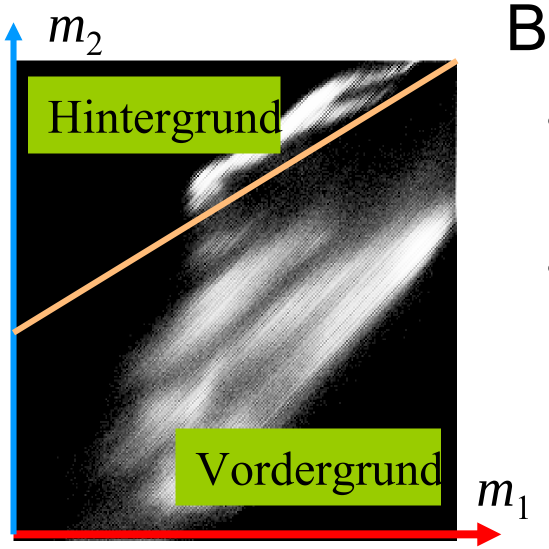




Projektionen des dreidimensionalen Merkmalsraums auf

- 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.

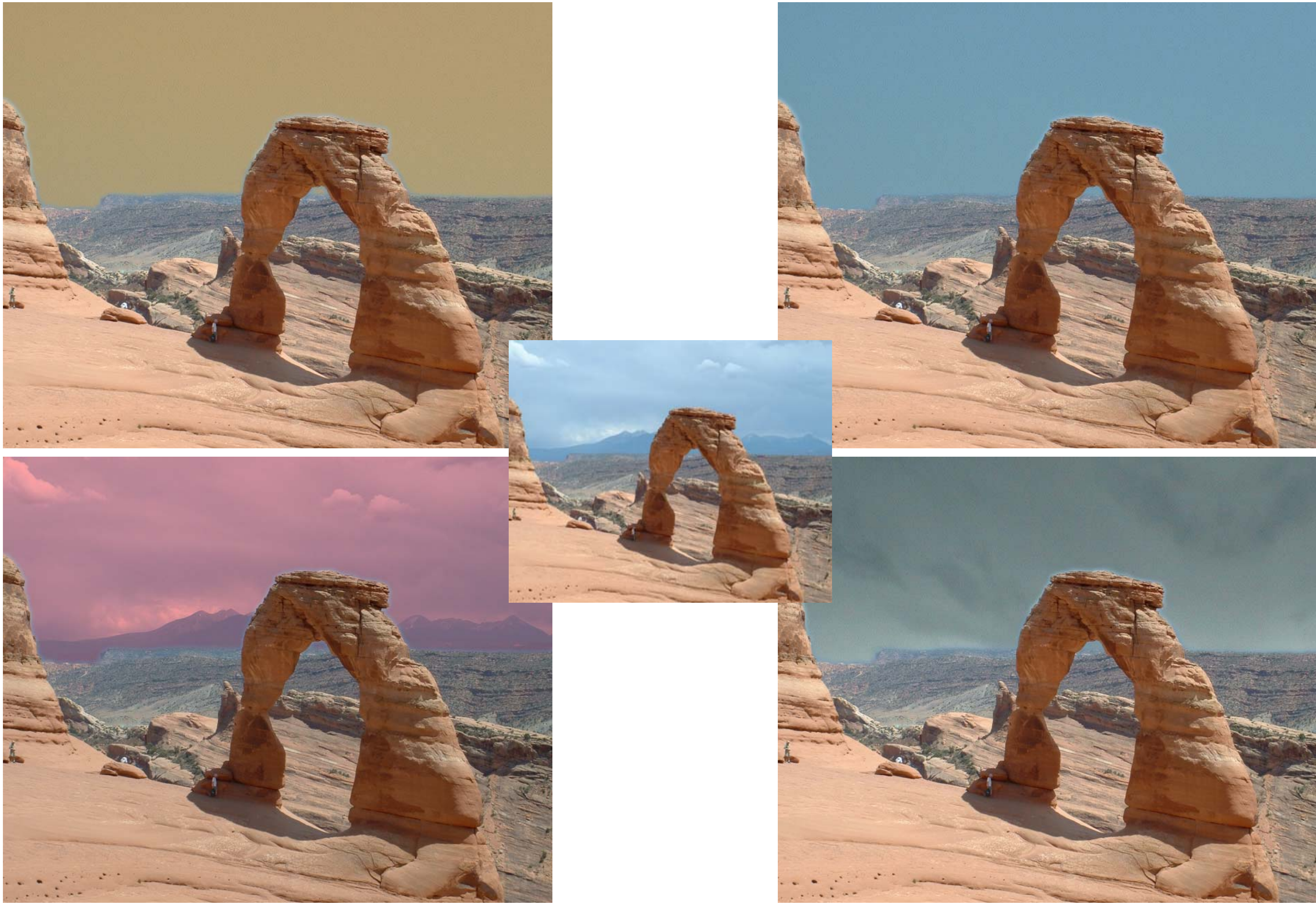
Beispiel:

- Entscheidungsgrenze ist eine Gerade $a \cdot m_1 + b \cdot m_2 + c = 0$.
- Pixelklassifikation:

  $p(i,j)$ = Vordergrund, falls $a \cdot m_1(p) + b \cdot m_2(p) + c < 0$,

  Hintergrund, sonst.

Klassifikations-ergebnis



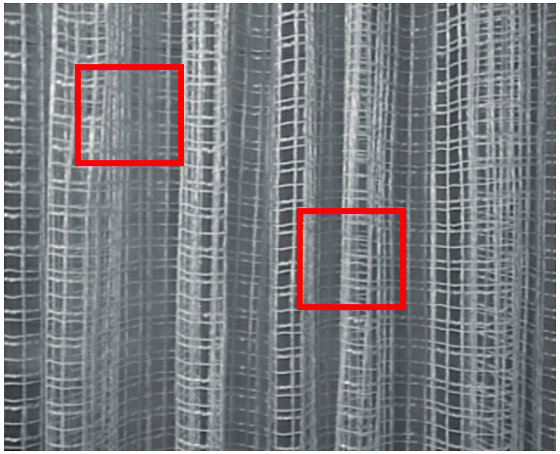
Texturmerkmale

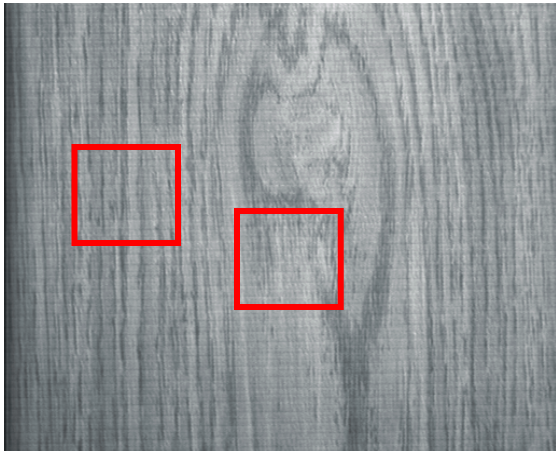
Objekte zeichnen sich nicht nur durch Hellig-keit und Farbe aus.

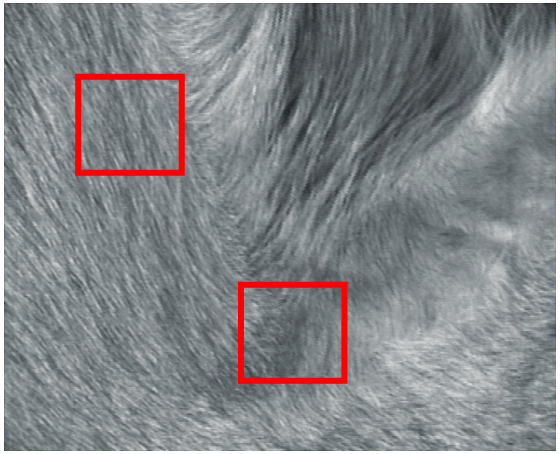
Textur: ein sich in einer bestimmten Region stochastisch oder deterministisch wiederholendes Muster (Invariante).

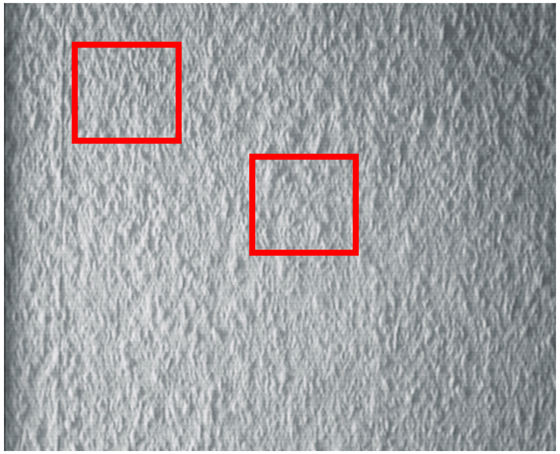
Texturmerkmal: messbare Charakteri-sierung der Invariante.

Textur beschreibt das **Objektinnere**.









Einfache Texturen - Varianz

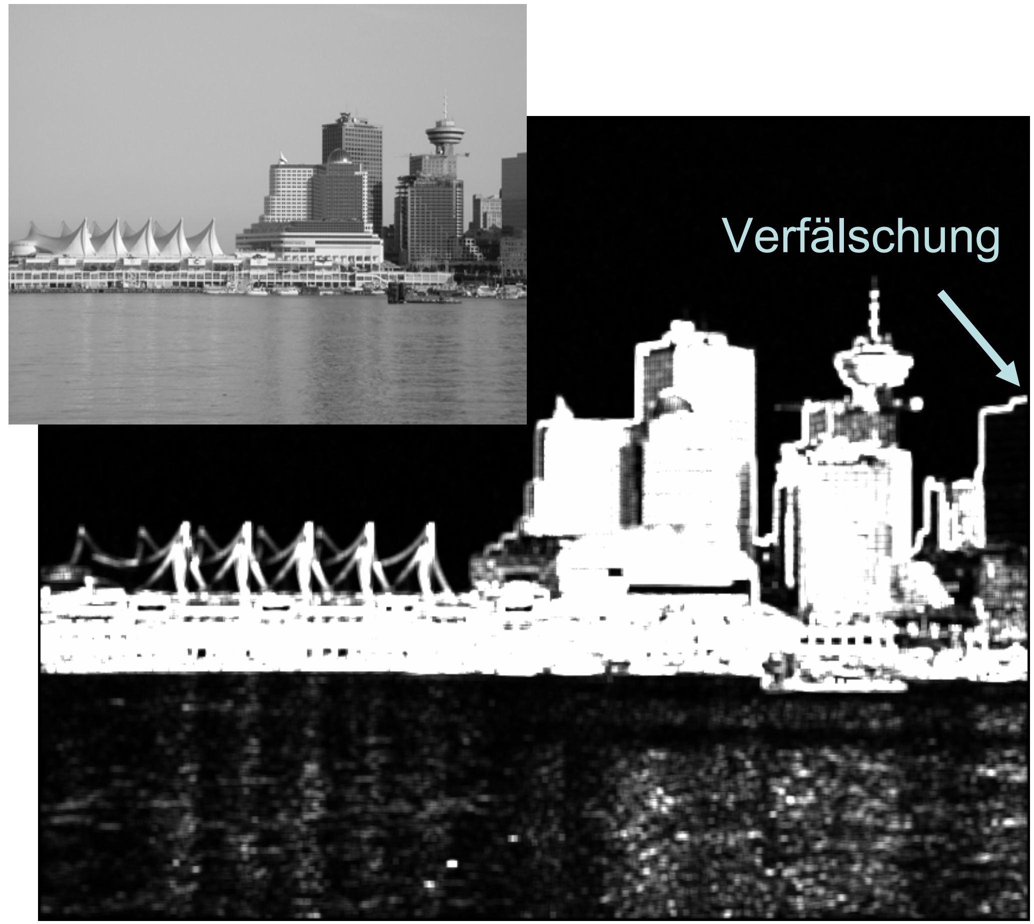
Lokale Varianz

$$V_n(i,j)=\frac{1}{n^2}\sum_{k=i-\frac{n}{2}}^{k=i+\frac{n}{2}}\sum_{l=j-\frac{n}{2}}^{l=j+\frac{n}{2}}\left(f(k,l)-\bar{f}_n(k,l)\right)^2,$$

$$\bar{f}_n(k,l)=\frac{1}{n^2}\sum_{k=i-\frac{n}{2}}^{k=i+\frac{n}{2}}\sum_{l=j-\frac{n}{2}}^{l=j+\frac{n}{2}}f(k,l)$$

Einfache Texturen können als Pixel-Merkmale verwendet werden.

aber: Merkmal ist an Kanten verfälscht! (warum ist das so?)

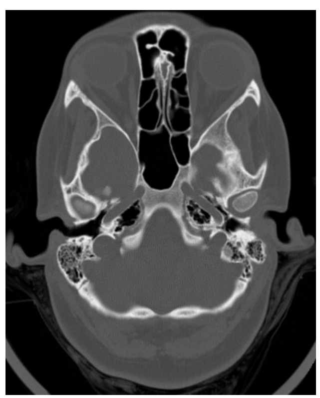


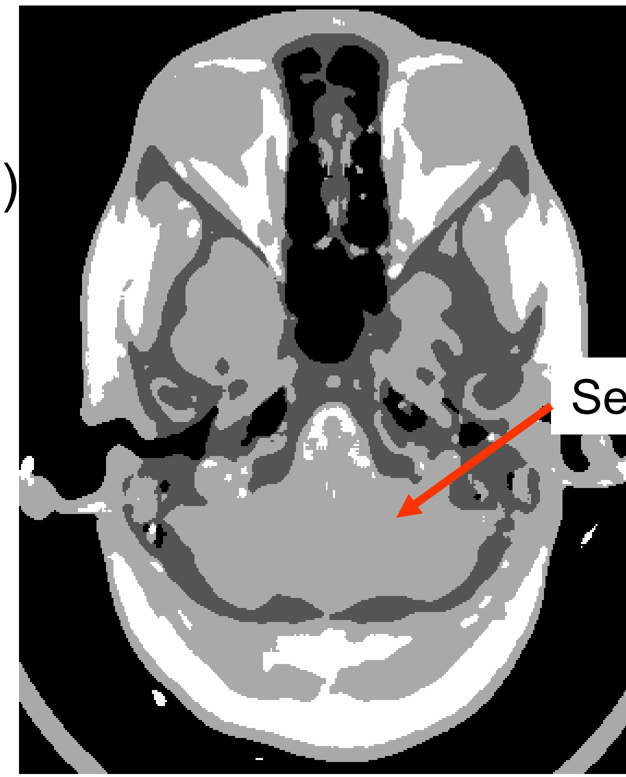
Merkmale von Regionen

Regionen: Resultat einer Segmentierung

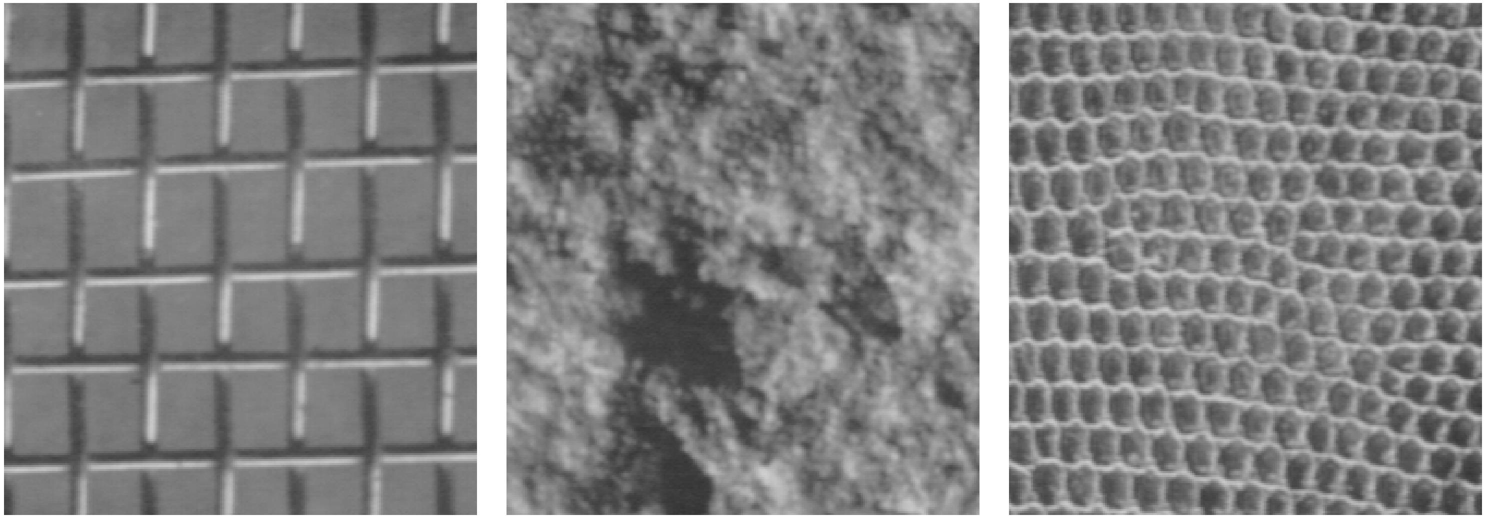
Regionenmerkmale:

- Geometrische Merkmale (Formmerkmale)
- Grauwertmerkmale (Texturmerkmale)





Texturmerkmale von Regionen



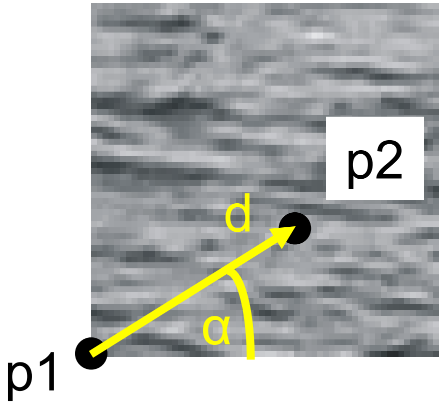
Mögliche Merkmale: - Richtungspräferenz

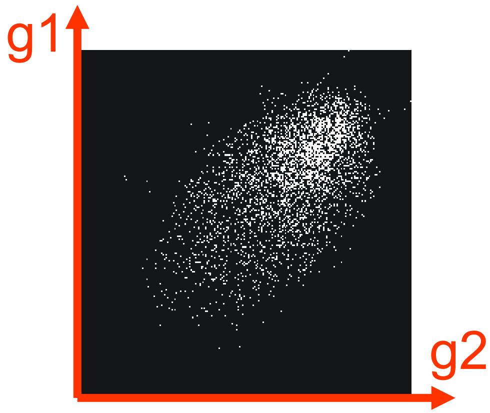
- Rauhigkeitsmaß

Merkmalsmaße z.B. die Haralick‘schen Texturmaße

Haralick´sche Texturmaße

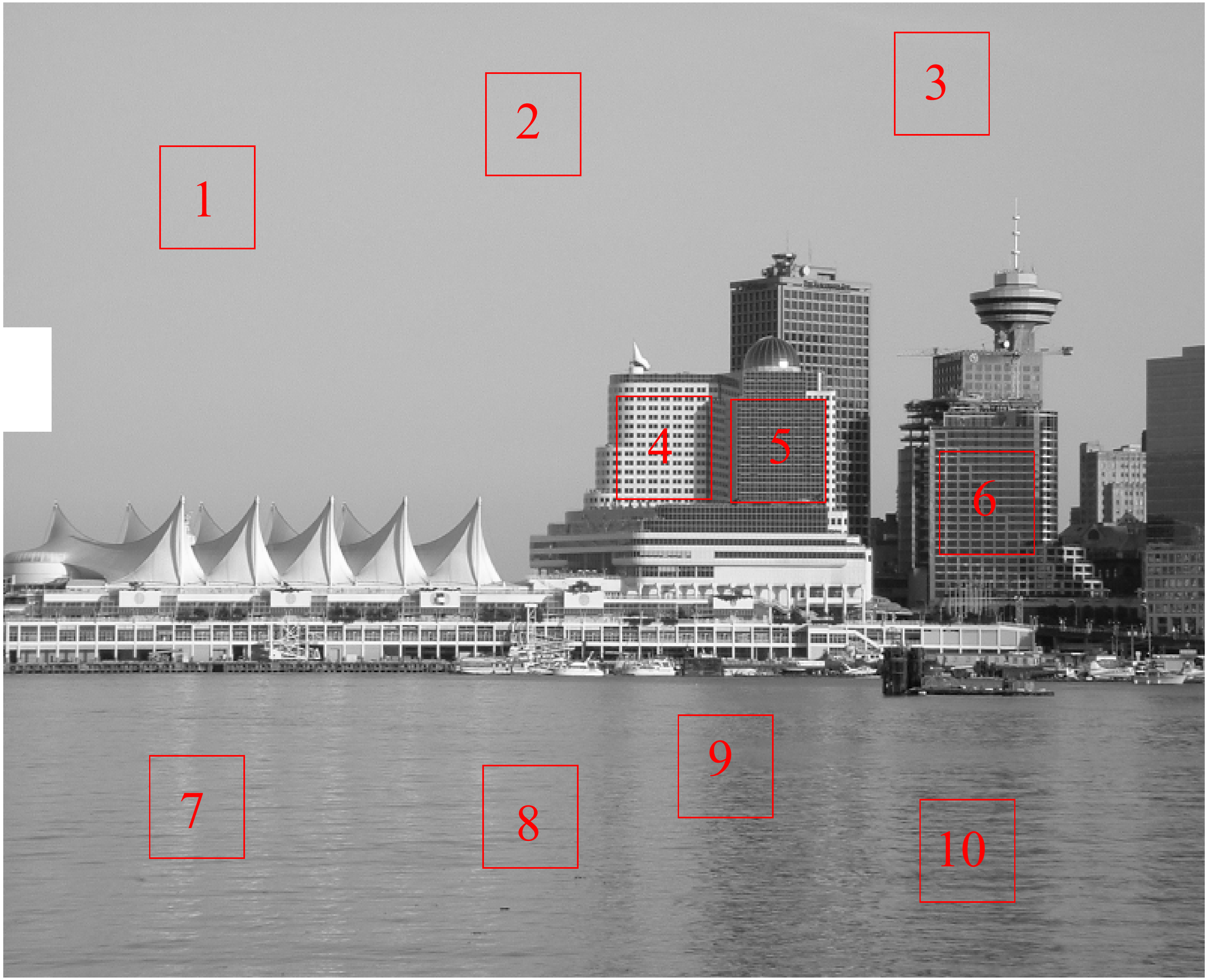
- **Co-Occurrence-Matrix =** Zweidim. Histogramm für Pixelpaare.
  Pixel p1 und p2 sind ein Paar, wenn sie Abstand *d* haben und auf einer Linie mit einem gegebenen Winkel α zur *x*-Achse liegen.
- Co-Occurence-Matrix repräsentiert die Korrelation zwischen Pixeln. (Wahrscheinlichkeit P, dass p1 und p2 Grauwerte g1 und g2 haben)
- Meist sind Pixel nicht über große Entfernungen korreliert, daher sind Werte *d*=1, *d*=2 üblich.
- Falls Korrelation über größere Entfernung vermutet wird, dann sollte eine Multiskalenstrategie angewendet werden.



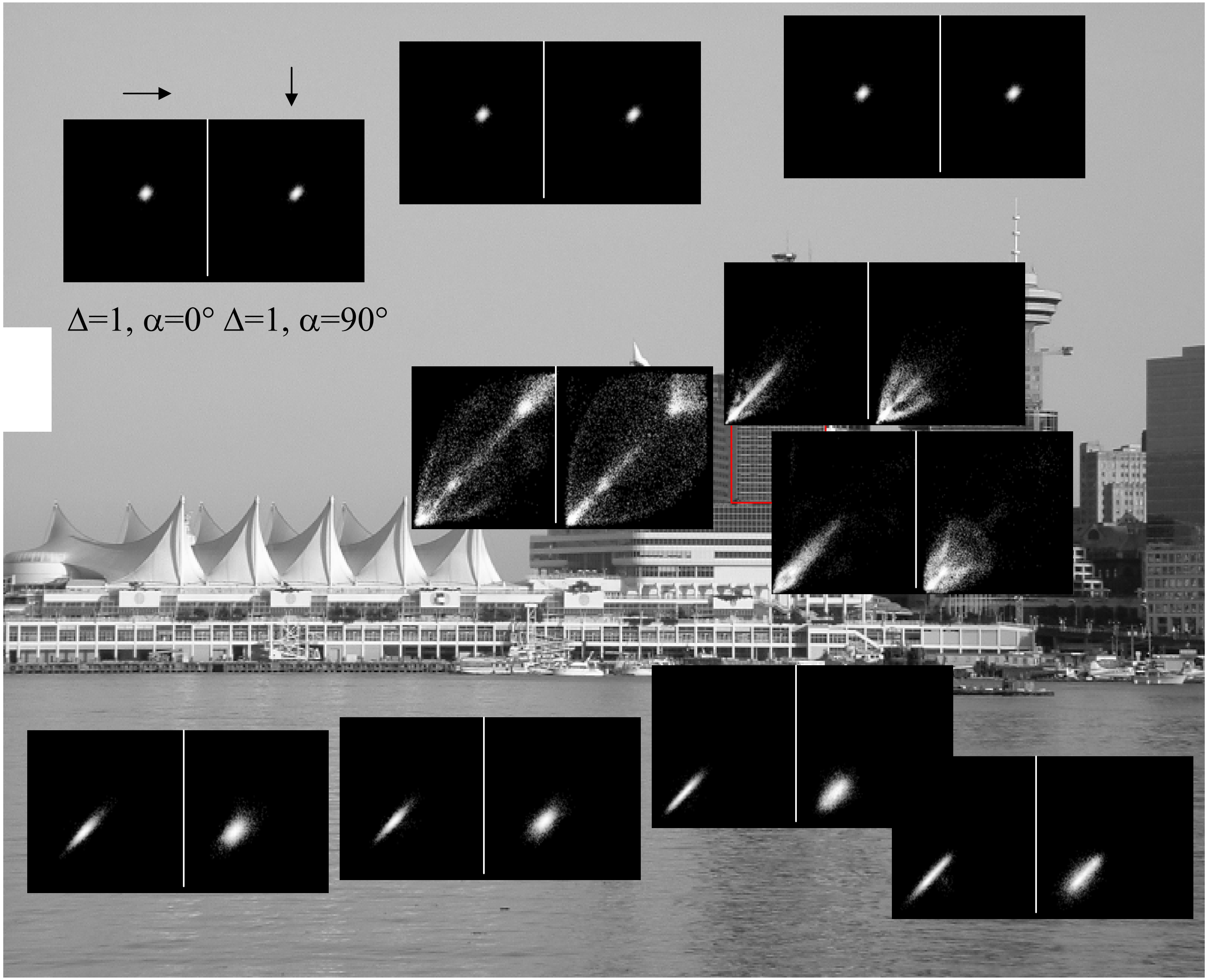


Robert Haralick

Beispiel



Beispiel



Haralick'sche Texturmaße

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

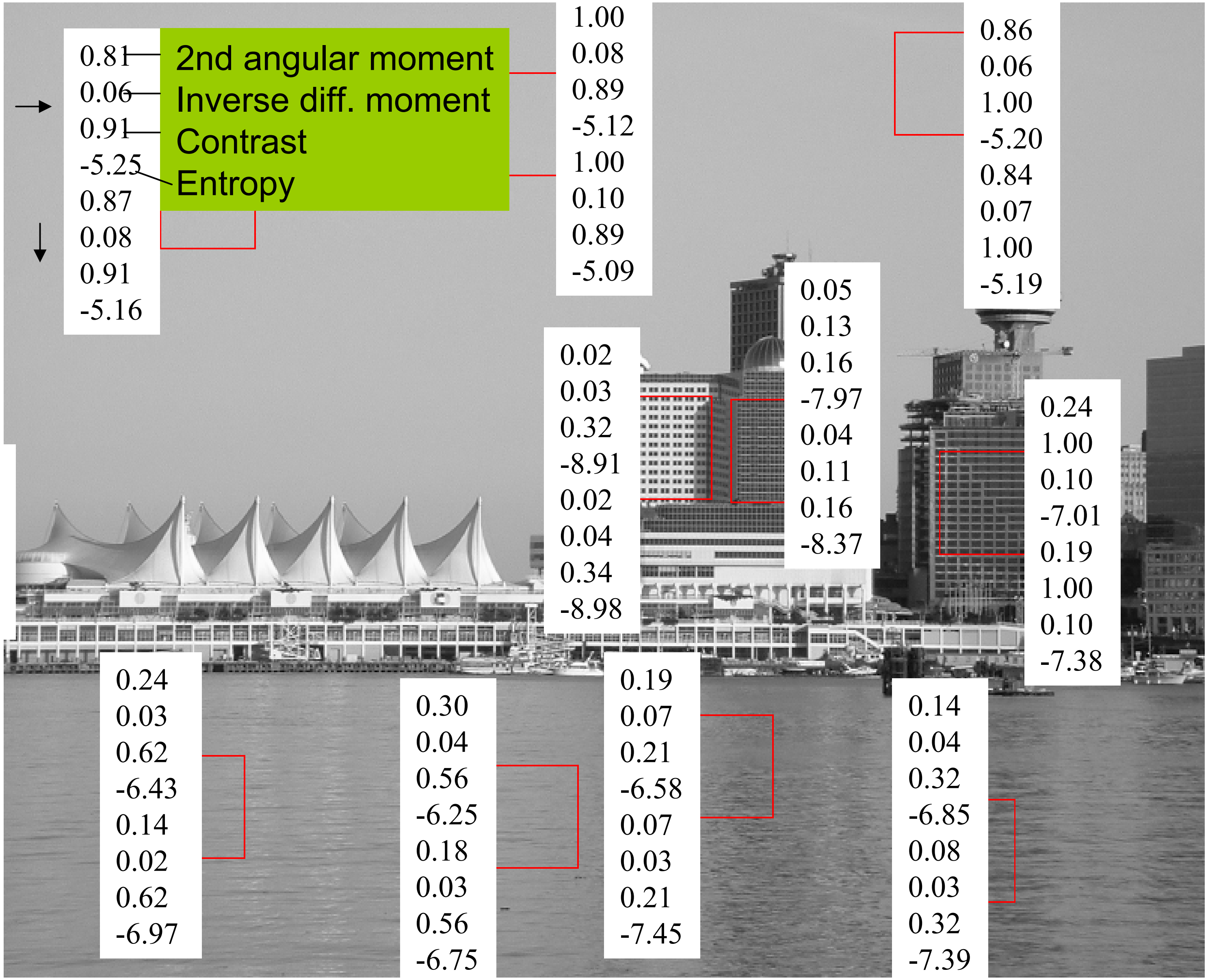
$$\sum_{g_1=0}^{K-1}\sum_{g_2=0}^{K-1} (g_1-g_2)^2 \cdot P_{\Delta,\alpha}(g_1,g_2)$$ Kontrast

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

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

- Lassen sich aus der Co-Occurrence-Matrix berechnen
- Liefern aussagekräftige Kennwerte für verschiedene Texturen

Textur maße


2nd angular moment
Inverse diff. moment
Contrast
Entropy
0.81 0.06 0.91 -5.25 0.87 0.08 0.91 -5.16
1.00 0.08 0.89 -5.12 1.00 0.10 0.89 -5.09
0.86 0.06 1.00 -5.20 0.84 0.07 1.00 -5.19
0.05 0.13 0.16 -7.97 0.04 0.11 0.16 -8.37
0.02 0.03 0.32 -8.91 0.02 0.04 0.34 -8.98
0.24 1.00 0.10 -7.01 0.19 1.00 0.10 -7.38
0.24 0.03 0.62 -6.43 0.14 0.02 0.62 -6.97
0.30 0.04 0.56 -6.25 0.18 0.03 0.56 -6.75
0.19 0.07 0.21 -6.58 0.07 0.03 0.21 -7.45
0.14 0.04 0.32 -6.85 0.08 0.03 0.32 -7.39

Formmerkmale

Formmerkmale beschreiben das Äußere von Segmenten
(Formmerkmale von Pixeln sind nicht sinnvoll ;-)

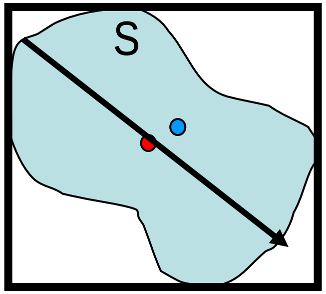
Allgemeine Merkmale:

- Mittelpunkt der Bounding Box
- Schwerpunkt *m* des Segments

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

Translationsinvariante Merkmale

- Bounding Box.
- Richtung der größten Ausdehnung und Ausdehnung in dieser Richtung.



Formmerkmale

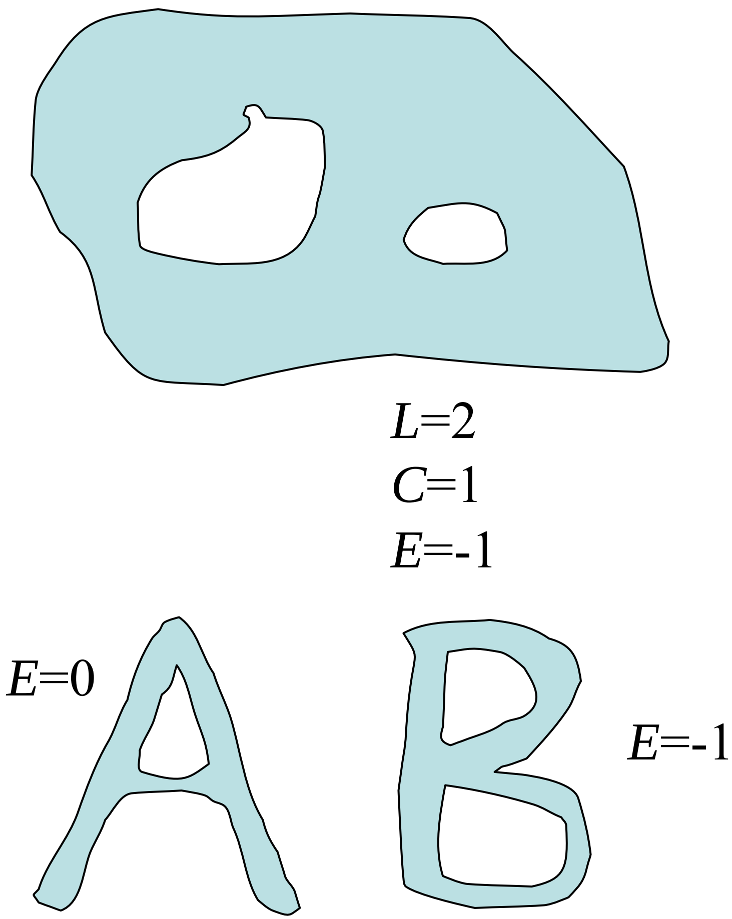
Translations- und rotationsinvariante Merkmale

- Eingeschlossene Fläche: F
- Länge des Randes: R
- Durchschnittliche Krümmungsänderung

Translations-, rotations- und skalierungsunabhängige Merkmale

- Kreisähnlichkeit = Verhältnis zwischen Fläche und Umfang):
  $(4\pi F)/R^2$

Topologische Formmerkmale



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, wieviele Merkmale?

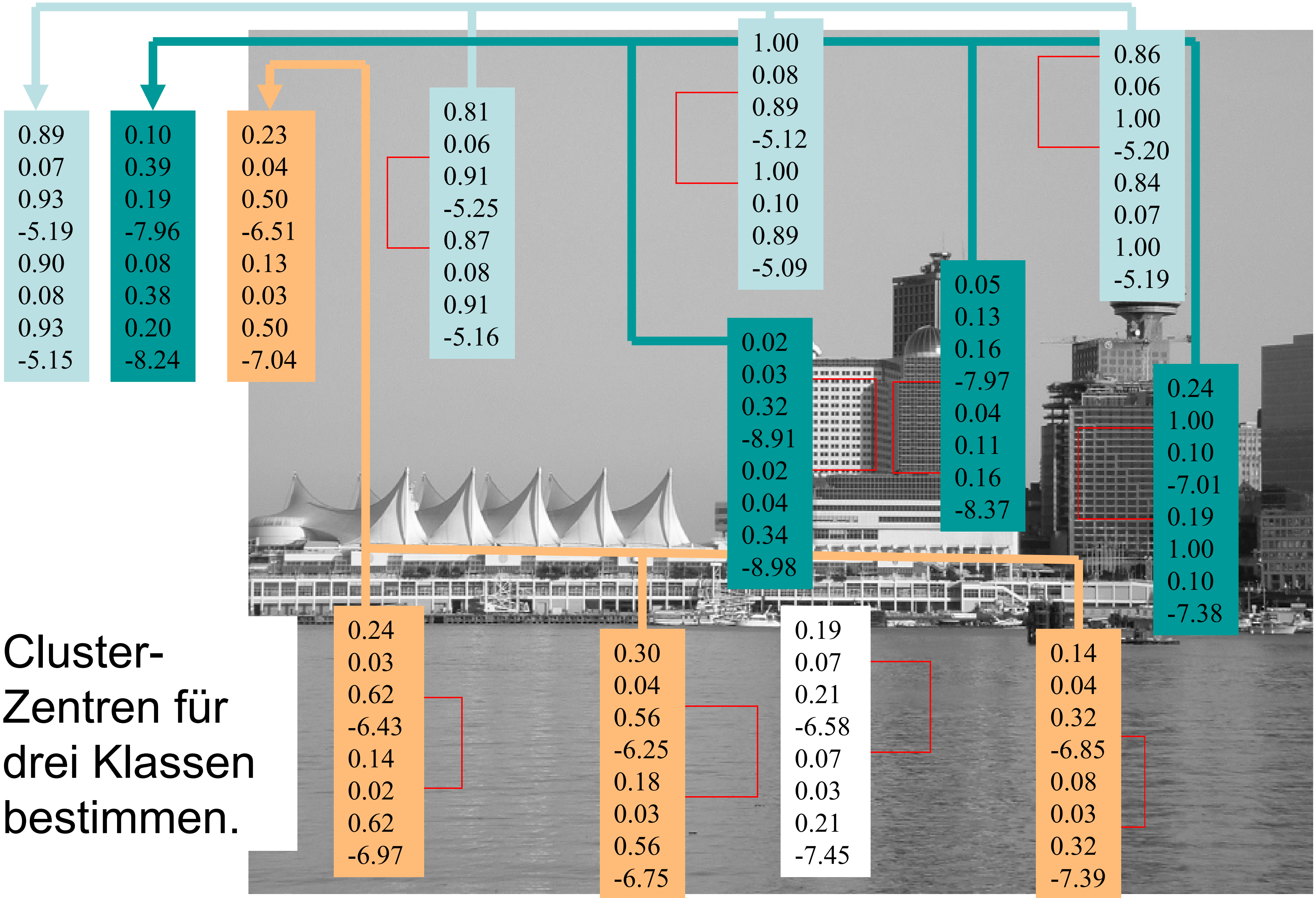
- Merkmale sollten aus den vermuteten Objekteigenschaften abgeleitet werden.
- Merkmale sollten voneinander unabhängig sein.
- Objekte der gleichen Klasse sollten im Merkmalsraum an ähnlichen Orten zu finden sei (Häufungspunkte der Klasse, Cluster).
- Die Unterscheidbarkeit nach Merkmalswert nimmt nicht ab, wenn die Dimension des Merkmalsraums erweitert wird.
- Die Suche nach Häufungspunkten im Merkmalsraum erfordert eine mit Dimension des Merkmalsraumes ansteigende Anzahl von Merkmalsträgern.
- Je weniger Merkmale zur Unterscheidung notwendig sind, desto effektiver wird die Entscheidungsfindung sein.

Klassifikation im *n*-dimensionalen Merkmalsraum

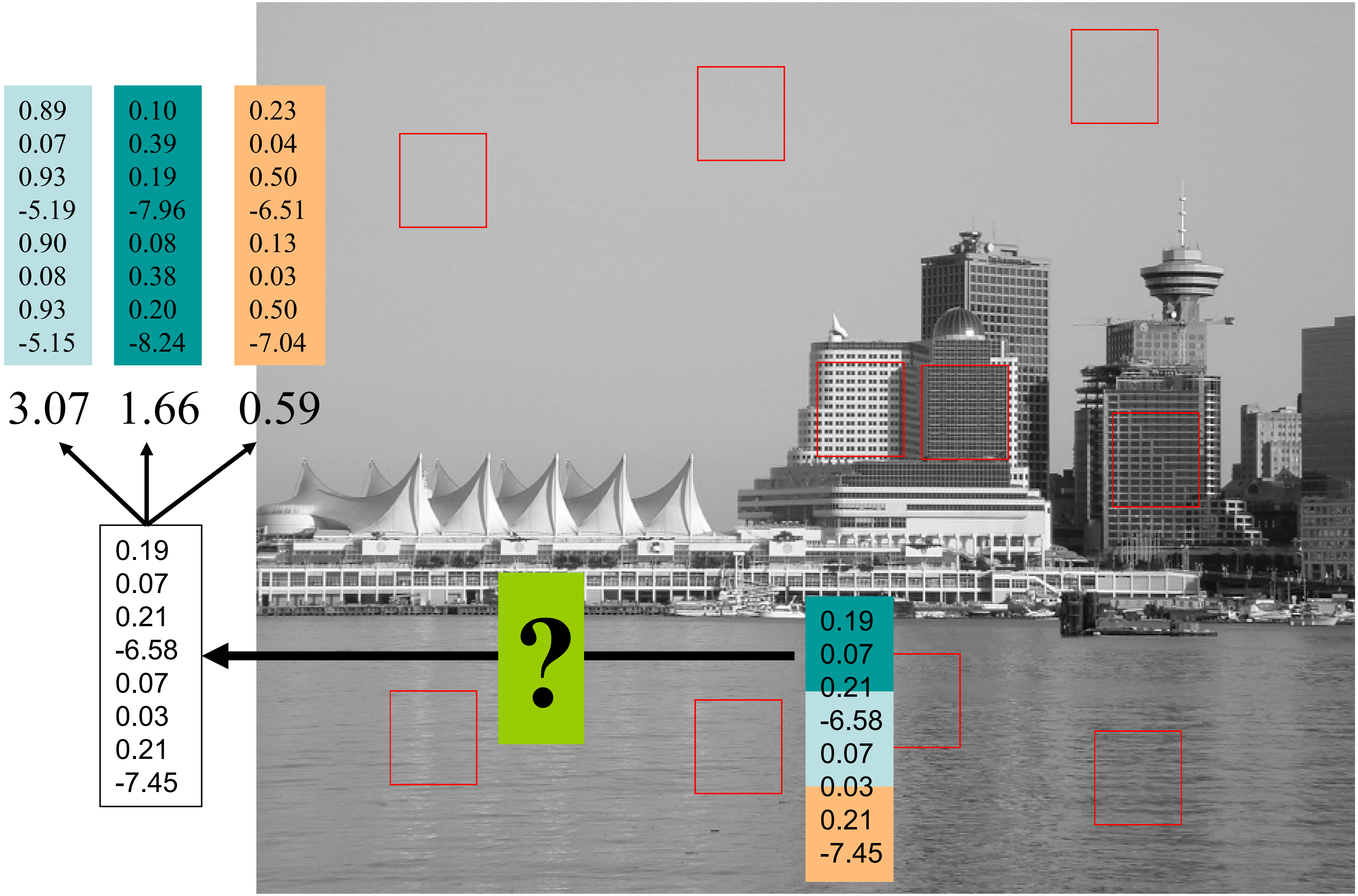
- Visuelle Suche nach Häufungspunkten oder Entscheidungsgrenzen ist nicht möglich.
- Automatische Verfahren:
  - Optimale Methoden: Minimierung der Wahrscheinlichkeit einer Fehlklassifikation.
  - Klassifikation nach Training: Entscheidung anhand von antrainierten Häufungspunkten oder Entscheidungsgrenzen.
  - 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 ist nicht ausreichend oder Stichproben sind nicht repräsentativ.
  - Unterschiedliche Skalierungen für die Merkmale.



Cluster-Zentren für drei Klassen bestimmen.


0.89
0.07
0.93
-5.19
0.90
0.08
0.93
-5.15
0.10
0.39
0.19
-7.96
0.08
0.38
0.20
-8.24
0.23
0.04
0.50
-6.51
0.13
0.03
0.50
-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 ist so gewählt, dass die Anzahl der Fehlklassifikationen minimal ist.

Lösungsansatz:

Merkmalsträger als Stichproben von statistischen Verteilungen der einzelnen Klassen im Merkmalsraum

Kriterium minimiert Wahrscheinlichkeit einer Fehlentscheidung

A-Posteriori-Wahrscheinlichkeit

- Mit welcher Wahrscheinlichkeit gehört ein Segment *s* mit Merkmalen *m* der Klasse *c* an:

$$P\left(s = c_i \middle| \vec{m}(s)\right)$$

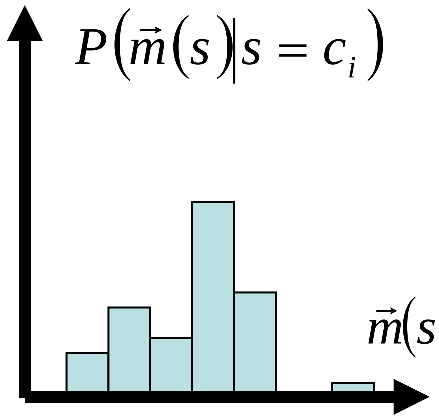
- Klassifikation: wähle diejenige Klasse, für die die Wahrscheinlichkeit am größten ist.
- Die Anzahl der Fehlentscheidungen ist minimal, da immer(!) die Lösung mit höchster Wahrscheinlichkeit gewählt wird.
- Problem: A-Posteriori-Wahrscheinlichkeit ist schwer zu bestimmen:
  - für jede Merkmalskombination müssen klassenabhängige Wahrscheinlichkeiten bestimmt werden.

Merkmalsverteilungsfunktion

- Leichter ist zu bestimmen, wie sich für jede Klasse separat die Merkmalswerte verteilen:

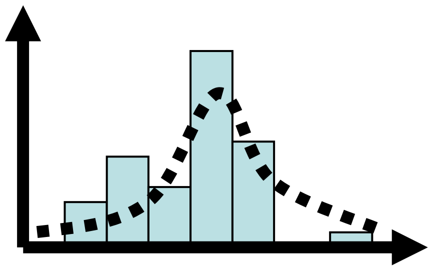
$$P\left(\vec{m}(s) \middle| s = c_i\right)$$

- 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.

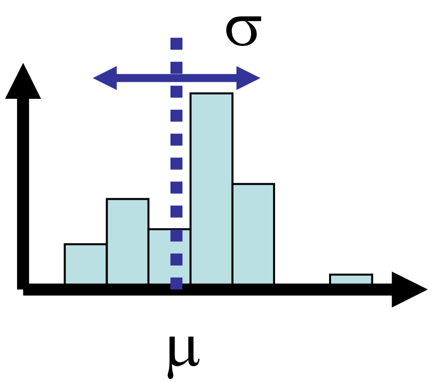


Schätzung der Merkmalsverteilungsfunktion

- 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

- Merkmalsverteilungsfunktion wird mit der A-Priori-Wahrscheinlichkeit gewichtet.

$$P(s = c_i)$$

- A-Priori-Wahrscheinlichkeit: Klassenwahrscheinlichkeit ohne Kenntnis der Merkmale
- Beispiel:
  - Segmente von Personen allgemein:
    $P(s=\text{Mann})=0.5$, $P(s=\text{Frau})=0.5$
  - In Informatikfakultät:
    $P(s=\text{Mann})=0.8$, $P(s=\text{Frau})=0.2$

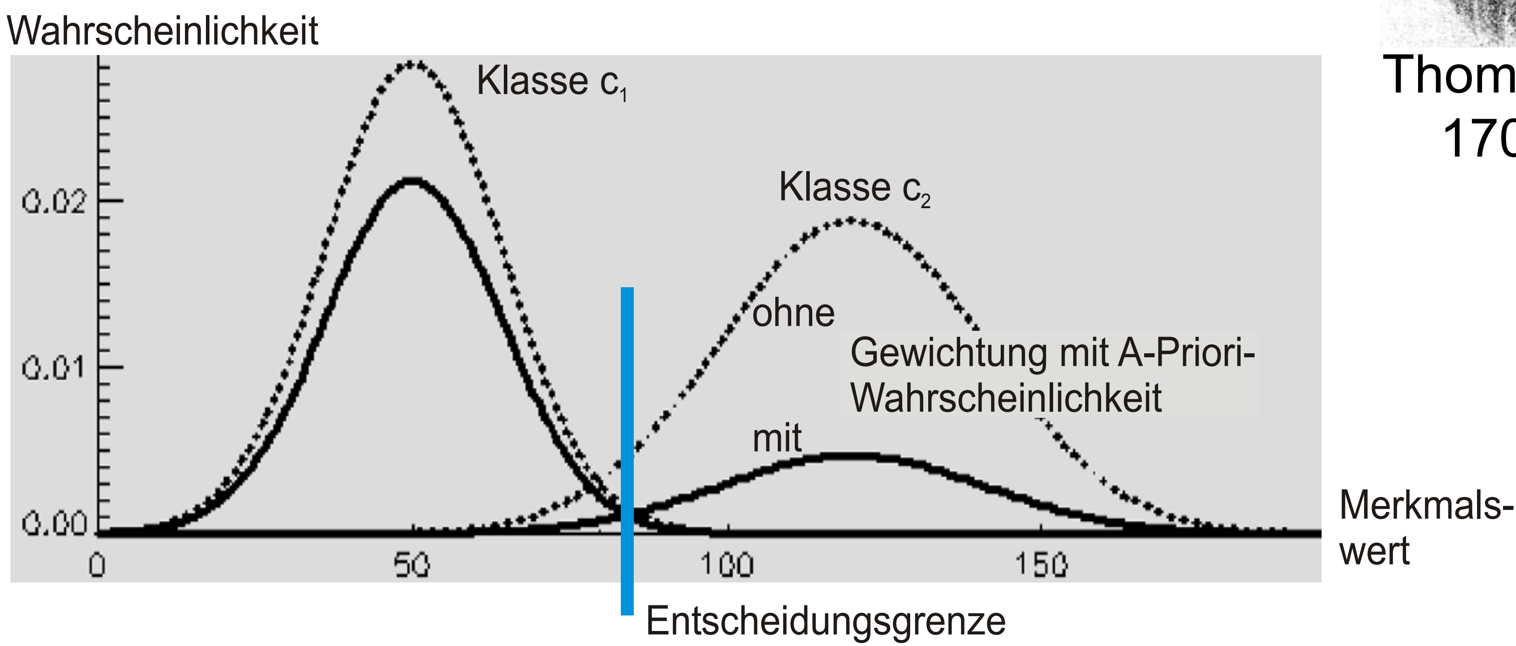
Schätzung der A-Priori-Wahrscheinlichkeit

- Bestimmung aus vorhandenem Wissen („es gibt etwa gleich viele Männer und Frauen")
- Bestimmung aus der Trainingsmenge
  - Trainingsmenge $T=T_1\cup T_2\cup\ldots$ ist repräsentativ für alle Klassen
  - Anteil der Elemente von T_i an T entspricht der A-Priori-Wahrscheinlichkeit, dass ein beliebig ausgewähltes Element der Klasse c_i angehört.

Satz von Bayes

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

Thomas Bayes, 1700-1761



$$P(s = c_i | \vec{m}(s)) = \frac{P(\vec{m}(s) | s = c_i) P(s = c_i)}{\sum_{k=0}^{K-1} P(\vec{m}(s) | s = c_k)}$$

Normierung, ergibt P(m(s))

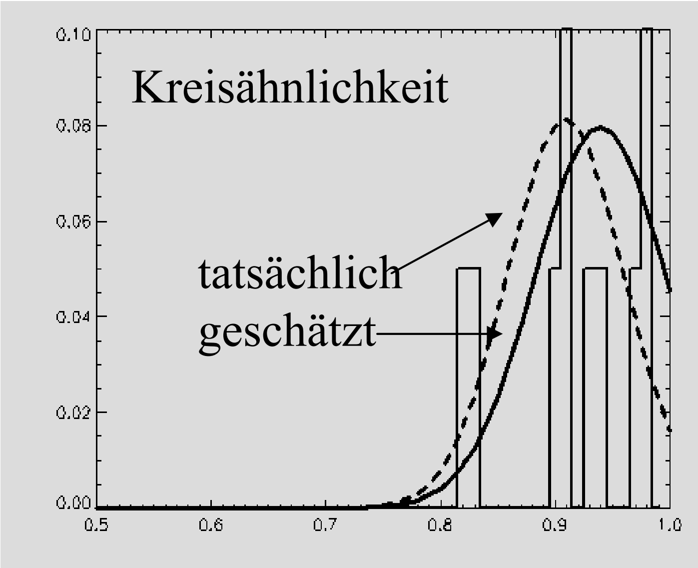
Beispiel

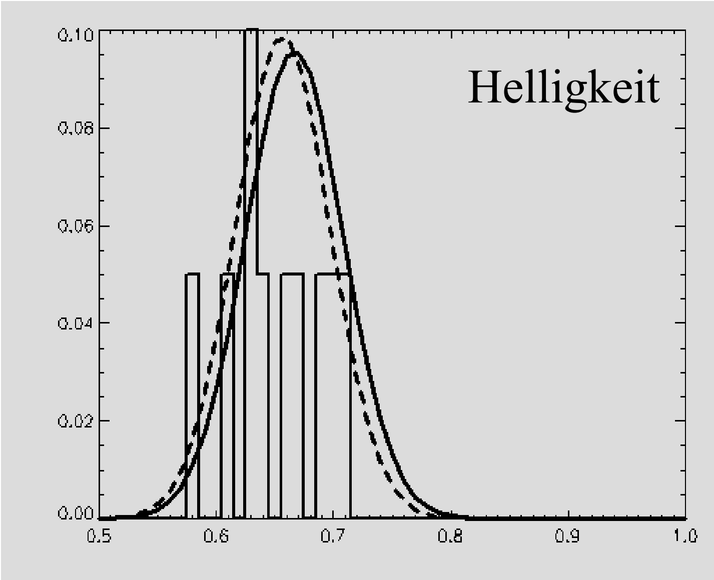
- Gegeben sei eine Menge von Bildern
- Jedes Bild enthält eine Frucht (*a*:Apfel oder *b*:Birne), die segmentiert wurde.
- Merkmale m_1 und m_2 sind m_1=Kreisähnlichkeit (Äpfel sind runder als Birnen) und m_2=Helligkeit (Äpfel sind dunkler als Birnen).
- Beide Merkmale sind unabhängig voneinander
- Es 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)

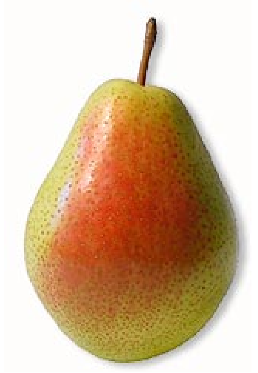
Merkmale in der Trainingsmenge

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



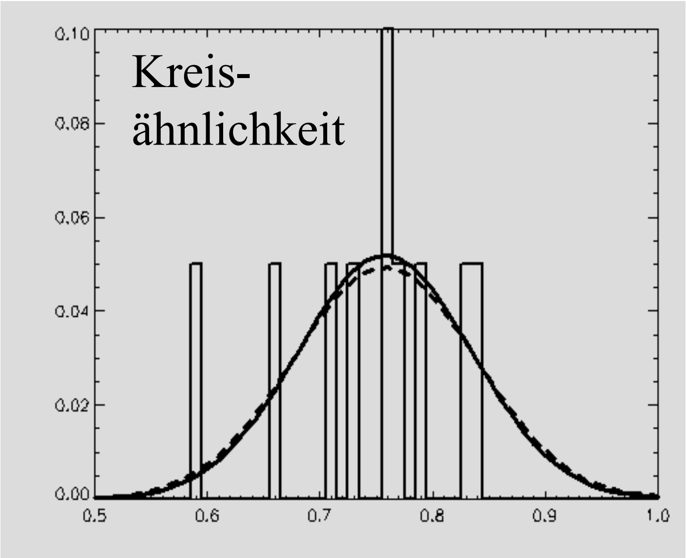


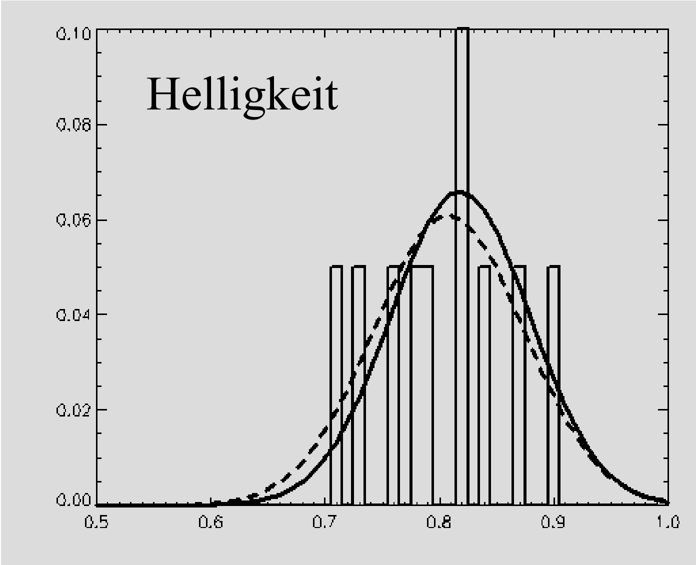
Merkmalsverteilungsfunktionen (Birnen)



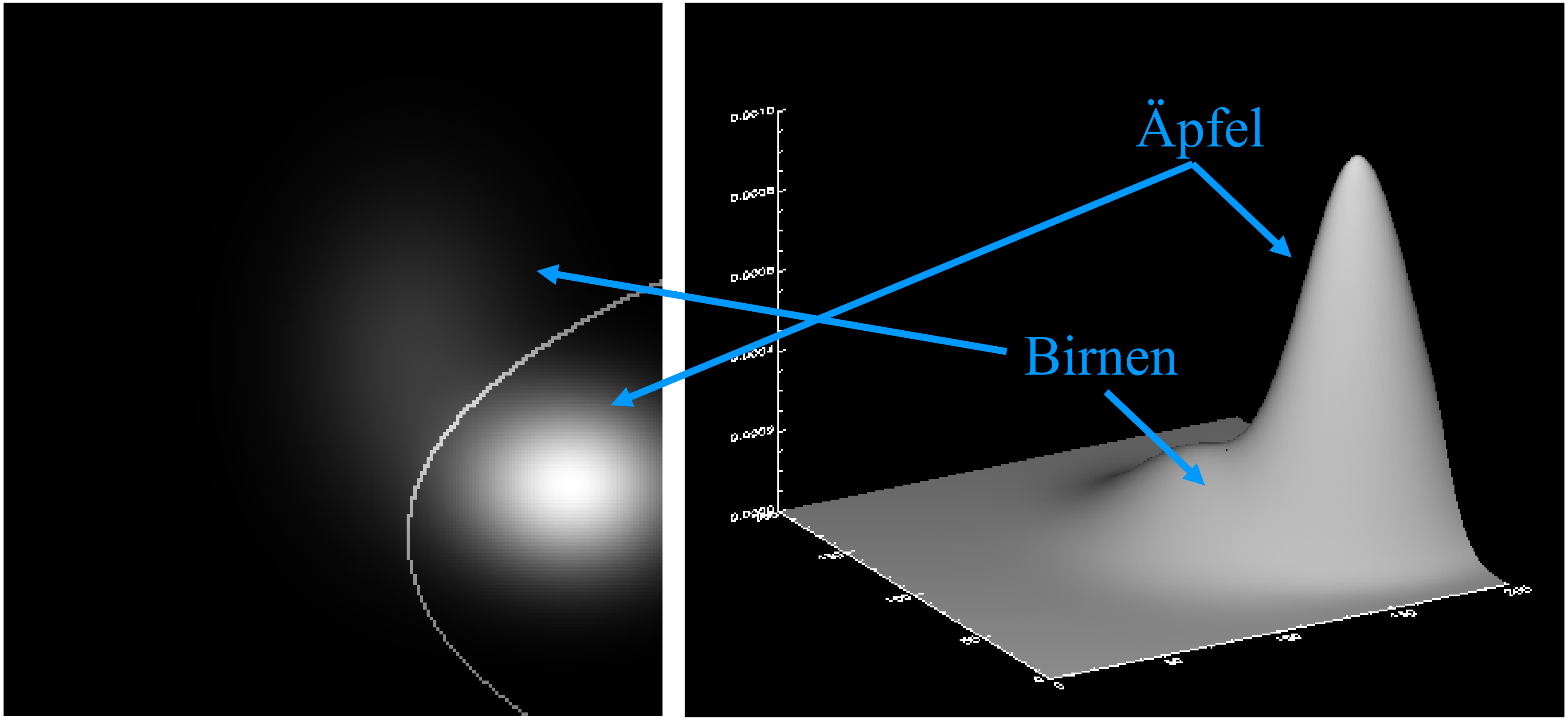
Merkmale in der Trainingsmenge

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





Überlagerte Klassenwahrscheinlichkeiten



Klassifikation

Aus den Parametern der Testdatenmengen berechnete Verteilungsfunktionen:

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

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

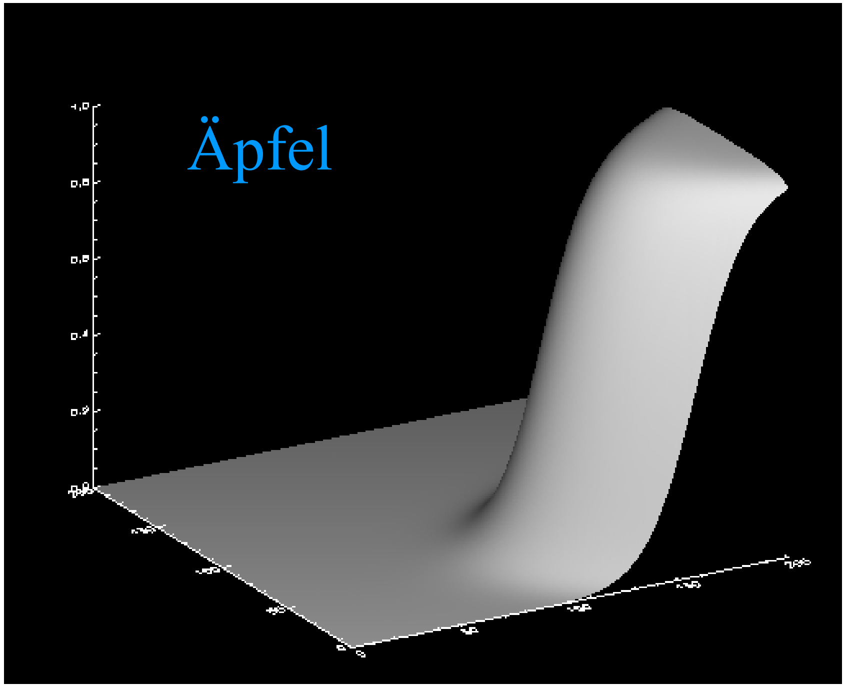
Klassifikation eines unbekannten Segments mit Werten (0.83, 0.72):

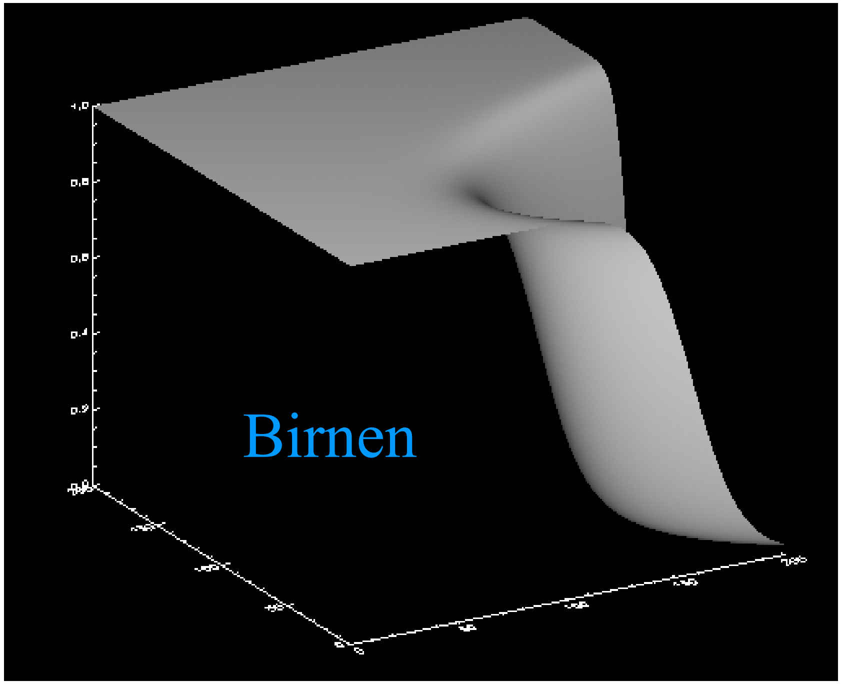
$$P\left(a \middle| \begin{pmatrix}0.83 & 0.72\end{pmatrix}\right) = \frac{P\left(\begin{pmatrix}0.83 & 0.72\end{pmatrix} \middle| a\right) \cdot 0.67}{P\left(\begin{pmatrix}0.83 & 0.72\end{pmatrix} \middle| a\right) \cdot 0.67 + P\left(\begin{pmatrix}0.83 & 0.72\end{pmatrix} \middle| b\right) \cdot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

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

Tatsächliche Wahrscheinlichkeiten:
P(a|(0.83 0.72))=0.60 und P(b|(0.83 0.72))=0.40 !

A-Posteriori-Wahrscheinlichkeiten





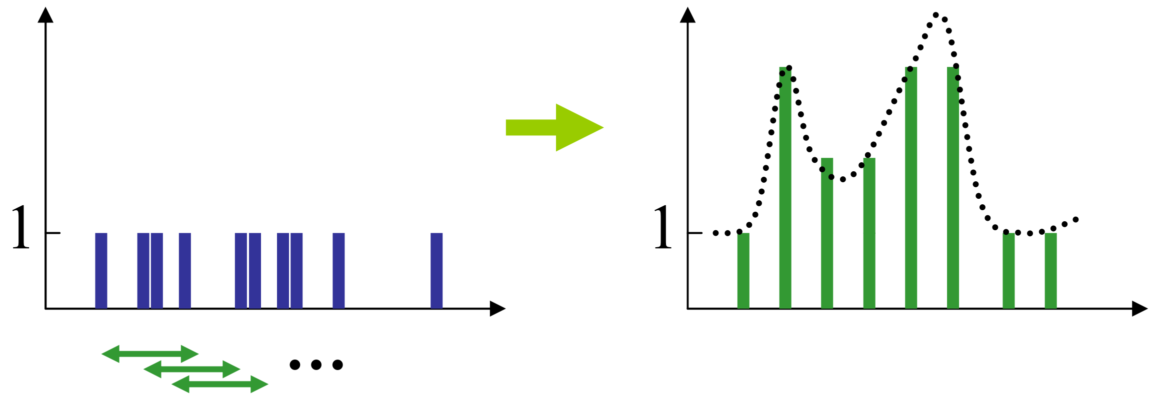
Direkte Schätzung der A-Posteriori-Wahrscheinlichkeiten (d.h. direkte Zuordnung zu Klassen)

- Beobachtung:
  - An den meisten Orten im Merkmalsraum ist der 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.

Schätzung eines Funktionswerts

Reeller Definitionsbereich und endliche Anzahl von Stichproben.

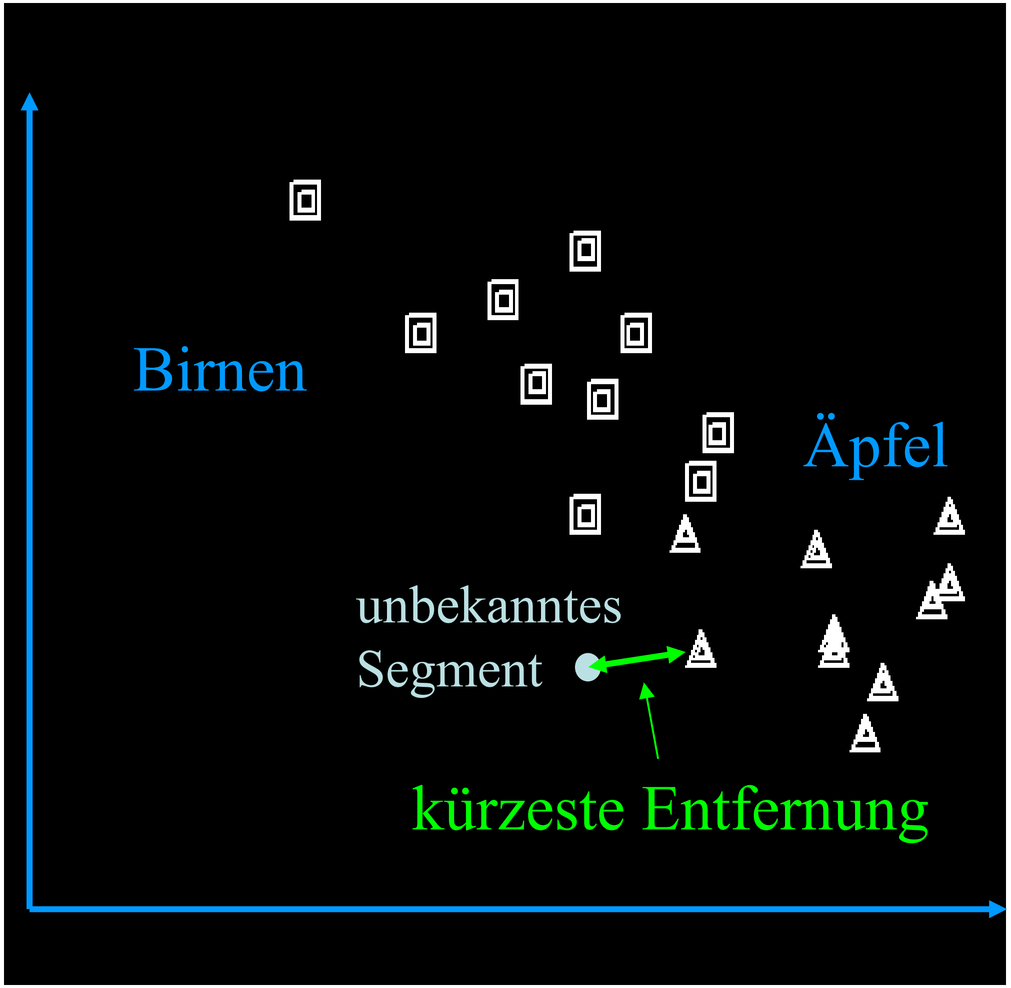
Schätzung erfolgt durch Mittelung über einen Unterbereich.



Problem: Schätzung wird unzuverlässiger, je weniger Stichproben in den Unterbereich fallen.

Lösung: Bereich von der lokalen Dichte der Stichproben abhängig machen
► k-Nearest-Neighbour-Classification

Single Nearest Neighbour

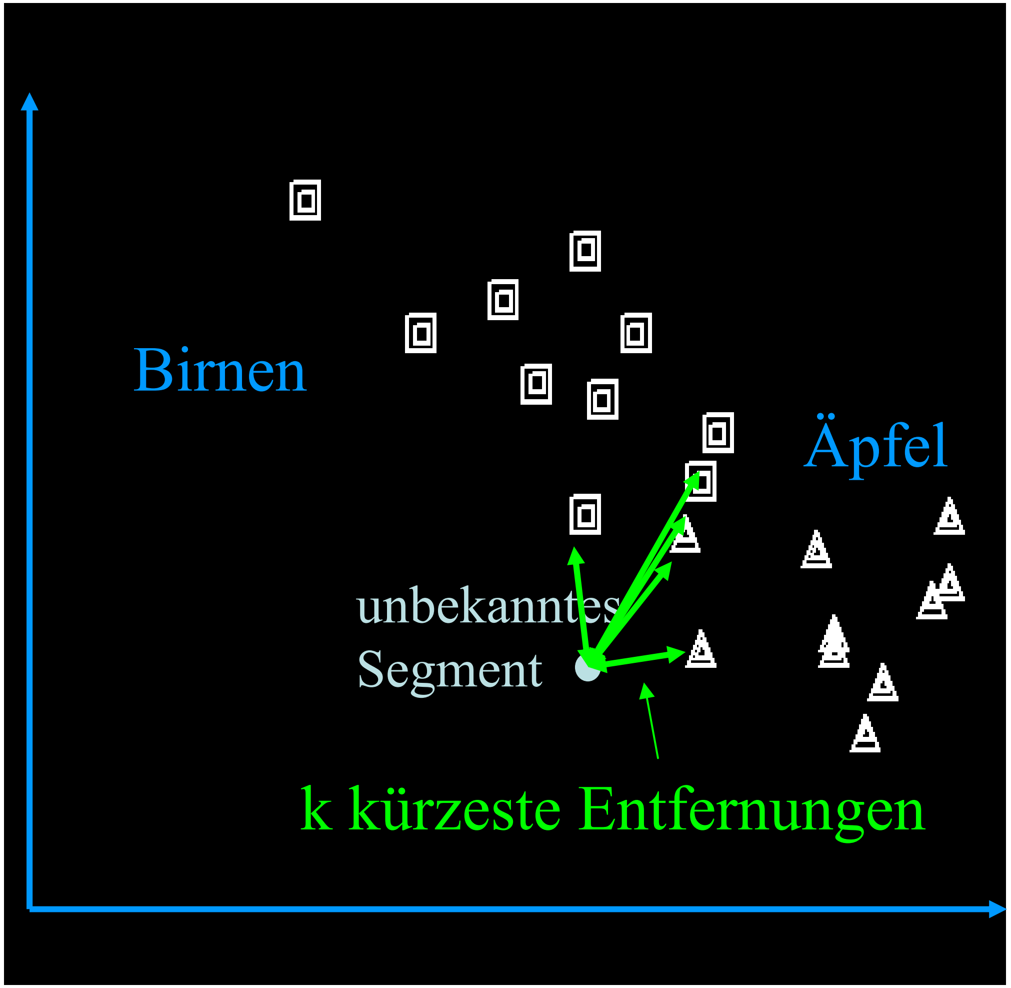


Für die zu bestimmende A-posteriori-Wahrscheinlichkeit wird die Umgebung **gerade so groß** gemacht, dass sie **eine klassifizierte Stichprobe** c_k umfasst.

Schätzung von $P(s=c_i|m(s))$:

P=1.0 für $c_i=c_k$
P=0.0 sonst

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

- Mit steigendem *k* wird die Schätzung der durchschnittlichen A-Posteriori-Wahrscheinlichkeit in der Umgebung ε, in der die Stichproben liegen, besser.
- Je dichter die Stichproben liegen, desto kleiner ist ε für ein gegebenes *k*.
- Falls *k* und die Stichprobendichte gegen unendlich gehen, nähert sich die Qualtität 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

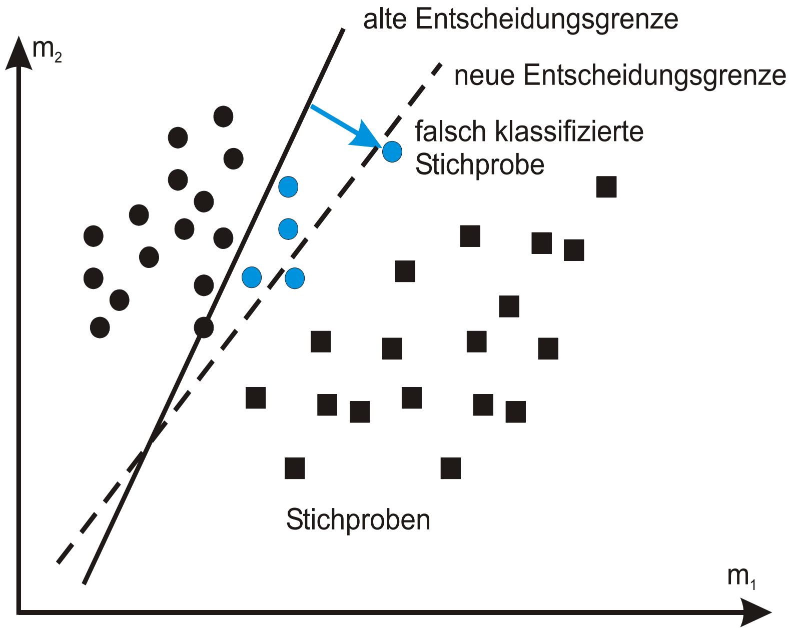
- Zwei-Klassen-Problem (Klassifikation nach zwei Klassen c_1 und c_2).
- Die Klassen sind linear separierbar, d.h. es gibt eine lineare Funktion

$$D(m_1,m_2,\ldots,m_\mathrm{k})=w_0+w_1m_1+\ldots+w_\mathrm{M}m_\mathrm{M}$$

  so dass $D()<0$ für die Merkmale aller Segmente aus c_1 und $D()>0$ für die Merkmale aller Segmente aus c_2.
- **Training**: Finde die Parameter $w_0,\ldots,w_\mathrm{M}$
- **Klassifikation**: Einsetzen der Merkmale eines unbekannten Segments in die Funktion und Entscheidung nach Funktionswert.

Entscheidungsgrenze finden

- Initialisiere die Gewichtungen $w_0,...,w_M$ mit kleinen Zufallswerten.
- Solange die Klassifikation nicht perfekt ist bzw. solange der durchschnittliche Klassifikationsfehler zu groß ist
- Wähle eine Stichprobe s aus dem Trainingsdatensatz und berechne $D(s)$
- Falls class(s) ≠ sign($D(s)$) dann (mit $a_1,a_2 > 0$)
  - $w_i := w_i + a_1 \cdot \text{class}(s) \cdot s_i$, für $i=1,..M$
  - $w_0 := w_0 + a_2 \cdot \text{class}(s)$



Lineare Entscheidungsgrenzen

- Effiziente Klassifikation für Zwei-Klassen-Problem.
- Abbruchkriterium notwendig, falls Klassen nicht linear separierbar sind.
- Mehr-Klassen-Problem: Paarweise Klassifikation und Zuordnung zu derjenigen Klasse, die bei keiner dieser Klassifikationen ausgeschlossen wurde. (→ denke an Intervallschachtelung)
- Erweiterungen/Alternativen
  - Nicht-lineare Funktionen über Einbettung in einen höher-dimensionalen Raum.
  - Diskriminanzfunktion: Parametrisierbare Funktion; Klassifikation nach Nähe zu prototypischen Funktionswerten

Beispielrechnung

- A-Priori-Wahrscheinl.
  - P(a)=P(b)=0.5
- Testmenge
  - N(a)=N(b)=4
- Segmentierung= ¼
- Klassifikation:
  - Preiswert, leicht!

Zutaten :
4 Äpfel
4 Birnen
1 Zimtstange
4 TL Zucker oder Süßstoff
2 TL Butter oder Margarine
4 Eier
25 cl Milch

Rezept :
Den Backofen auf 200°C (Th. 6) vorheizen. Äpfel und Birnen schälen, das Kerngehäuse entfernen und vierteln.
In einer Pfanne die Butter mit der Hälfte des Zuckers erwärmen, die Früchte und die Zimtstange dazugeben und 10 Min. auf kleiner Flamme kochen.
Die Zimtstange entfernen und den Rest in eine Auflaufform geben. Die Eier verrühren, den restlichen Zucker und die Milch dazugeben und auf die Früchte gießen. 15 bis 20 Minuten backen.

Tipps :
Getränkeempfehlung: Sekt oder Champagner