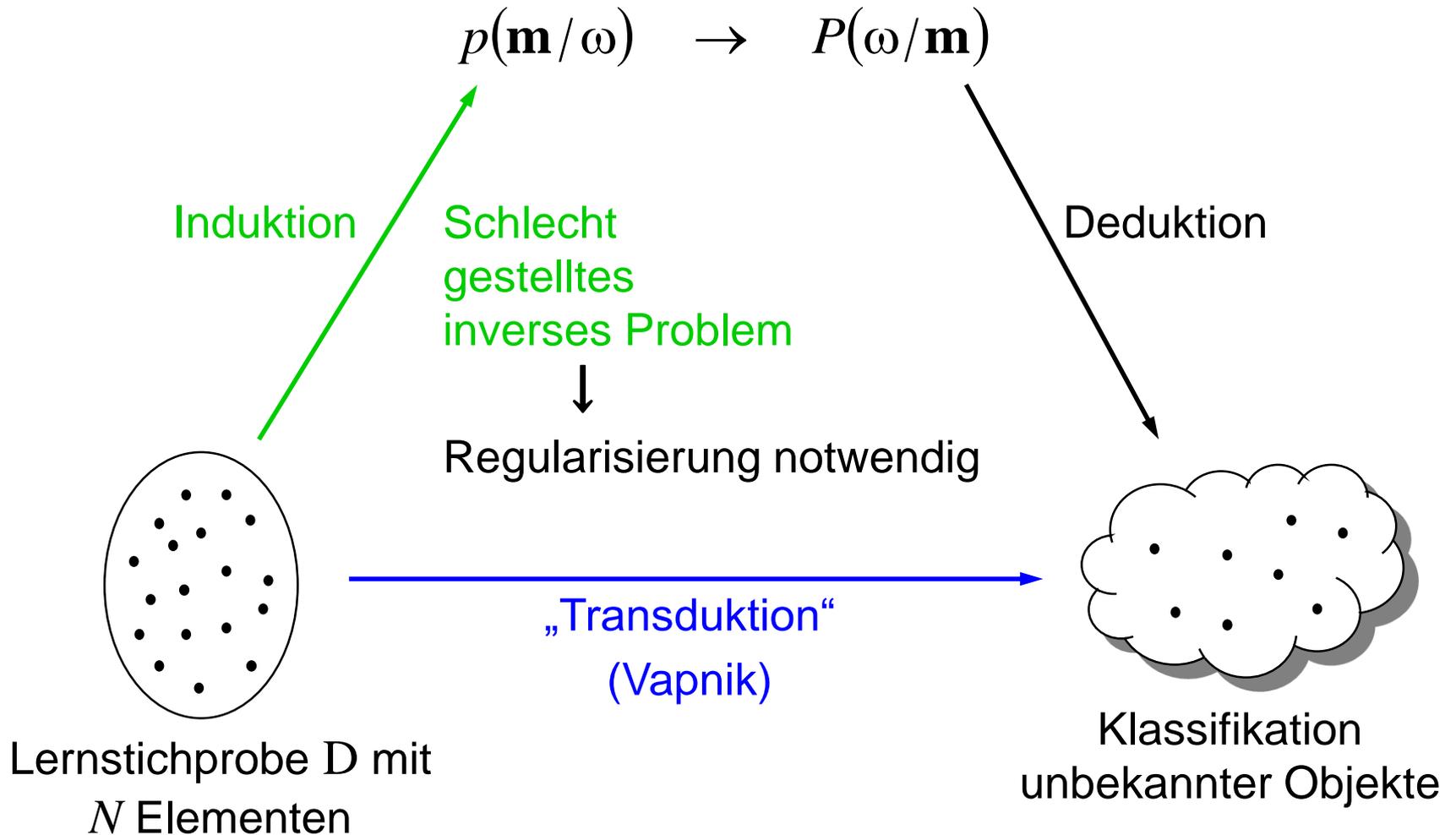


---

# 5. Parameterfreie Methoden

## 5. Parameterfreie Methoden

### Möglichkeiten zur Inferenz:



### Exkurs: Schlecht gestellte inverse Probleme

- Die Dichte  $p(\mathbf{m}|\omega)$  hat i. Allg. unendlich viele Freiheitsgrade, jedoch stehen nur  $N$  Stichproben zur Schätzung zur Verfügung.
- Die Inferenz von  $y$  auf  $x$  gegeben eine Abbildung  $y = A(x)$  ist ein **schlecht gestelltes inverses Problem** (ill posed inverse problem), wenn mindestens eine der folgenden 3 Eigenschaften auf  $A^{-1}$  zutrifft:
  - $A^{-1}$  nicht wohldefiniert
  - $x = A^{-1}(y)$  nicht eindeutig
  - $A^{-1}$  nicht stetig

Lösung mittels **Regularisierung**:

- Zusätzliche Annahmen machen
- Vorwissen einbringen
- Wunscheigenschaften als Nebenbedingungen formulieren

Unser Beispiel:  $A : p(\mathbf{m} | \omega) \mapsto D$

Regularisierung z.B. durch **parametrisierte Familien von WDFen**

## 5. Parameterfreie Methoden

---

### Probleme:

- Funktionale Form der WDFen  $p(\mathbf{m}|\omega_i)$  unbekannt
- Praktische Aufgabenstellungen sind nicht immer adäquat mit den gängigen parametrisierten, unimodalen WDFen beschreibbar.
- WDFen hochdimensionaler Merkmalsvektoren sind nicht immer adäquat mit Produkten eindimensionaler, parametrisierter WDFen beschreibbar.

### Annahmen für dieses Kapitel:

- Überwachtes Lernen
- Funktionale Form der WDFen  $p(\mathbf{m}|\omega_i)$  unbekannt; Es wird **keine parametrisierte Familie von WDFen angenommen**.
- Die Aufgabe der Schätzung der  $p(\mathbf{m}|\omega_i)$  zerfällt in  $c$  gleichartige Einzelaufgaben für jede der Klassen  $\omega_i$ ,  $i = 1, \dots, c$ . **Im Folgenden wird daher die Kennzeichnung der Klasse unterdrückt und nur  $p(\mathbf{m})$  notiert.**

## 5. Parameterfreie Methoden

---

Wahrscheinlichkeit, dass  $\mathbf{m}$  in die Region A fällt.

$$P := P(A) = \int_A p(\mathbf{m}) d\mathbf{m}$$

$P(A)$  entspricht einer geglätteten Version der WDF  $p(\mathbf{m})$ . Diese geglättete Version kann durch Schätzung von  $P(A)$  geschätzt werden.

Sei  $\{\mathbf{m}_1, \dots, \mathbf{m}_N\}$  i.i.d.

Wahrscheinlichkeit, dass  $k$  der  $N$  Daten in die Region A fallen, beträgt:

$$P(k) = \binom{N}{k} P^k (1 - P)^{N-k}$$

Binomialverteilung

$$E\{k\} = NP$$

## 5. Parameterfreie Methoden

### Annahmen:

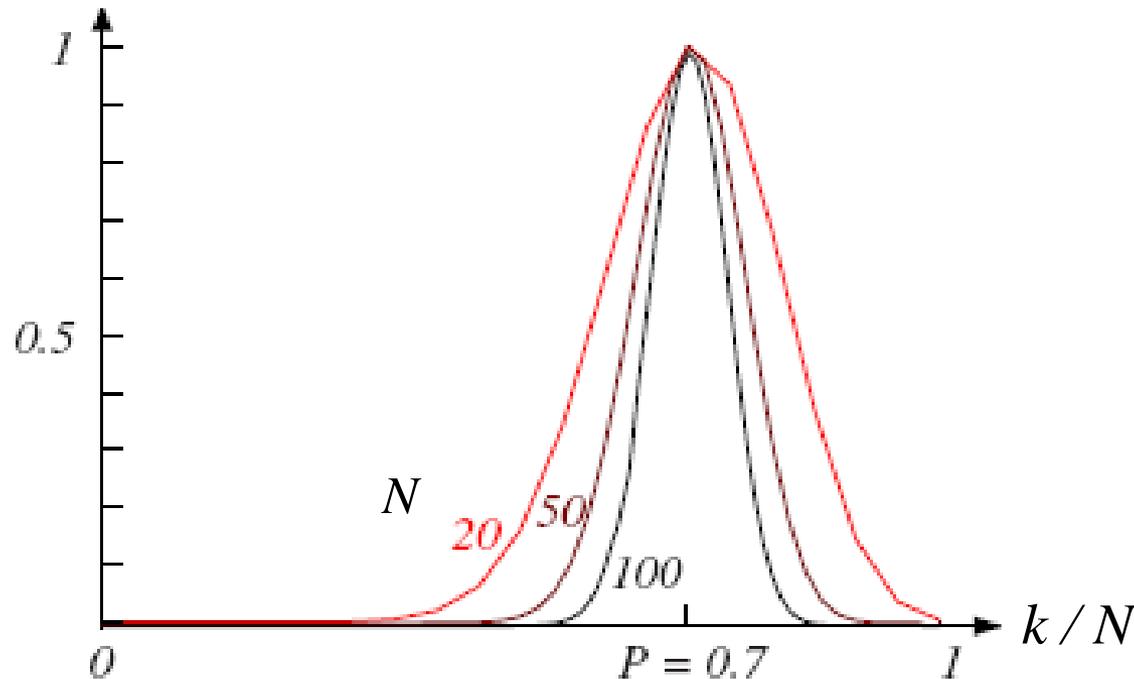
- $p(\mathbf{m})$  sei stetig innerhalb  $A$ .
- $A$  sei klein genug, dass  $p(\mathbf{m})$  innerhalb  $A$  näherungsweise konstant ist.
- $V$ : Volumen der Region  $A$

$$\begin{aligned} \Rightarrow & \int_A p(\mathbf{m}') d\mathbf{m}' \approx p(\mathbf{m})V \\ & P = \int_A p(\mathbf{m}') d\mathbf{m}' \\ & P = \frac{E\{k\}}{N} \approx \frac{k}{N} \end{aligned} \quad \Rightarrow \quad p(\mathbf{m}) \approx \frac{k/N}{V}$$

## 5. Parameterfreie Methoden

Binomialverteilung zeigt beim Erwartungswert ein scharf ausgeprägtes Maximum  $\rightarrow k/N$  ist eine geeignete Schätzung für  $P$ .

Bsp.: Auf  $\max = 1$  normierte Binomialverteilungen für verschiedene  $N$  und  $P = 0.7$



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

## 5. Parameterfreie Methoden

### Diskussion:

$$V \text{ fest: } \frac{P}{V} = \frac{P(A)}{V(A)} = \frac{\int_A p(\mathbf{m}') d\mathbf{m}'}{\int_A d\mathbf{m}'} \quad \text{schätzt geglättete Version von } p(\mathbf{m})$$

Übergang:  $V \rightarrow 0, N \rightarrow \infty$ , so dass  $P/V \rightarrow p(\mathbf{m})$

$A_1, A_2, \dots$  sei Folge von Regionen.

$V_N$ : Volumina der  $A_N$

$k_N$ : Zahl der Stichproben, die in  $A_N$  fallen.

$$\hat{p}_N(\mathbf{m}) := \frac{k_N / N}{V_N}$$

Notwendig für Konvergenz  
für  $N \rightarrow \infty$ :

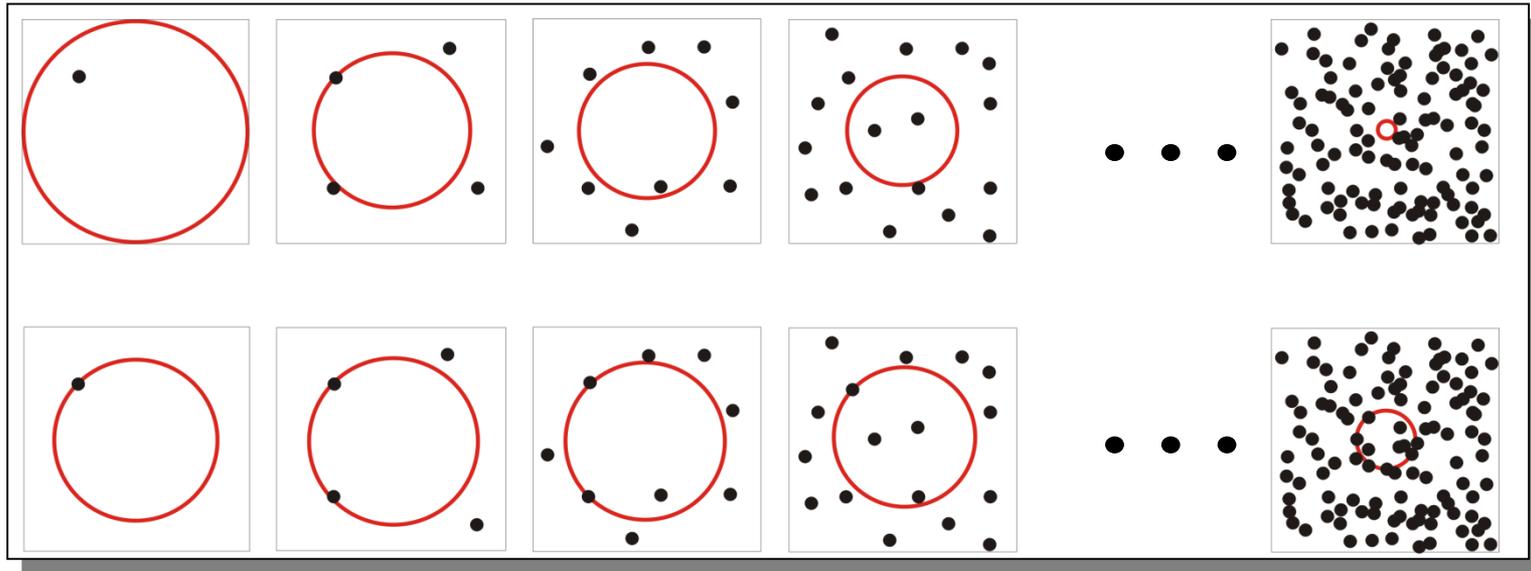
$$\left\{ \begin{array}{l} \lim_{N \rightarrow \infty} V_N = 0 \\ \lim_{N \rightarrow \infty} k_N = \infty \\ \lim_{N \rightarrow \infty} k_N / N = 0 \end{array} \right.$$

Ohne Herleitung: Diese Bedingungen sind sogar hinreichend für die Konvergenz nach Wahrscheinlichkeit für alle  $\mathbf{m}$ , an denen  $p(\mathbf{m})$  stetig ist.

## 5. Parameterfreie Methoden

### Ansätze:

- **Parzen Fenster:** Volumen als Funktion von  $N$  z.B.  $V_N \propto 1/\sqrt{N}$   
Die Volumina sind eine deterministische Funktion der Stichprobengröße und hängen nicht von den Daten ab.
- **Nächste Nachbarn:** Anzahl als Funktion von  $N$  z.B.  $k_N \propto \sqrt{N}$   
Die Volumina hängen von den Daten ab und sind mithin Zufallsvariablen.



## 5.1. Parzen Fenster

$A_N$  wird als  $d$ -dimensionaler Würfel gewählt

$h_N$  : Kantenlänge

$V_N = h_N^d$  : Volumen

$\varphi(\mathbf{u}) := \text{rect}(\mathbf{u}) = \begin{cases} 1 & \text{für } |u_j| \leq \frac{1}{2}, \quad \forall j = 1, \dots, d \\ 0 & \text{sonst} \end{cases}$  : Fensterfunktion

$\Rightarrow \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right) = \begin{cases} 1 & \text{für } \mathbf{m} \in \text{Würfel } A_N \text{ um } \mathbf{m}_i \\ 0 & \text{sonst} \end{cases}$

$\Rightarrow \left. \begin{aligned} k_N &= \sum_{i=1}^N \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right) \\ \hat{p}_N(\mathbf{m}) &= \frac{k_N}{N} \end{aligned} \right\} \Rightarrow \hat{p}_N(\mathbf{m}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right)$

Entspricht einer Approximation der WDF durch Interpolation zwischen den Stützstellen  $\mathbf{m}_i$  mittels der Funktion  $\varphi(\cdot)$ .

## 5.1. Parzen Fenster

Schreibweise:

$$\left. \begin{aligned} \delta_N(\mathbf{m}) &:= \frac{1}{V_N} \varphi\left(\frac{\mathbf{m}}{h_N}\right) \\ \hat{p}_N(\mathbf{m}) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right) \end{aligned} \right\} \Rightarrow \hat{p}_N(\mathbf{m}) = \frac{1}{N} \sum_{i=1}^N \delta_N(\mathbf{m} - \mathbf{m}_i)$$

$$\delta_N(\mathbf{m}) \xrightarrow{N \rightarrow \infty} \delta(\mathbf{m}) \quad \text{Dirac'sche "Deltafunktion"}$$

**Verallgemeinerung der Fensterfunktion:**

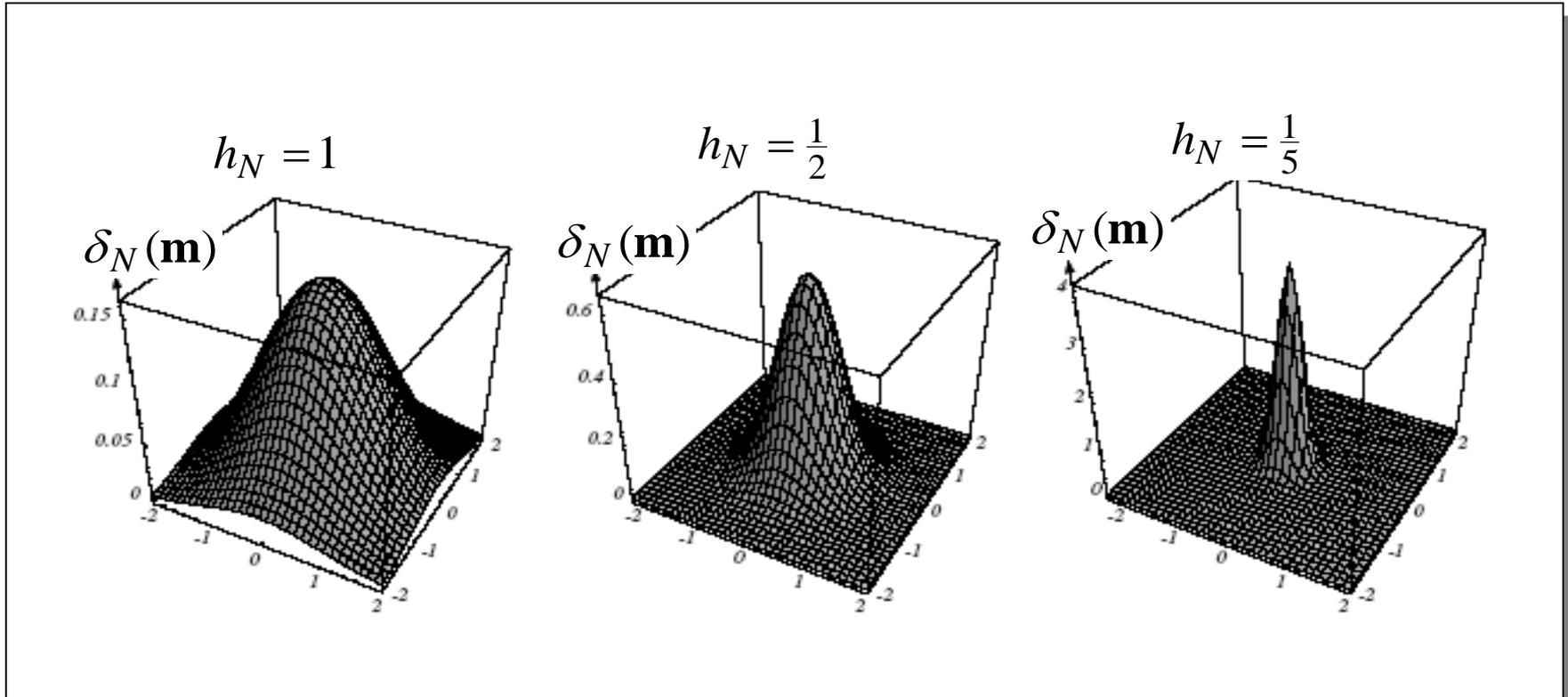
Aus:

- $\delta_N(\mathbf{m}) \geq 0$
  - $\int \delta_N(\mathbf{u}) d\mathbf{u} = 1$
- d.h.: formal ist  $\delta_N(\cdot)$  eine WDF, } folgt  $\int \hat{p}_N(\mathbf{m}) d\mathbf{m} = 1$  und  $\hat{p}_N(\mathbf{m}) \geq 0$   
D.h.  $\hat{p}_N(\cdot)$  ist formal eine WDF.

**Verallgemeinerung:** für  $\delta_N(\cdot)$  kann beliebige WDF gewählt werden.

## 5.1. Parzen Fenster

**Beispiel:** 
$$\delta_N(\mathbf{m}) := \frac{1}{2\pi h_N^2} \exp\left(-\frac{1}{2h_N^2} \|\mathbf{m}\|^2\right)$$

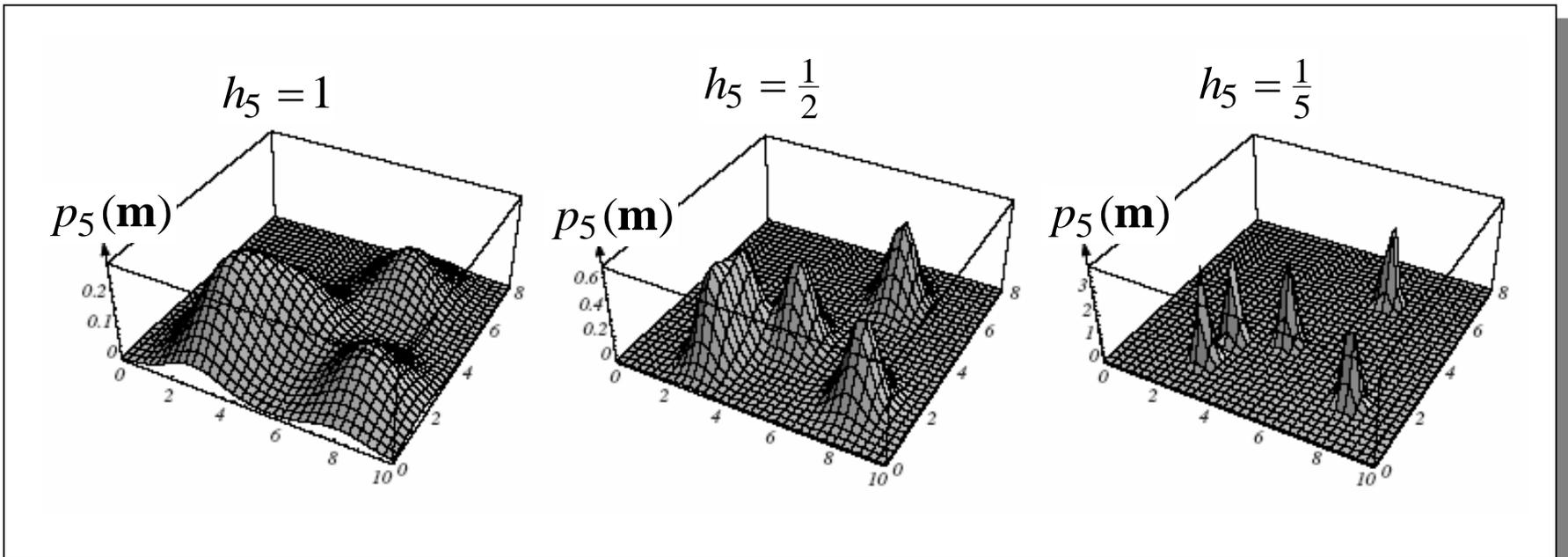


Quelle: R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification

## 5.1. Parzen Fenster

**Beispiel:** 
$$\delta_N(\mathbf{m}) := \frac{1}{2\pi h_N^2} \exp\left(-\frac{1}{2h_N^2} \|\mathbf{m}\|^2\right)$$

Parzen Fenster Schätzung für fünf Stichproben,  $N = 5$ .



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

## 5.1. Parzen Fenster

### Konvergenz:

$\hat{p}_N(\mathbf{m})$  habe den Erwartungswert  $\mu_N(\mathbf{m})$  und die Varianz  $\sigma_N^2(\mathbf{m})$

$\hat{p}_N(\mathbf{m}) \rightarrow p(\mathbf{m})$  im quadratischen Mittel, wenn:

$$\begin{cases} \lim_{N \rightarrow \infty} \mu_N(\mathbf{m}) = p(\mathbf{m}) \\ \lim_{N \rightarrow \infty} \sigma_N^2(\mathbf{m}) = 0 \end{cases}$$

### Konvergenz des Erwartungswertes:

$$\mu_N(\mathbf{m}) = \mathbb{E}\{\hat{p}_N(\mathbf{m})\} = \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left\{ \frac{1}{V_N} \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right) \right\} = \int \frac{1}{V_N} \varphi\left(\frac{\mathbf{m} - \mathbf{v}}{h_N}\right) p(\mathbf{v}) d\mathbf{v}$$

$$= \int \delta_N(\mathbf{m} - \mathbf{v}) p(\mathbf{v}) d\mathbf{v} = \delta_N(\mathbf{m}) \underbrace{*}_{\text{Faltung}} p(\mathbf{m}) \quad \Rightarrow$$

Wenn  $V_N \rightarrow 0$ , dann  $\delta_N(\mathbf{m}) \rightarrow$  Diracsche „Deltafunktion“  $\delta(\mathbf{m})$

$\delta(\mathbf{m})$  ist das Einselement der Faltungsoperation.

$$\Rightarrow \boxed{\mu_N(\mathbf{m}) \xrightarrow{V_N \rightarrow 0} p(\mathbf{m})}$$

## 5.1. Parzen Fenster

### Konvergenz der Varianz:

$$\sigma_N^2(\mathbf{m}) = \text{Var}\{\hat{p}_N(\mathbf{m})\} = \text{E}\{(\hat{p}_N(\mathbf{m}) - \mu_N(\mathbf{m}))^2\}$$

$$= \sum_{i=1}^N \text{E}\left\{\left(\frac{1}{NV_N} \varphi\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right) - \frac{1}{N} \mu_N(\mathbf{m})\right)^2\right\}$$

wg. der Unabhängigkeit der Stichproben:  
Varianz = Summe der Varianzen der Summanden

$$= N \text{E}\left\{\frac{1}{N^2 V_N^2} \varphi^2\left(\frac{\mathbf{m} - \mathbf{m}_i}{h_N}\right)\right\} - \frac{1}{N} \mu_N^2(\mathbf{m})$$

$$= \frac{1}{NV_N} \int \frac{1}{V_N} \varphi^2\left(\frac{\mathbf{m} - \mathbf{v}}{h_N}\right) p(\mathbf{v}) d\mathbf{v} - \frac{1}{N} \mu_N^2(\mathbf{m}) \leq \frac{1}{NV_N} \int \frac{1}{V_N} \varphi^2\left(\frac{\mathbf{m} - \mathbf{v}}{h_N}\right) p(\mathbf{v}) d\mathbf{v}$$

Unter der Voraussetzung:  $\sup_{\mathbf{u}} \{\varphi(\mathbf{u})\} < \infty$  gilt die Abschätzung:

$$\sigma_N^2(\mathbf{m}) \leq \frac{\sup\{\varphi(\cdot)\} \mu_N(\mathbf{m})}{NV_N}$$

## 5.1. Parzen Fenster

### Konvergenz – Zusammenfassung:

$$\mu_N(\mathbf{m}) = \delta_N(\mathbf{m}) * p(\mathbf{m})$$

$$\sigma_N^2(\mathbf{m}) \leq \frac{\sup\{\varphi(\cdot)\}\mu_N(\mathbf{m})}{NV_N}$$

Konvergenz im quad. Mittel:  $\hat{p}_N(\mathbf{m}) \rightarrow p(\mathbf{m})$ , wenn:

$$\lim_{N \rightarrow \infty} V_N = 0$$

$$\sup_{\mathbf{u}} \{\varphi(\mathbf{u})\} < \infty$$

$$\lim_{N \rightarrow \infty} NV_N = \infty$$

$$\lim_{\|\mathbf{u}\| \rightarrow \infty} \varphi(\mathbf{u}) \prod_{i=1}^d u_i = 0$$

Die letzte Voraussetzung wurde in unserer Beweisskizze nicht verwendet, wird aber bei einer strengen Beweisführung benötigt.

## 5.1. Parzen Fenster

---

### Konvergenz – Zusammenfassung:

$V_N$  muss für  $N \rightarrow \infty$  langsamer als  $1/N$  gegen Null streben.

### Beispiel:

$$\varphi(u) := \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

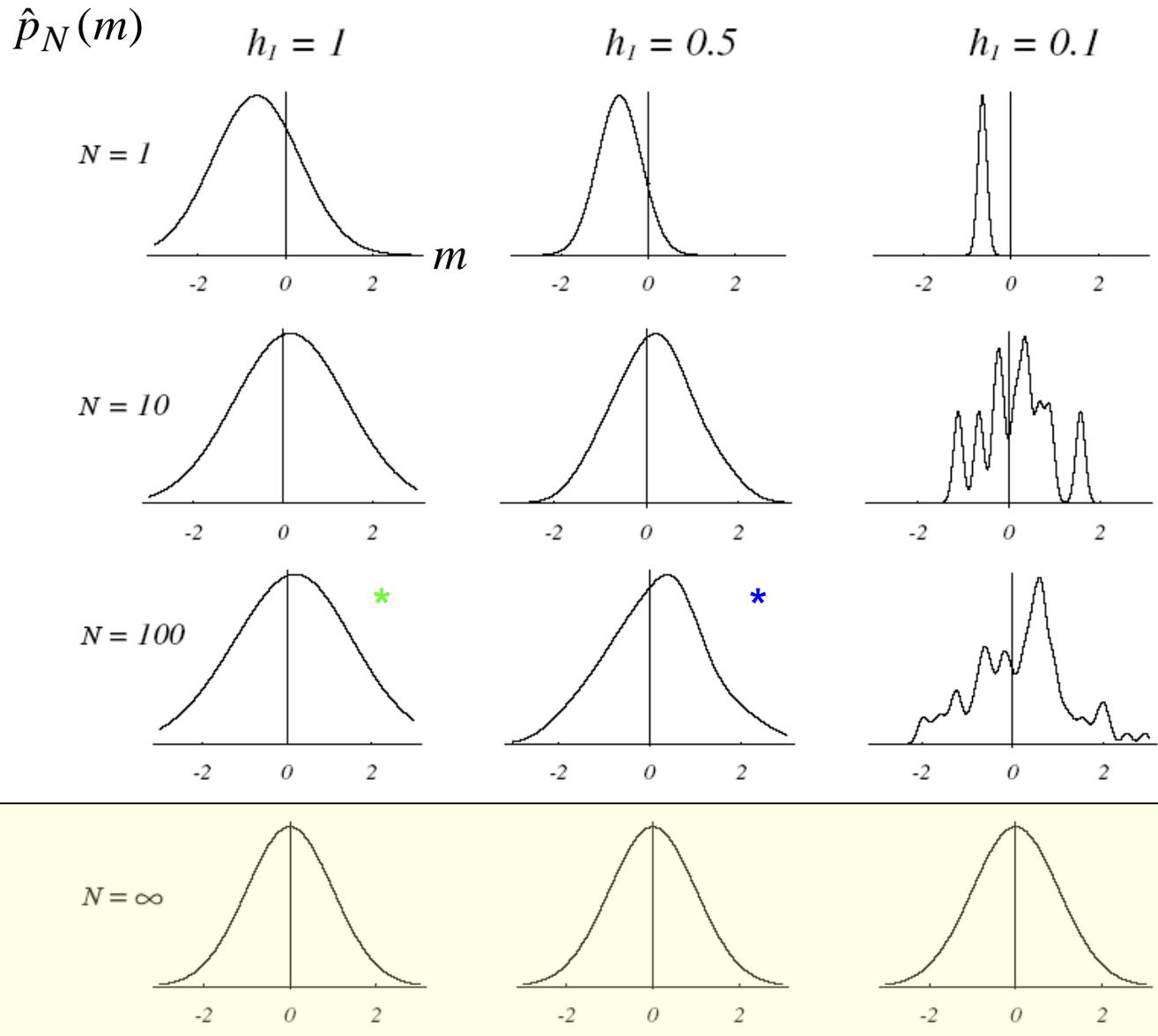
$$h_N := \frac{h_1}{\sqrt{N}}, \text{ mit } h_1 \text{ Designparameter}$$

$$\hat{p}_N(m) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_N} \varphi\left(\frac{m - m_i}{h_N}\right) = \frac{1}{h_1 \sqrt{N}} \sum_{i=1}^N \varphi\left(\frac{m - m_i}{h_1} \sqrt{N}\right)$$

# 5.1. Parzen Fenster

**Beispiel:**

Dimension  $d = 1$



→ Starke Abhängigkeit von der Wahl der Anfangsgröße  $h_1$

$p(m)$

$N = \infty$

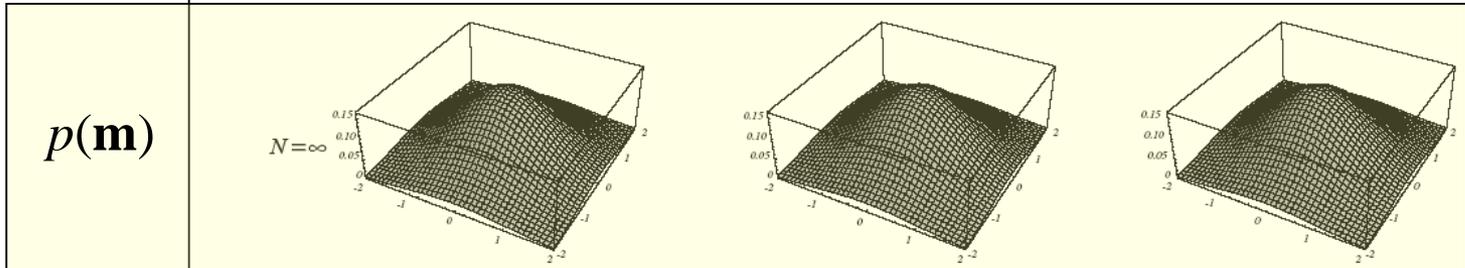
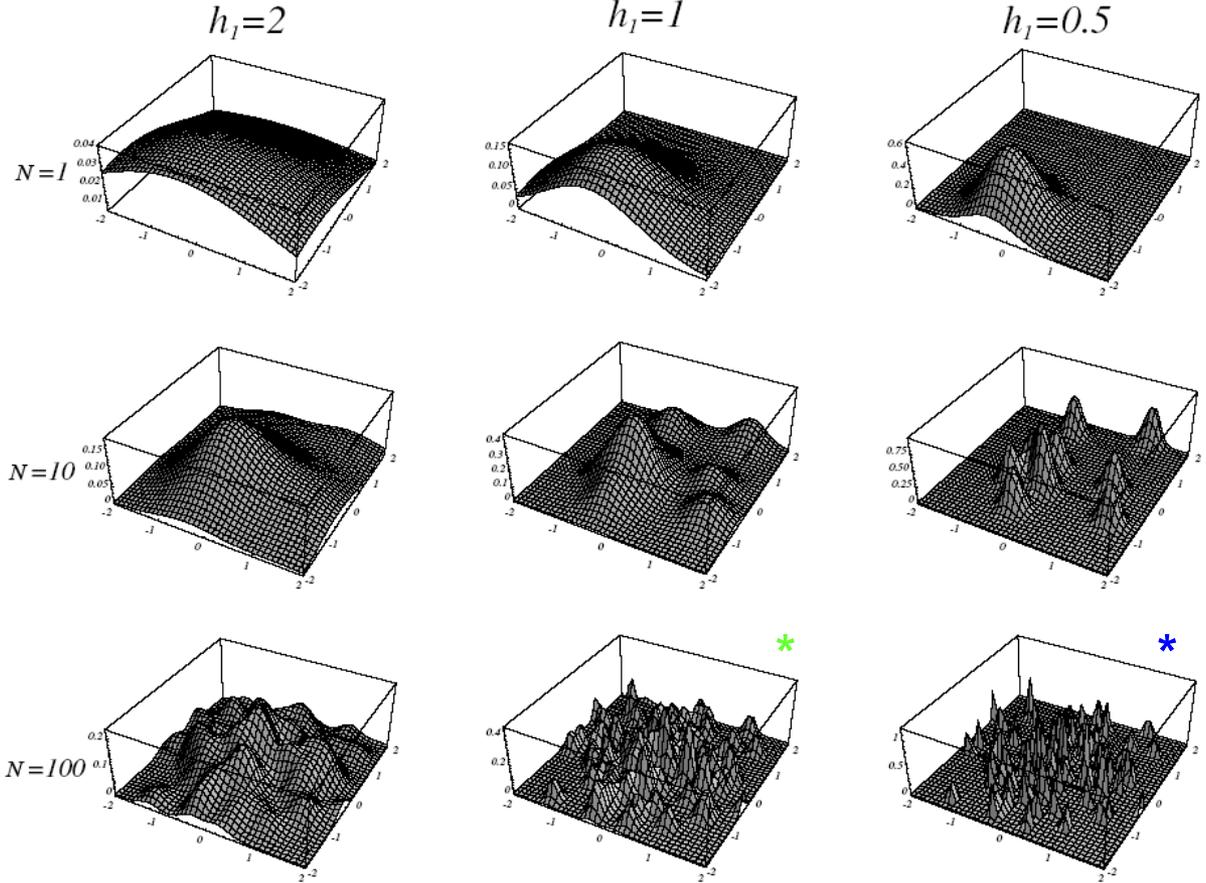
# 5.1. Parzen Fenster

**Beispiel:**  $d = 2$

$$h_N := \frac{h_1}{\sqrt[4]{N}}$$

$$\Rightarrow V_N = \frac{h_1^2}{\sqrt{N}}$$

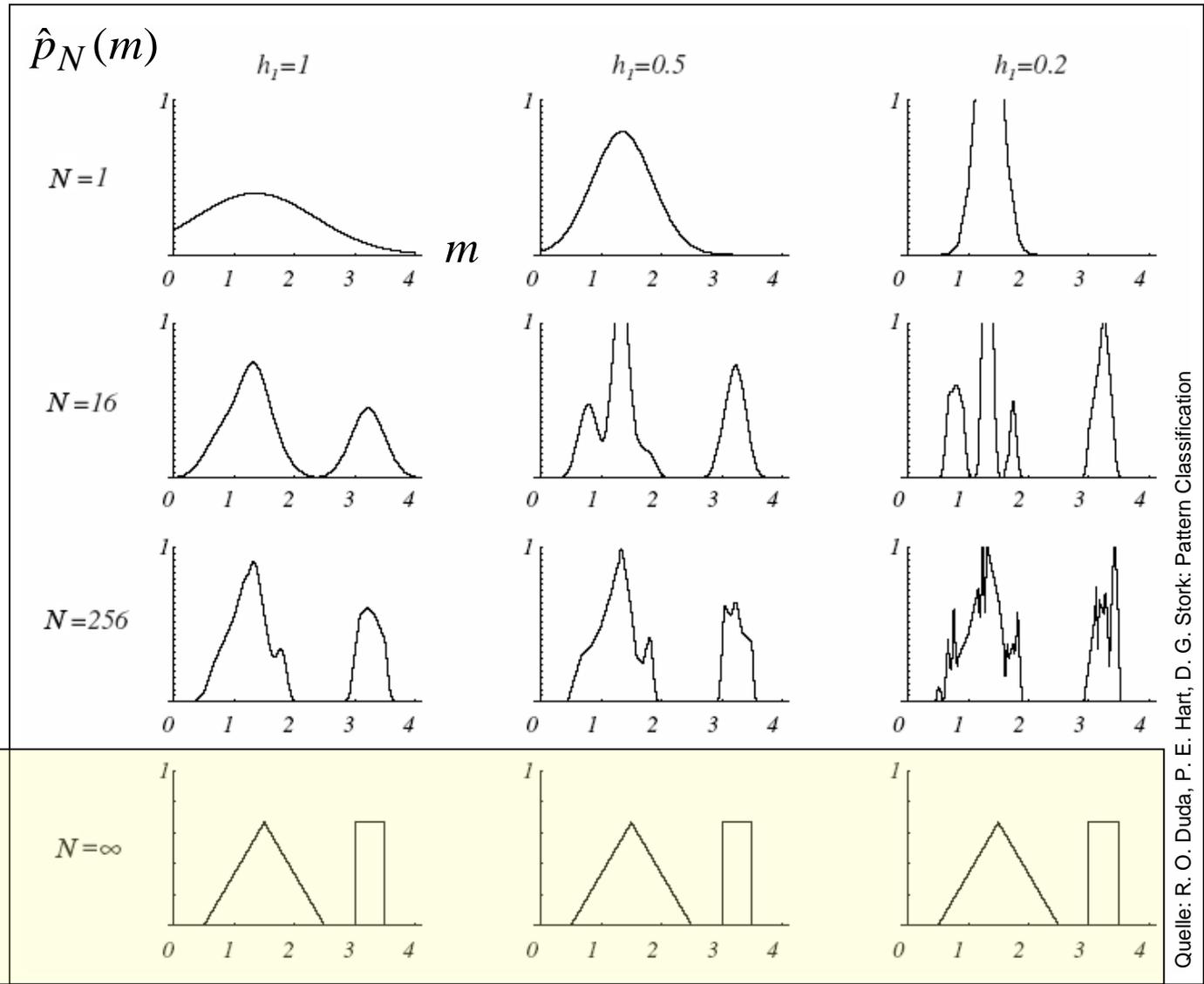
- Starke Abhängigkeit von der Wahl der Anfangsgröße  $h_1$
- Höhere Dimension  $d=2$  bedingt geringere räumliche Dichte der Stichproben und damit schlechtere Konvergenz. Vergleiche z.B.: \* und \*



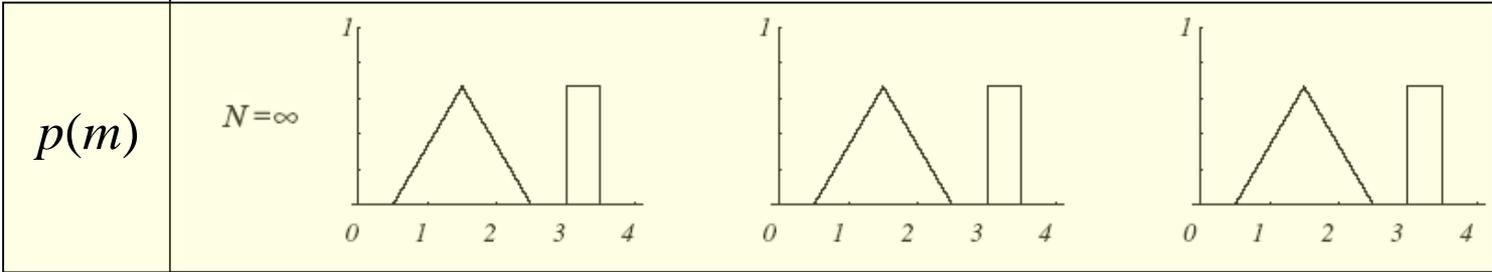
# 5.1. Parzen Fenster

## Beispiel:

Dimension  $d = 1$ ,  
bimodale WDF



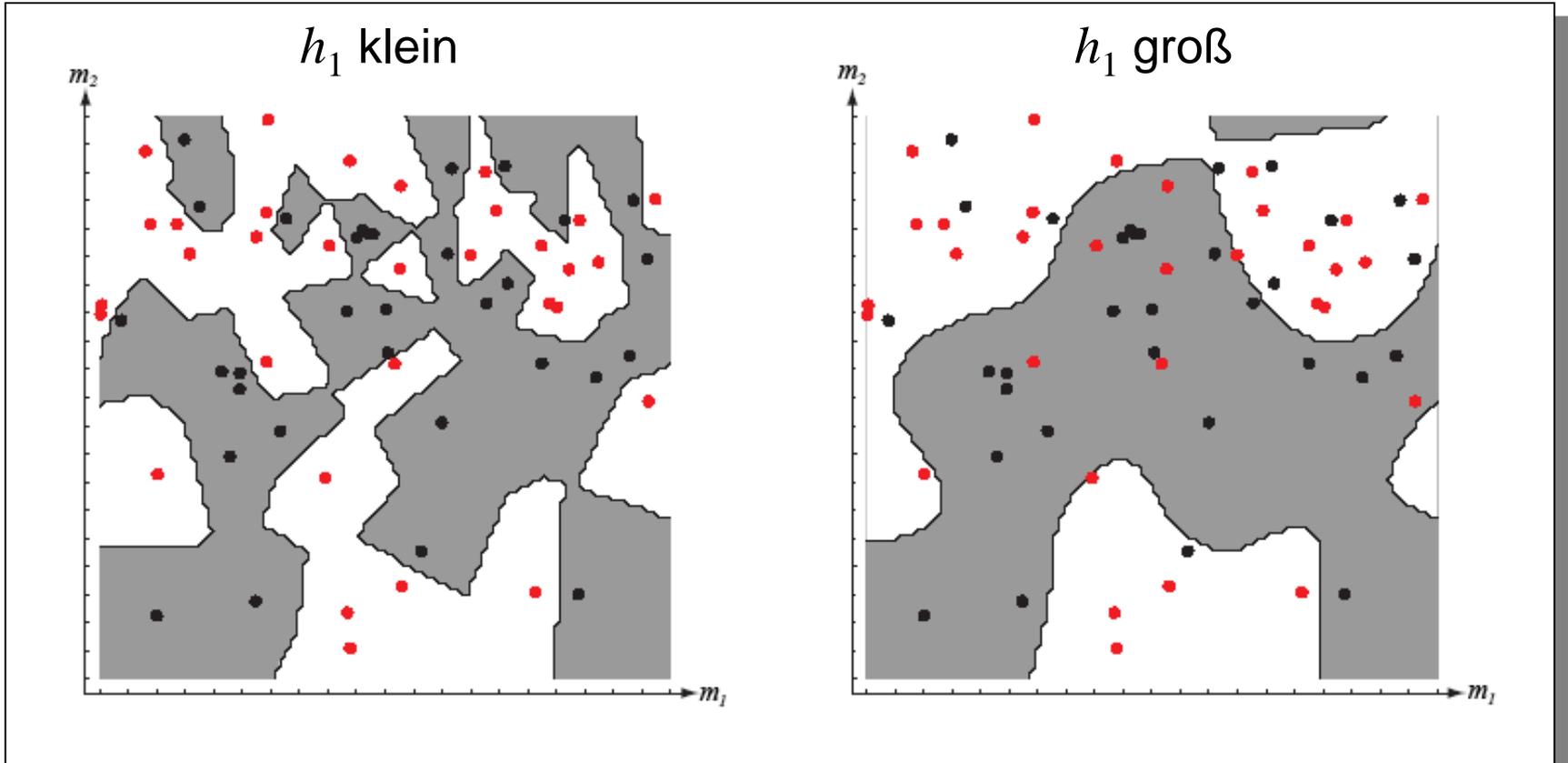
→ Starke Abhängigkeit  
von der Wahl der  
Anfangsgröße  $h_1$



## 5.1. Parzen Fenster

### Beispiel:

Entscheidungsgebiete auf Basis der Parzen Fenster Methode,  $c = 2$ ,  $d = 2$ .



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

## 5.1. Parzen Fenster

---

### Diskussion:

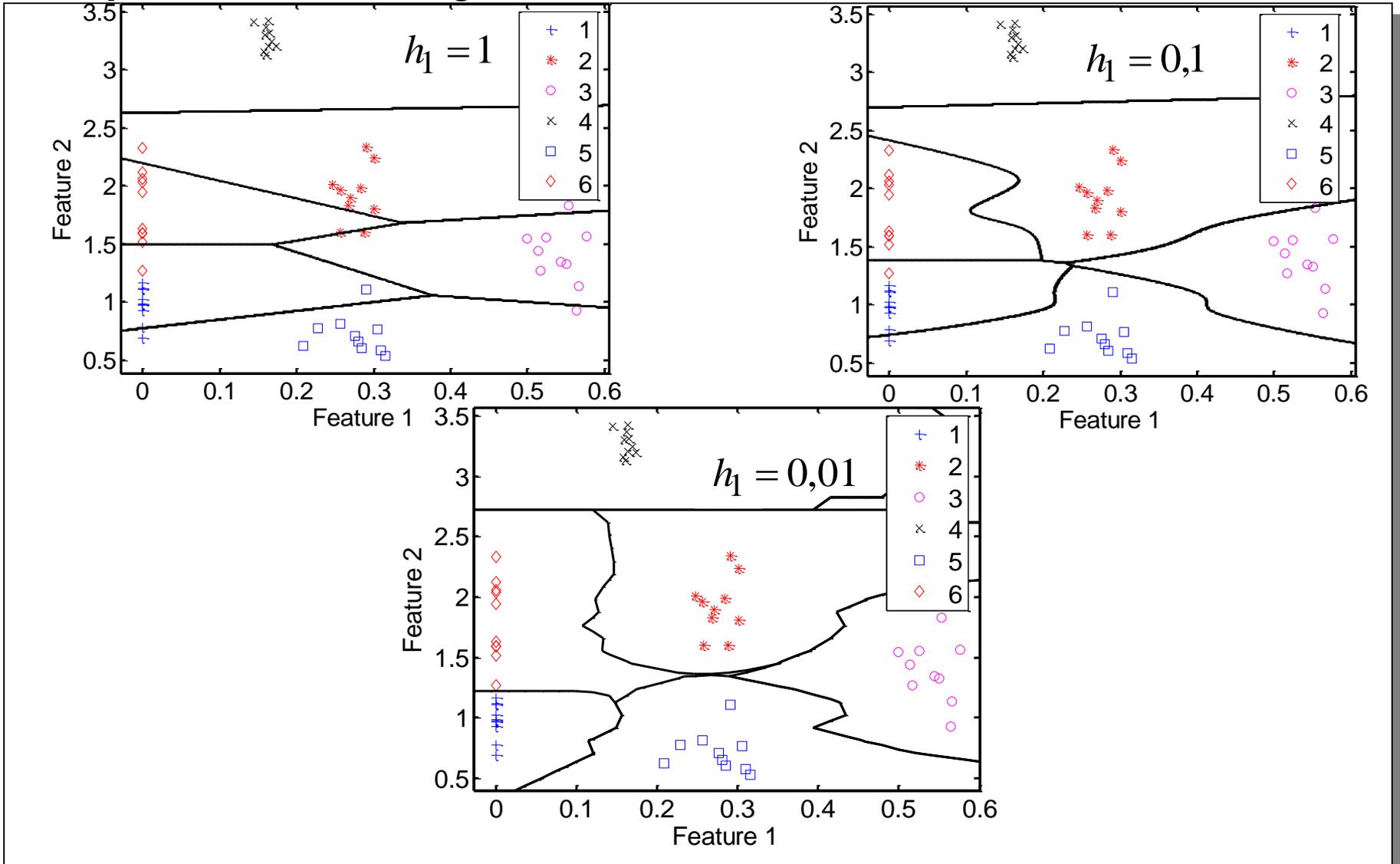
**Universalität:** Parzen Fenster Methode kommt ohne jegliches Vorwissen über die zugrunde liegenden WDFen aus. Kompliziert geformte, multi-modale WDFen können geschätzt werden.

**Parameterwahl:** Ergebnisse der Parzen Fenster Methode hängen für endliche Datensätze sehr stark von der Wahl des Anfangsvolumens  $V_1$  bzw. von der damit assoziierten Länge  $h_1$  ab.

**Datenunabhängige Regionengröße:** Regionen des Merkmalsraumes mit unterschiedlich dicht liegenden Stichproben werden durch die bei festem  $N$  für den ganzen Merkmalsraum konstante Größe der Region  $A_N$  gleich behandelt.

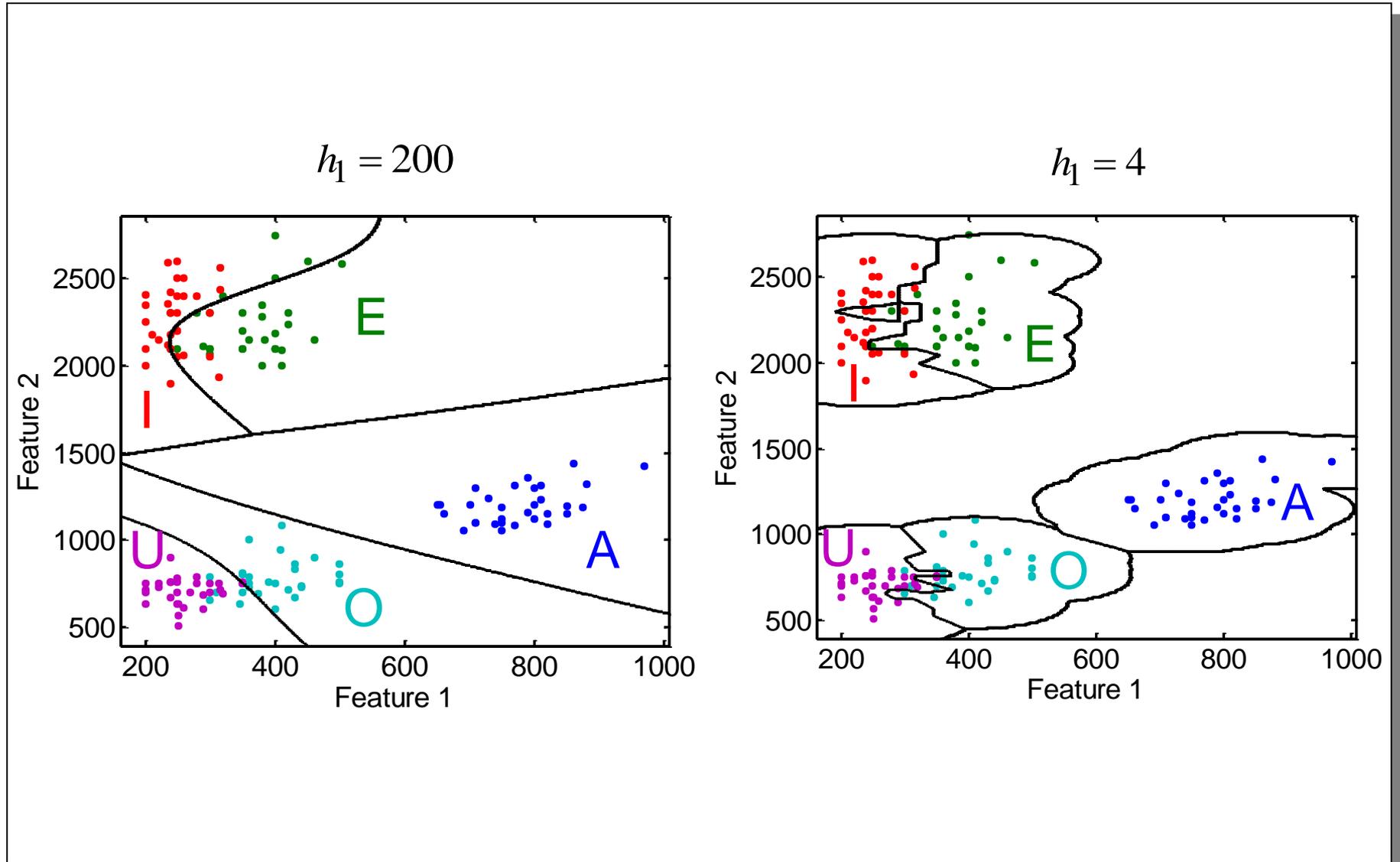
# 5.1. Parzen Fenster

## Beispiel: Griffserkennung Hand mit Parzen Klassifikator $N = 10$



## 5.1. Parzen Fenster

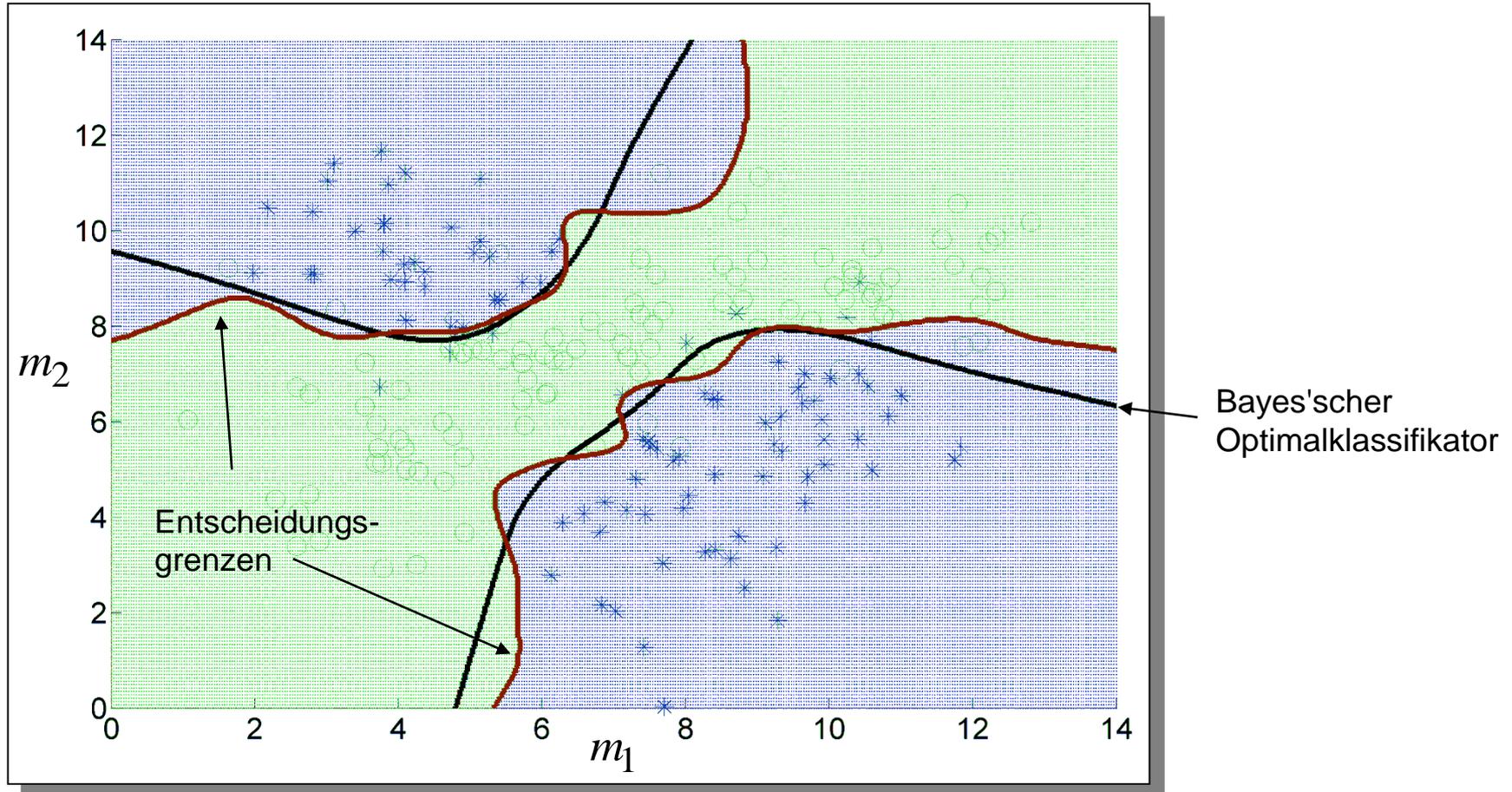
**Beispiel:** Vokalererkennung mit Parzen Klassifikator  $N = 30$



## 5.1. Parzen Fenster

**Beispiel:** Parzen Fenster Klassifikator  $h_1 \approx 0,5$

Lernstichprobe  $D$ :  $N = 200$  ( $N_1=100, N_2=100$ )

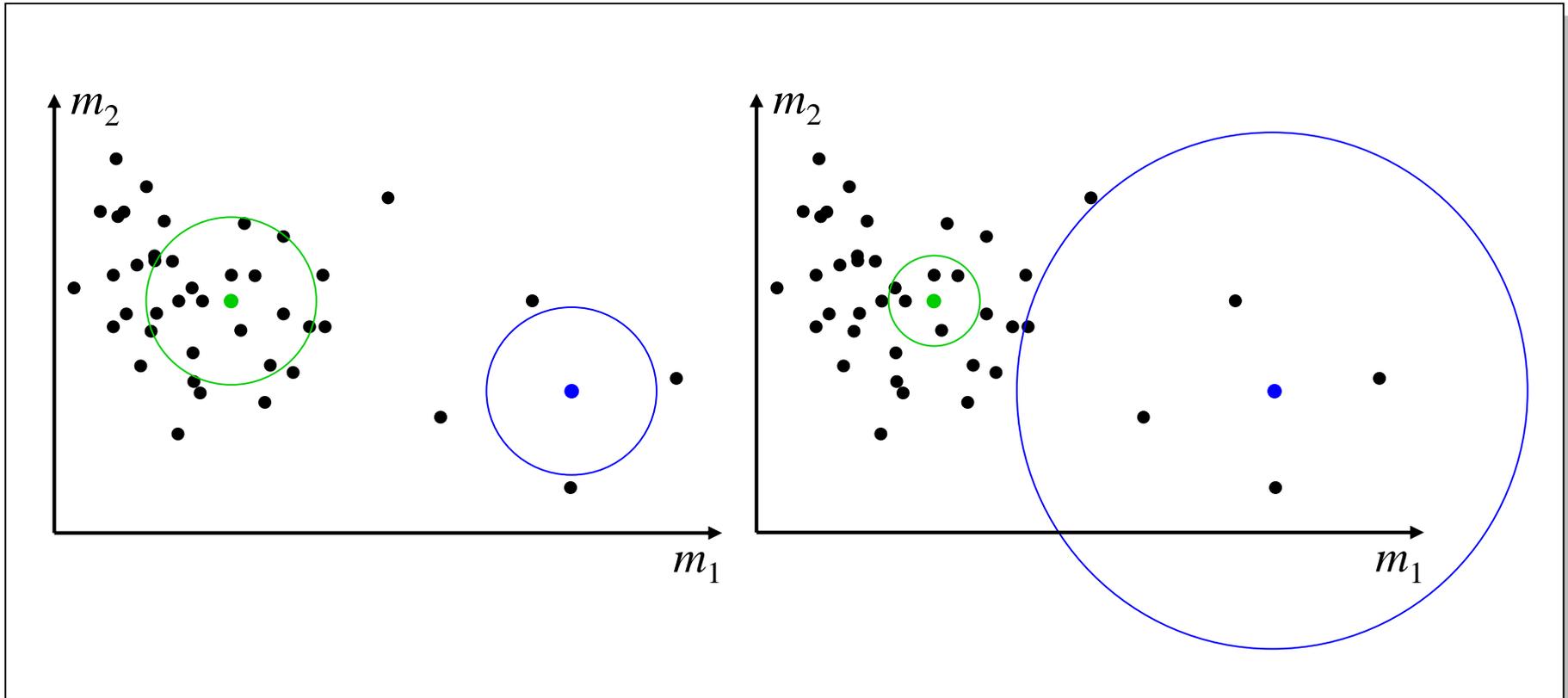


Klassifikation der Teststichprobe  $\rightarrow$  Testfehler = 10%

Asymptotischer Testfehler  $\approx 11,91\%$

## 5.2. $k_N$ - Nächste Nachbarn WDF Schätzung

### Diskussion:



Parzen Fenster Ansatz:  
 $V_N$  konstant bezüglich  $\mathbf{m}$ .

$k_N$ -Nächste Nachbarn Ansatz: hier für  $k_N = 5$ . Die Regionen passen sich der lokalen Dichte der Stichproben an.

## 5.2. $k_N$ - Nächste Nachbarn WDF Schätzung

---

### Darstellung:

- Sei  $A_k$  eine um  $\mathbf{m}$  zentrierte Zelle des Merkmalsraumes.
- Die Zelle werde ausgedehnt, bis sie  $k_N$  Stichproben enthält.

### Diskussion:

- Zelle ist klein (groß), wenn die WDF in der Nähe von  $\mathbf{m}$  groß (klein) ist.

### Forderungen:

- $\lim_{N \rightarrow \infty} k_N = \infty$  sichert, dass  $k_N / N$  eine gute Schätzung für  $P$  ist.  
(Stochastischer Fehler  $\rightarrow 0$ )
- $\lim_{N \rightarrow \infty} k_N / N = 0$ :  $k_N$  wächst langsam genug, dass Zellengröße  $\rightarrow 0$  geht.  
(Systematischer Fehler  $\rightarrow 0$ )

$$\Rightarrow \hat{p}_N(\mathbf{m}) \rightarrow p(\mathbf{m})$$

Ohne Herleitung: Diese Forderungen sind sogar hinreichend für die Konvergenz nach Wahrscheinlichkeit für alle  $\mathbf{m}$ , an denen  $p(\mathbf{m})$  stetig ist.

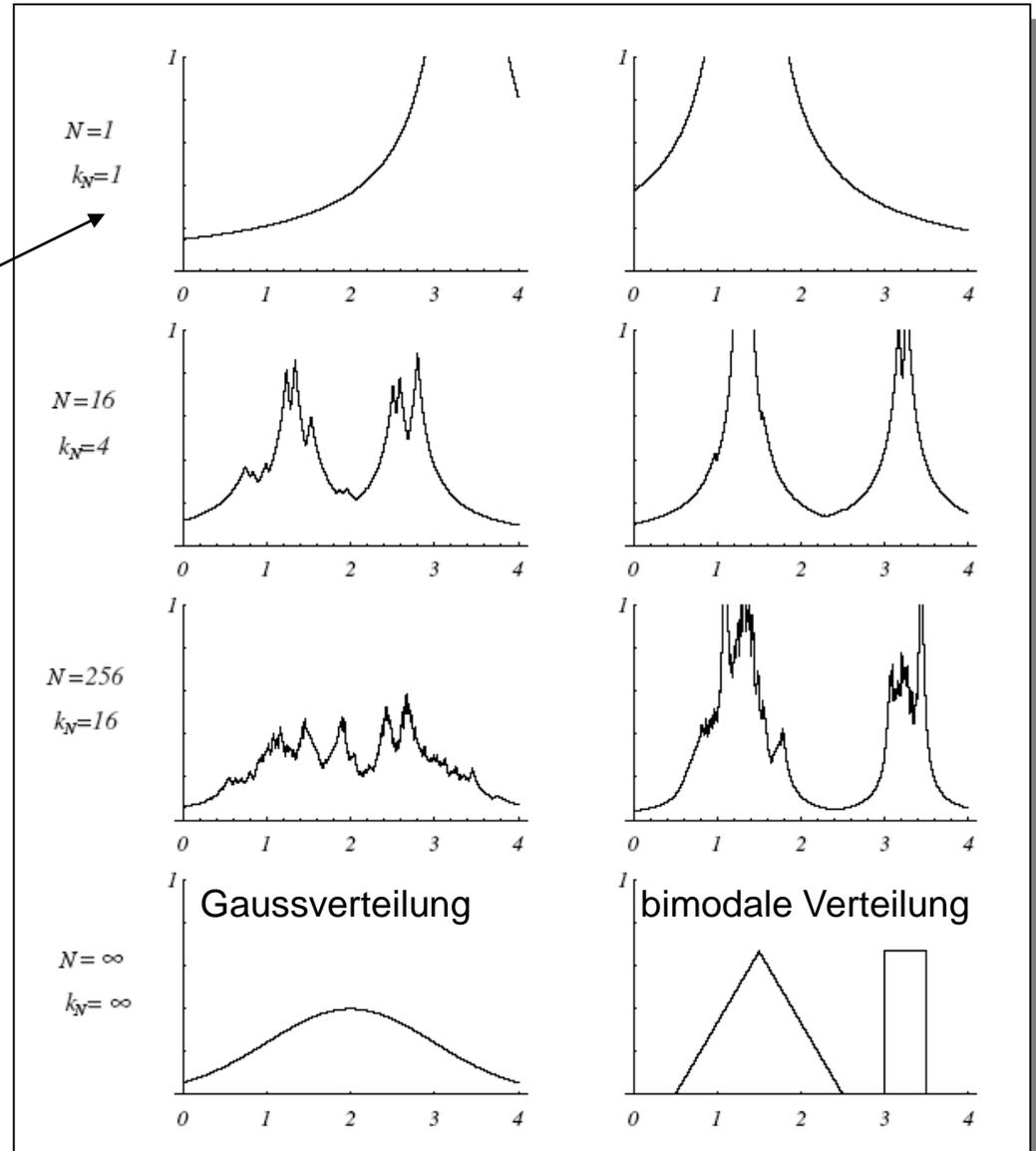
## 5.2. $k_N$ - Nächste Nachbarn WDF Schätzung

Beispiel,  $d = 1$ ,  $k_N := \sqrt{N}$

$$N = 1, \quad k_1 = 1$$

$$\hat{p}_1(m) = \frac{1}{2|m - m_1|}$$

→ Geschätzte WDFen sind stetig aber an den Knicken nicht differenzierbar. Knicke fallen i. Allg. nicht mit den Stichproben zusammen.



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

## 5.3. Nächste Nachbarn Klassifikation

### Heuristische Überlegung:

Sei  $A$  ein um  $\mathbf{m}$  zentriertes Gebiet mit dem Volumen  $V$ , in das  $k$  Stichproben fallen.  $k_i$  dieser Stichproben stammen von der Klasse  $\omega_i$ . Die Verbund-WV der Merkmale und Klassen lässt sich dann schätzen mittels:

$$\hat{P}_N(\mathbf{m}, \omega_i) = \frac{k_i / N}{V}$$

$$\Rightarrow \hat{P}_N(\omega_i | \mathbf{m}) = \frac{\hat{P}_N(\mathbf{m}, \omega_i)}{\sum_{j=1}^c \hat{P}_N(\mathbf{m}, \omega_j)} = \frac{k_i}{k}$$

In Regionen mit hoher A Posteriori Wahrsch.  $P(\omega_i | \mathbf{m}) > P(\omega_j | \mathbf{m}) \quad \forall j \neq i$  ist  $k_i$  tendentiell am größten. D.h. von den  $k$  nächsten Nachbarn zu  $\mathbf{m}$  werden die zur Klasse  $\omega_i$  gehörenden Stichproben tendenziell den größten Anteil stellen.

→ Entscheidung  $\hat{\omega}(\mathbf{m}) = \omega_i$  erscheint plausibel.

Spezialfall:  $k = \text{konstant} = 1 \rightarrow$  Nächster Nachbar Klassifikator

## 5.3. Nächste Nachbarn Klassifikation

**Gegeben:**  $D = \{\mathbf{m}_1, \dots, \mathbf{m}_N\}$  : klassifizierte Lernstichprobe

**Nächster Nachbar Klassifikation:**

$$\hat{\omega}(\mathbf{m}) := \omega_i \iff \arg \min_{\mathbf{m}_j \in D} \{ \|\mathbf{m} - \mathbf{m}_j\| \} \in D_i$$

**Entscheidungsfunktionen:**

$$k_i(\mathbf{m}) := \max_{\tilde{\mathbf{m}} \in D_i} \{ \|\mathbf{m} - \tilde{\mathbf{m}}\|^{-1} \} = \frac{1}{\min_{\tilde{\mathbf{m}} \in D_i} \{ \|\mathbf{m} - \tilde{\mathbf{m}}\| \}}$$

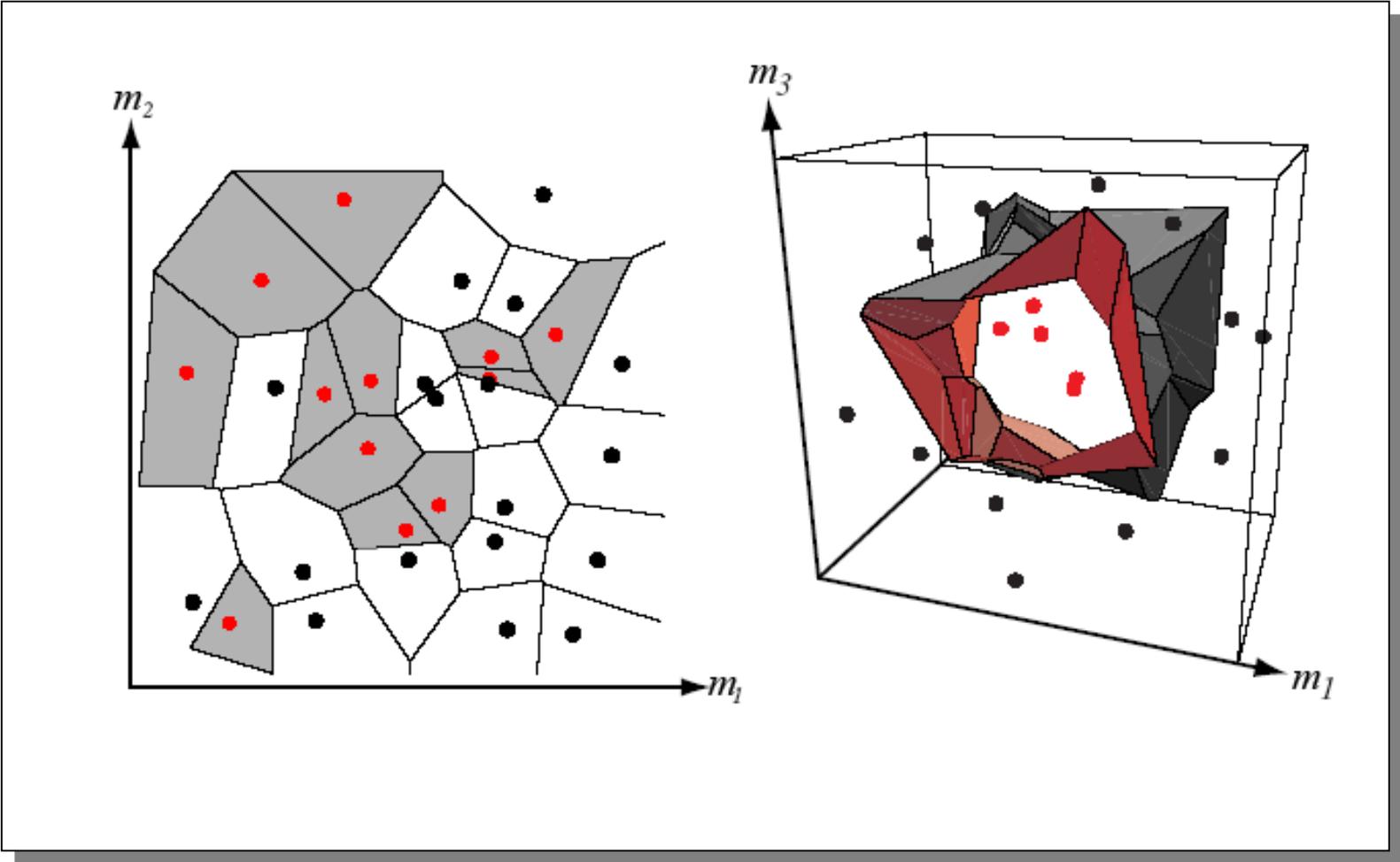
⇒ Merkmalsraum wird in  $N$  um die  $\mathbf{m}_i \in D$  zentrierte Zellen partitioniert.

→ Voronoi Mosaik

**Designfreiheitsgrad: Metrik**

# 5.3. Nächste Nachbarn Klassifikation

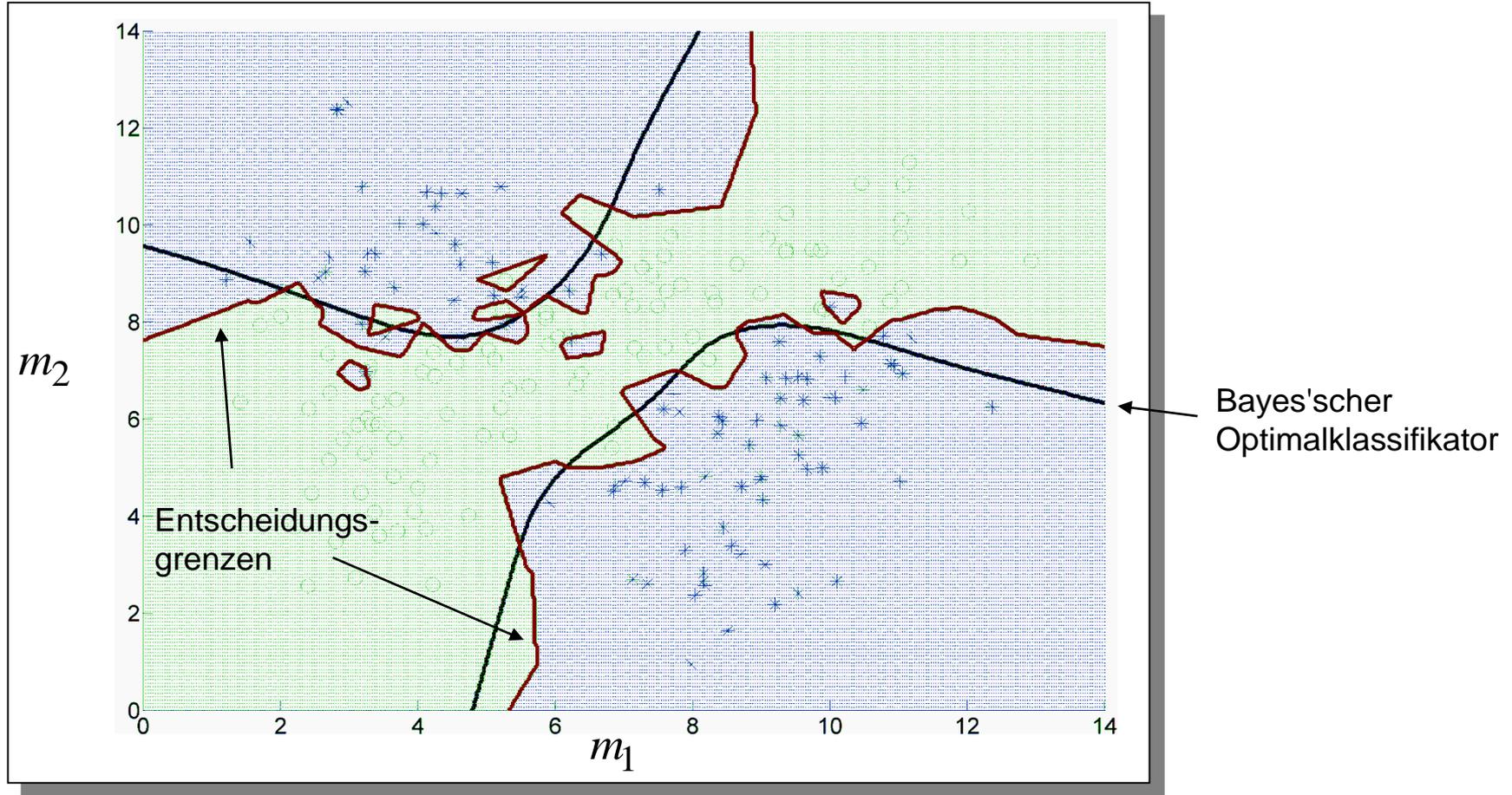
## Beispiel: Voronoi Mosaik für Euklidische Metrik



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

## 5.3. Nächste Nachbarn Klassifikation

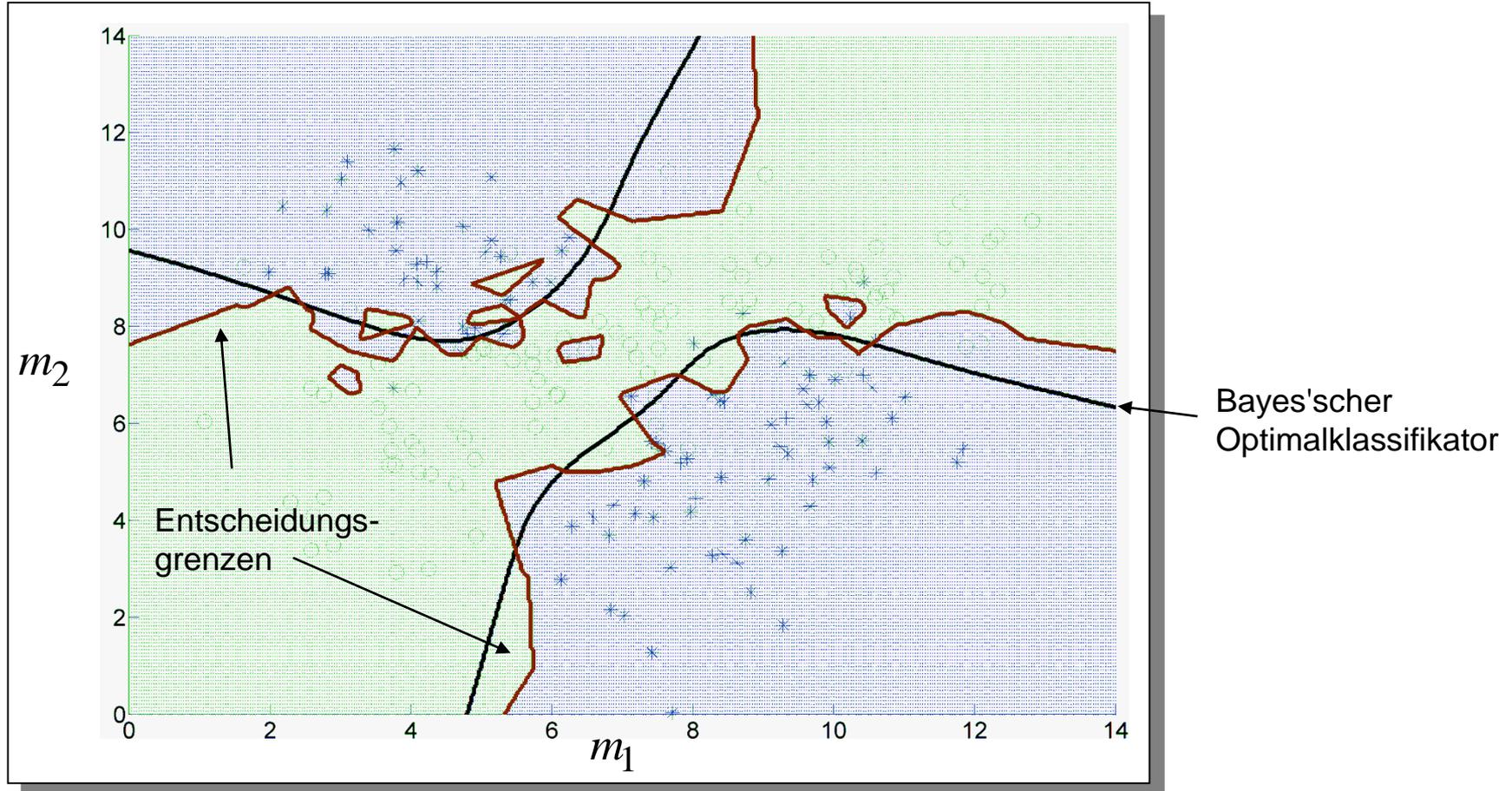
**Beispiel:** Nächste Nachbarn Klassifikator ( $k = 1$ ), Euklidische Metrik



Lernstichprobe D

## 5.3. Nächste Nachbarn Klassifikation

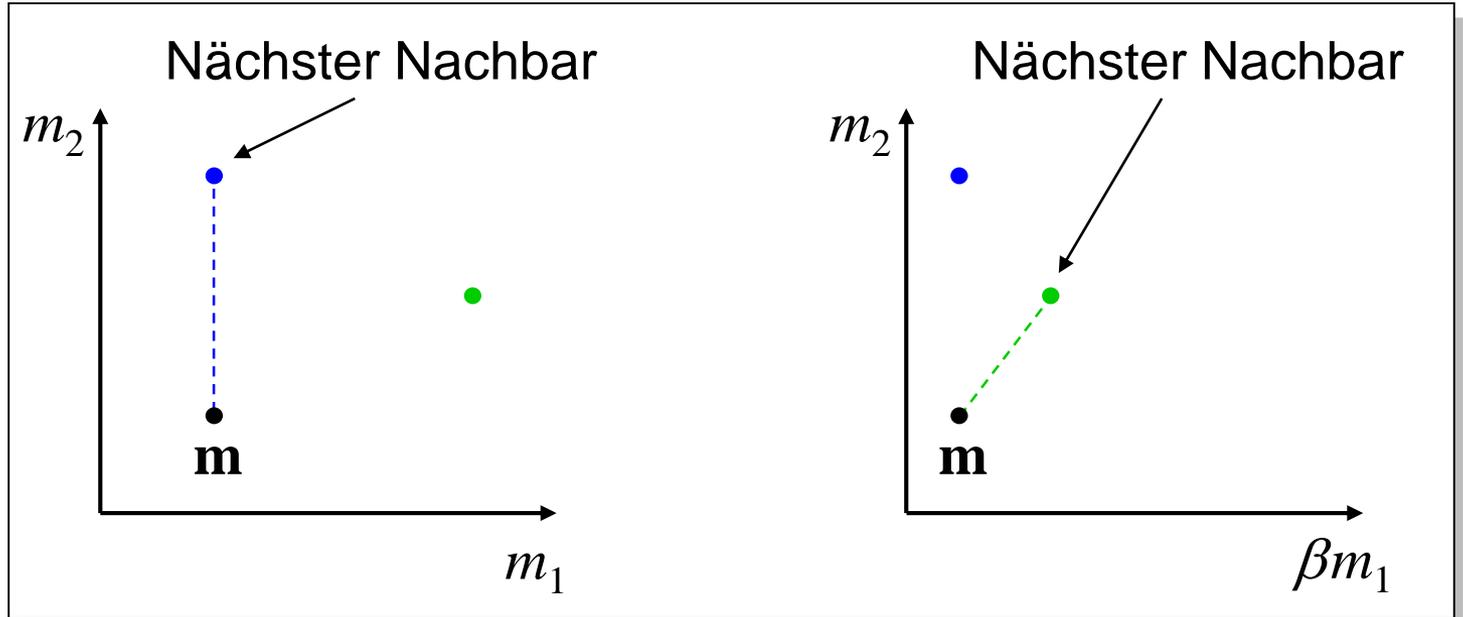
**Beispiel:** Nächste Nachbarn Klassifikator ( $k = 1$ ), Euklidische Metrik



Klassifikation der Teststichprobe  $\rightarrow$  Testfehler = 11,5%  
Asymptotischer Testfehler  $\approx$  14,48%

## 5.3. Nächste Nachbarn Klassifikation

**Bemerkung:** Skalierung von Merkmalen



## 5.3. Nächste Nachbarn Klassifikation

### Fehlerwahrscheinlichkeit:

$P$ : Fehlerwahrscheinlichkeit des Nächster Nachbar Klassifikators für  $N \rightarrow \infty$

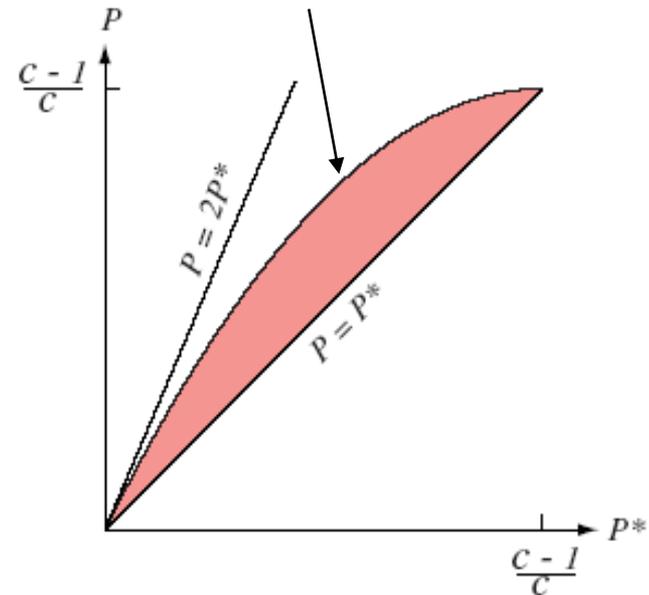
$P^*$ : Fehlerwahrscheinlichkeit des Bayes'schen Optimalklassifikators

Es gilt:

$$P^* \leq P \leq P^* \left( 2 - \frac{c}{c-1} P^* \right)$$

Asymptotisch steckt im nächsten Nachbar gewissermaßen bereits die Hälfte der für eine optimale Klassifikation notwendigen Information.

Hier kommen alle denkbaren asymptotischen Fehlerwahrscheinlichkeiten des Nächsten Nachbar Klassifikators zu liegen.



Quelle: R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification; dort kann auch die Herleitung nachgelesen werden.

## 5.3. Nächste Nachbarn Klassifikation

---

$k$  - **Nächste Nachbarn Klassifikator**:  $\mathbf{m}$  wird die am häufigsten vorgefundene Klasse innerhalb der  $k$  nächsten Nachbarn zugeordnet.

Entscheidungsfunktionen:

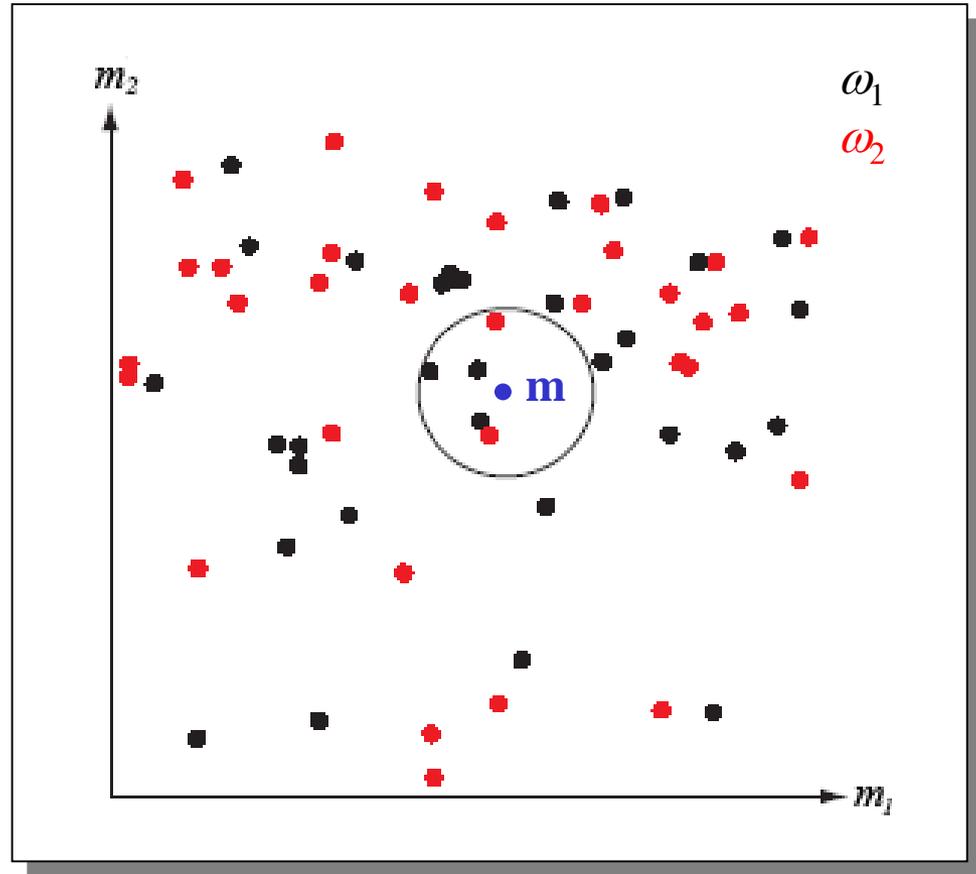
$$k_i(\mathbf{m}) := \left| \left\{ \tilde{\mathbf{m}} \mid \tilde{\mathbf{m}} \in \text{NN}_k(\mathbf{m}) \wedge \tilde{\mathbf{m}} \in D_i \right\} \right|$$

Mit  $\text{NN}_k(\mathbf{m})$ : Menge der  $k$  nächsten Nachbarn zu  $\mathbf{m}$  innerhalb  $D$ .

## 5.3. Nächste Nachbarn Klassifikation

Beispiel:  $k = 5$ ,  $c = 2$ ,  $d = 2$

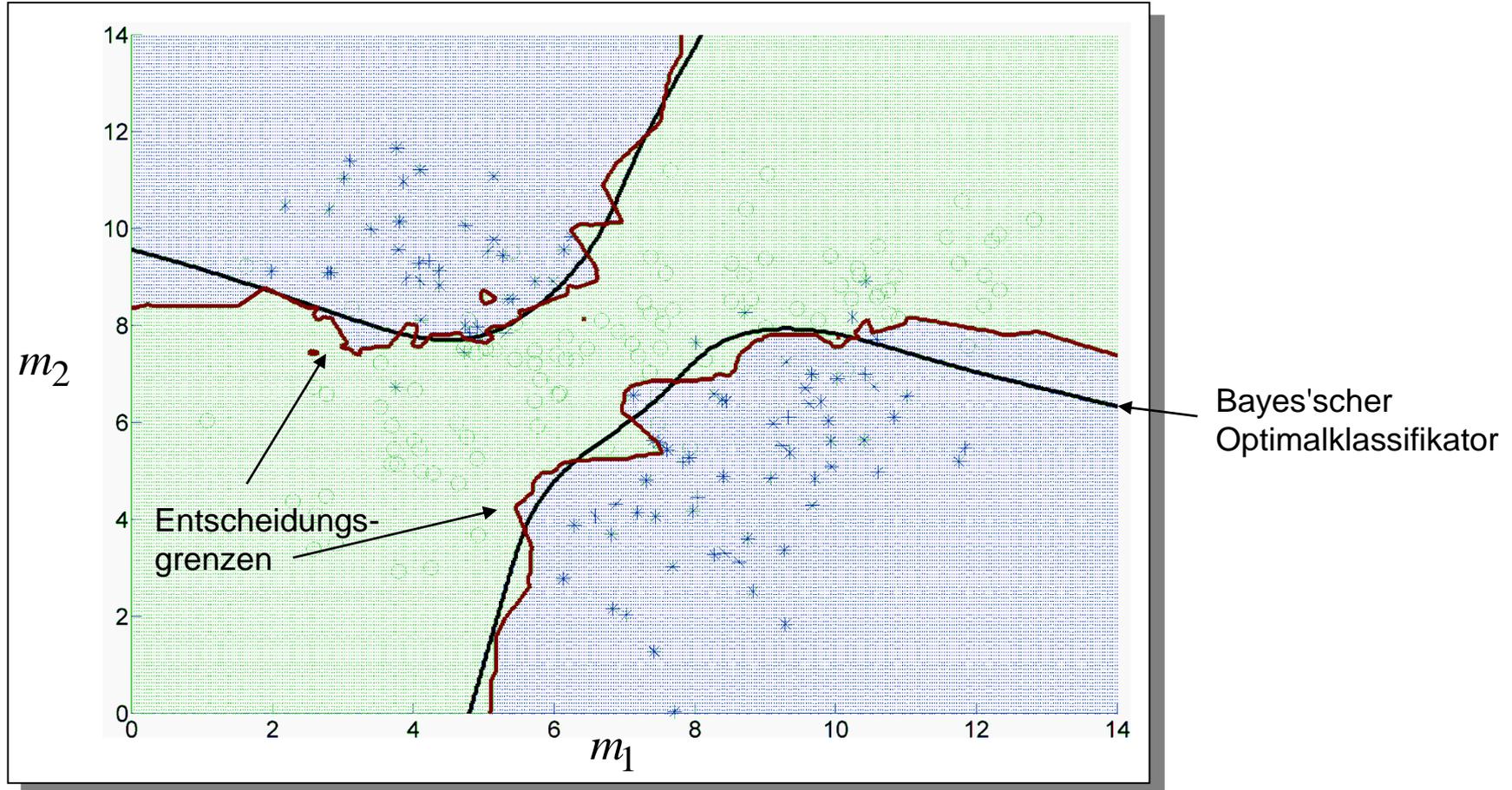
$m$  wird  $\omega_1$  zugeordnet.



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

## 5.3. Nächste Nachbarn Klassifikation

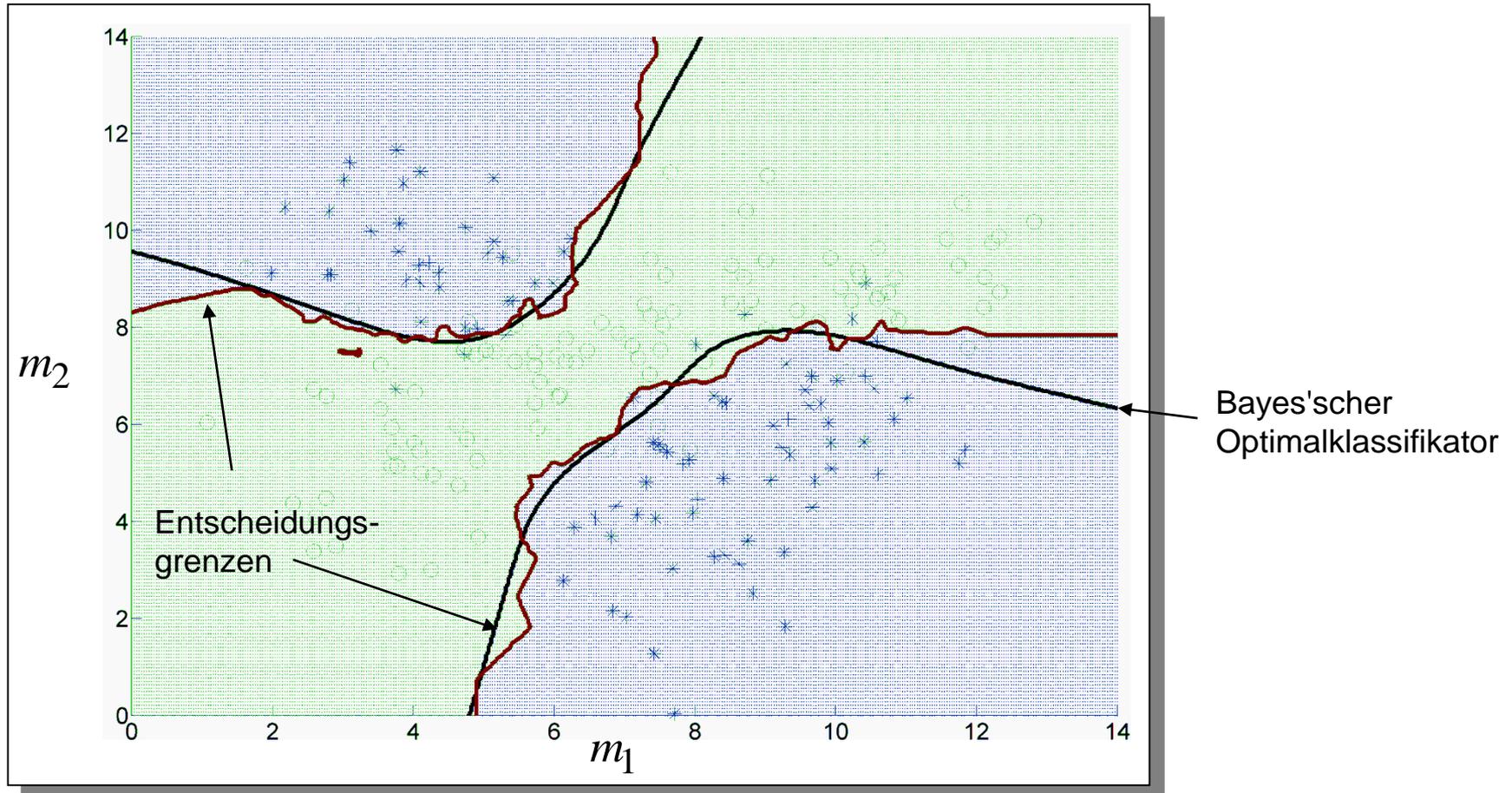
**Beispiel: 3 - Nächste Nachbarn Klassifikator ( $k = 3$ ) , Euklidische Metrik**



Klassifikation der Teststichprobe  $\rightarrow$  Testfehler = 9%  
Asymptotischer Testfehler  $\approx$  12,09%

## 5.3. Nächste Nachbarn Klassifikation

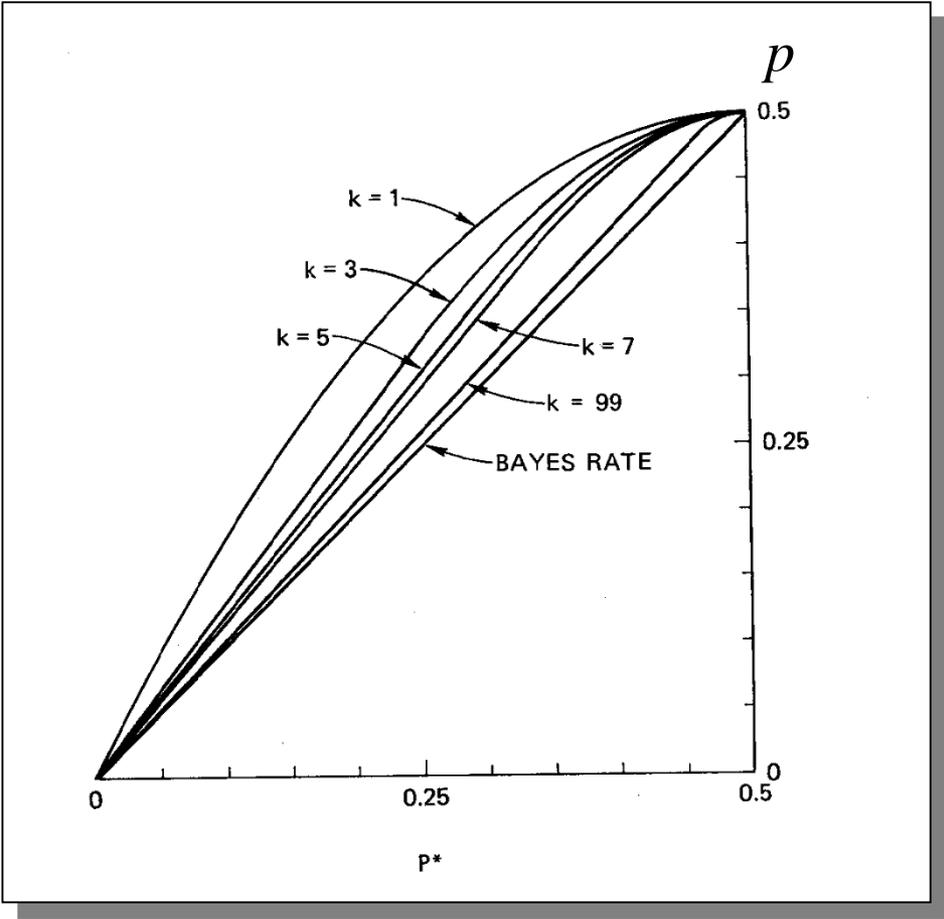
Beispiel: 5 - Nächste Nachbarn Klassifikator ( $k = 5$ ), Euklidische Metrik



Klassifikation der Teststichprobe  $\rightarrow$  Testfehler = 9,5%  
Asymptotischer Testfehler  $\approx$  11,59%

# 5.3. Nächste Nachbarn Klassifikation

Fehlerwahrscheinlichkeit:  $k$  - Nächste Nachbarn Klassifikator  $c = 2$



Quelle: R. O. Duda, P. E. Hart: Pattern Classification, 1973