# Überblick

- **Grundlagen**
  - Einführung in die automatische Mustererkennung
  - Grundlagen der Wahrscheinlichkeitsrechnung
- **Klassifikation bei bekannter Wahrscheinlichkeitsverteilung**
  - Entscheidungstheorie
  - Bayes-Klassifikator
  - Entscheidungsfunktionen bei gaußverteilten Daten
- **Überwachtes Lernen bei unbekannter Verteilung der Daten**
  - Nichtparametrische Klassifikation
  - Probleme bei hochdimensionalen Daten
  - Lineare Klassifikation, Perzeptron
  - Nicht linear trennbare Systeme
  - Nichtlineare Klassifikatoren
  - Vergleich von Klassifikatoren
- **Unüberwachtes Lernen**
  - Hauptkomponentenanalyse
  - K-Means-Clustering
  - Agglomeratives Clustern

# Entscheidungstheorie

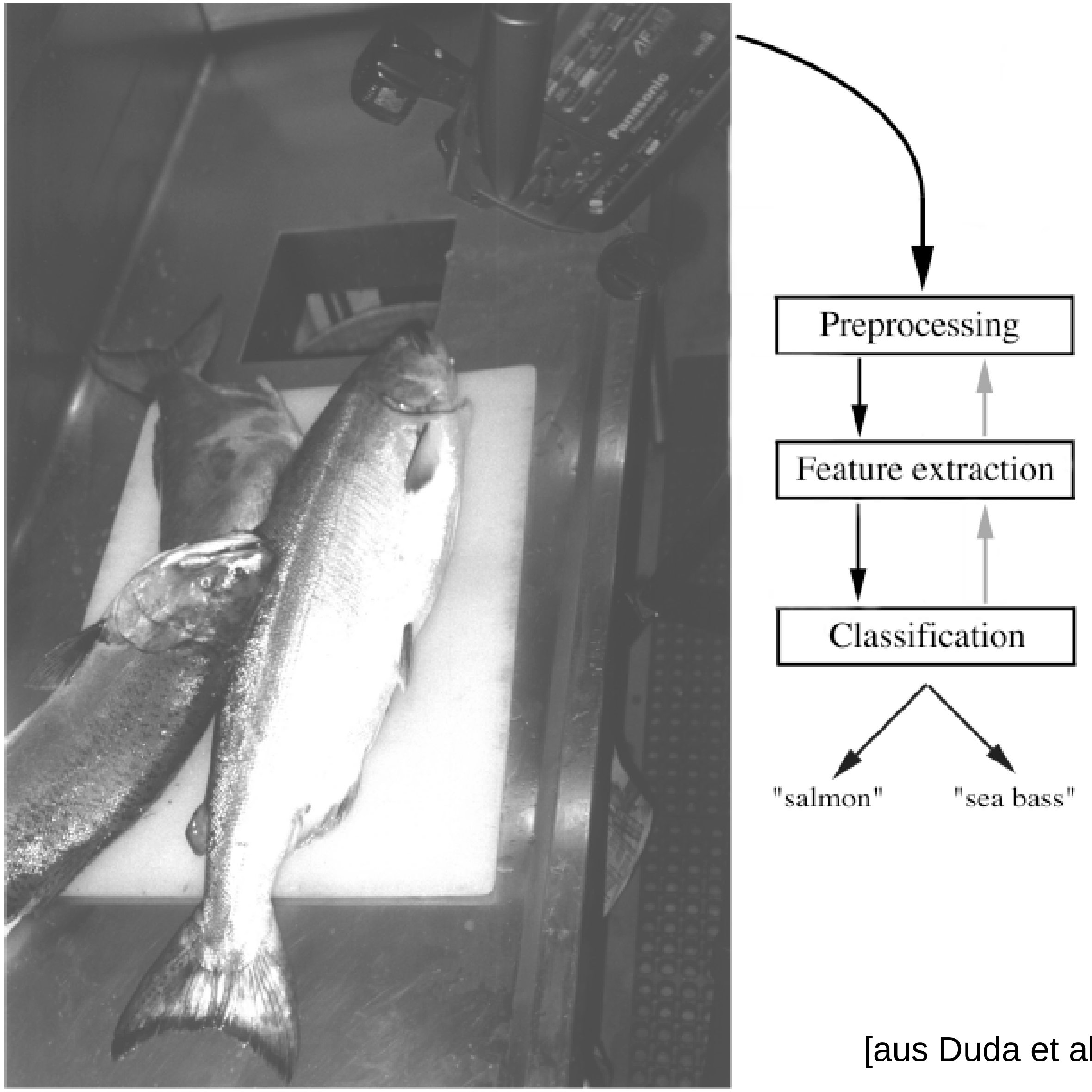
- Die Entscheidungstheorie bewertet die Kosten einer Klassifikationsentscheidung mit Hilfe von Wahrscheinlichkeitsaussagen.

- Grundannahme: Alle Wahrscheinlichkeiten seien im vornhinein bekannt.

- bildet die formale Grundlage der Mustererkennung

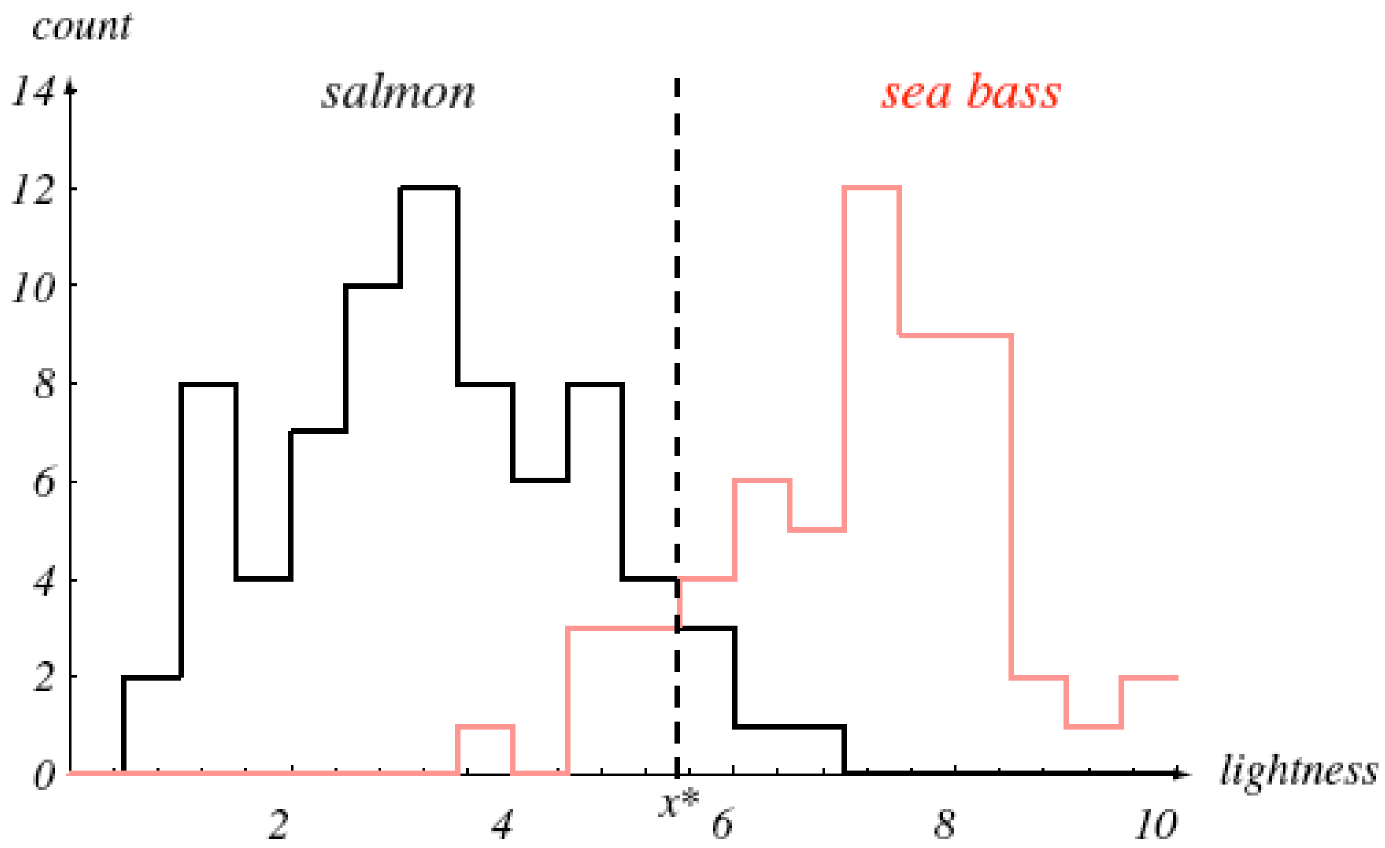
- formalisiert den „gesunden Menschenverstand“.

# Beispiel für eine Mustererkennungsaufgabe




[aus Duda et al., 2001]

# Beispiel für eine Mustererkennungsaufgabe




[aus Duda et al., 2001]

# Beispiel für entscheidungstheoretischen Ansatz

- Annahme: Reihenfolge der Fische auf dem Band sei zufällig, d.h. nicht vorhersagbar.
- In der Entscheidungstheorie heißt das: Die Natur ist in einer von zwei möglichen Zuständen $\omega$ („state of nature“):

  Lachs: $\omega = \omega_1$

  Wolfsbarsch: $\omega = \omega_2$

  Weil $\omega$ unvorhersagbar ist, muß $\omega$ als Zufallsvariable beschrieben werden.
- Annahme: Es gebe eine A-priori-Wahrscheinlichkeit $p(\omega_1)$ für Lachs und $p(\omega_2)$ für Wolfsbarsch. Bei 2 Zuständen gilt

  $$p(\omega_1) + p(\omega_2) = 1 .$$

  Die A-priori-Wahrscheinlichkeit beschreibt das Vorwissen über das Entscheidungsproblem.

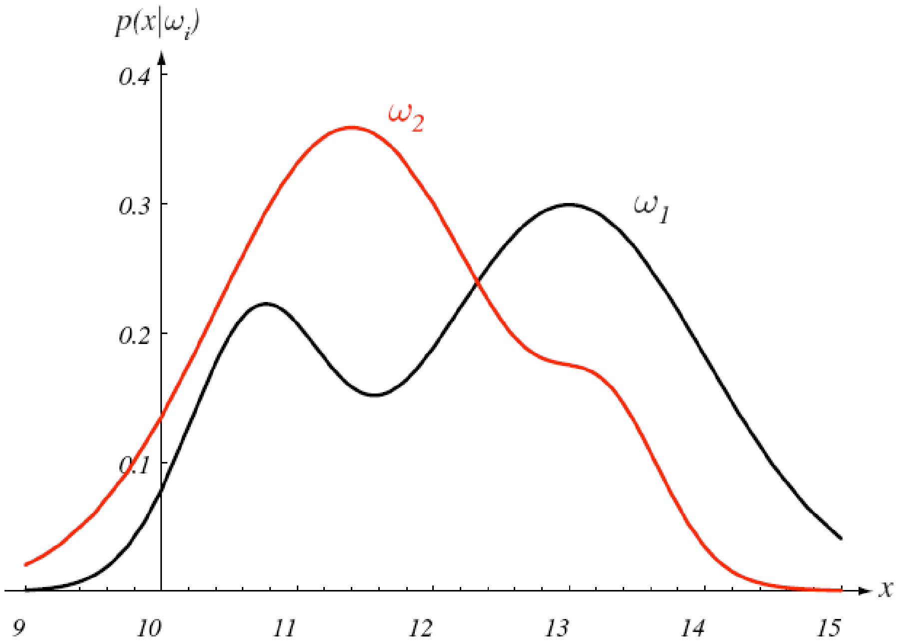
# Klassenbedingte Wahrscheinlichkeit

- Wenn jeder Fehler gleich viel kostet, und wenn Band nicht beobachtet werden kann, ist die beste Entscheidungsregel:

  „Entscheide für $\omega_1$ falls $p(\omega_1) > p(\omega_2)$, sonst für $\omega_2$"

- Gibt es (fehlerbehaftete) Messungen eines diskriminativen Merkmals x, dann hängt die Messung vom Naturzustand $\omega$ ab. Beschreibung durch bedingte Wahrscheinlichkeit

  $p(x|\omega_1)$ (klassenbedingte Wahrscheinlichkeit)



# Satz von Bayes

Wie beeinflußt die Messung x die Schätzung des tatsächlichen Naturzustandes?

Gemeinsame Verteilung von $\omega_j$ und x:

$$p(\omega_j, x) = p(x|\omega_j)\, p(x|\omega_j) = p(\omega_1|x)\, p(x) \quad \text{(Kettenregel)}$$

Satz von Bayes:

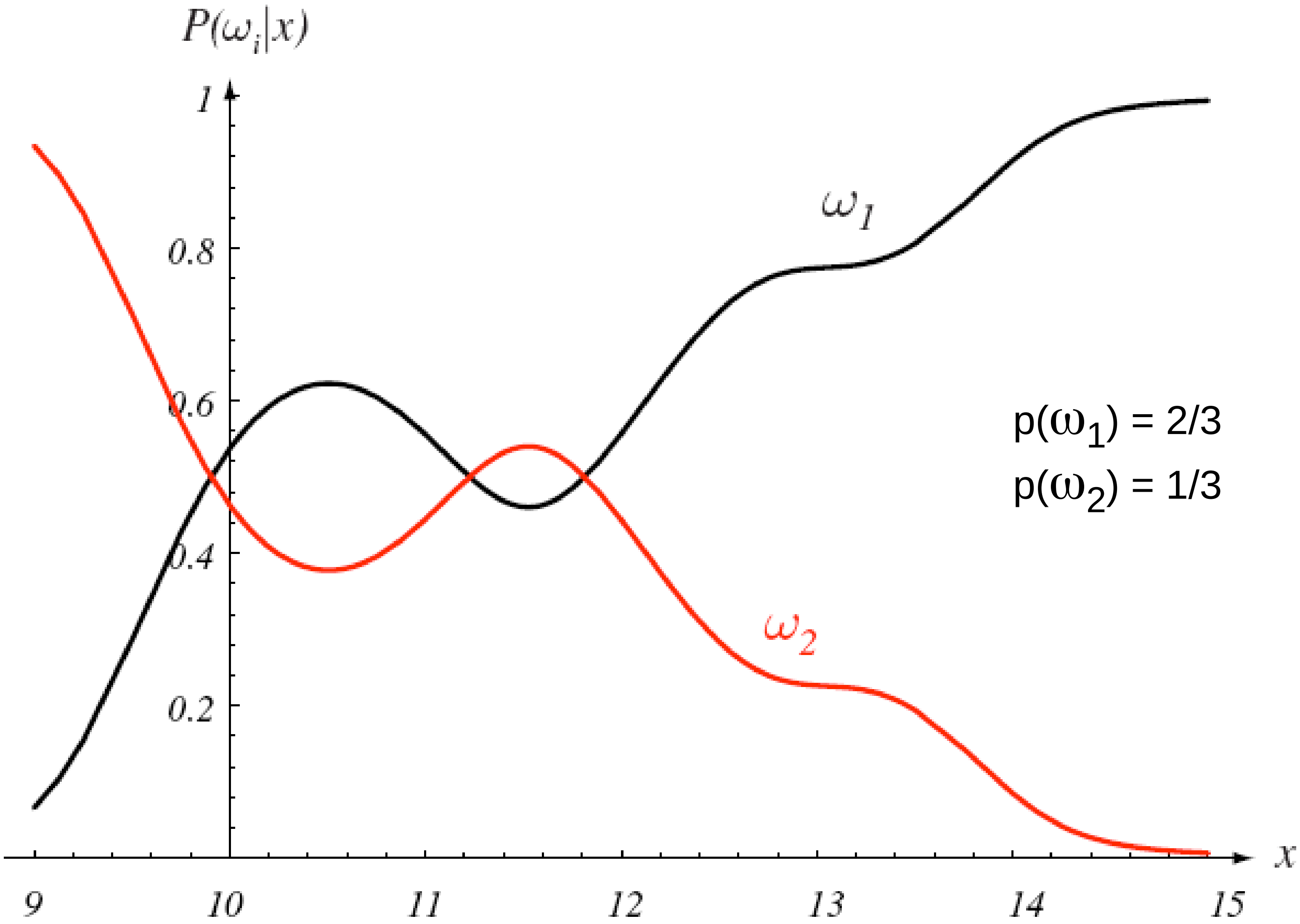
$$p(\omega_j \mid x) = \frac{p(x \mid \omega_j)\, p(\omega_j)}{p(x)}$$

mit $p(x) = \sum_{j=1}^{2} p(x \mid \omega_j)\, p(\omega_j)$

In Worten:

$$A - posteriori - Wahrscheinlichkeit = \frac{Likelihood \times A - priori - Wahrscheinlichkeit}{Evidenz}$$

# A-posteriori-Wahrscheinlichkeit (Beispiel)




$p(\omega_1) = 2/3$

$p(\omega_2) = 1/3$

# Bayessche Entscheidungsregel

Nach einer Messung x ist die Fehlerwahrscheinlichkeit

$$p(Fehler \mid x) = \begin{cases} p(\omega_1 \mid x) & bei\ Entscheidung\ für\ \omega_1 \\ p(\omega_2 \mid x) & bei\ Entscheidung\ für\ \omega_2 \end{cases}$$

Die Fehlerwahrscheinlichkeit für gegebenes x wird minimiert durch die Bayessche Entscheidungsregel:

Entscheide für $\omega_1$ wenn $p(\omega_1|x) > p(\omega_2|x)$, sonst für $\omega_2$ bzw.

Entscheide für $\omega_1$ wenn $p(x|\omega_1)\, p(\omega_1) > p(x|\omega_2)\, p(\omega_2)$ , sonst für $\omega_2$

Die Bayessche Entscheidungsregel minimiert auch die durchschnittliche Fehlerwahrscheinlichkeit über alle x:

$$p(Fehler) = \frac{1}{|X|} \sum_{x \in X} p(Fehler, x) = \frac{1}{|X|} \sum_{x \in X} p(Fehler \mid x)\, p(x)$$

# Aufgabe

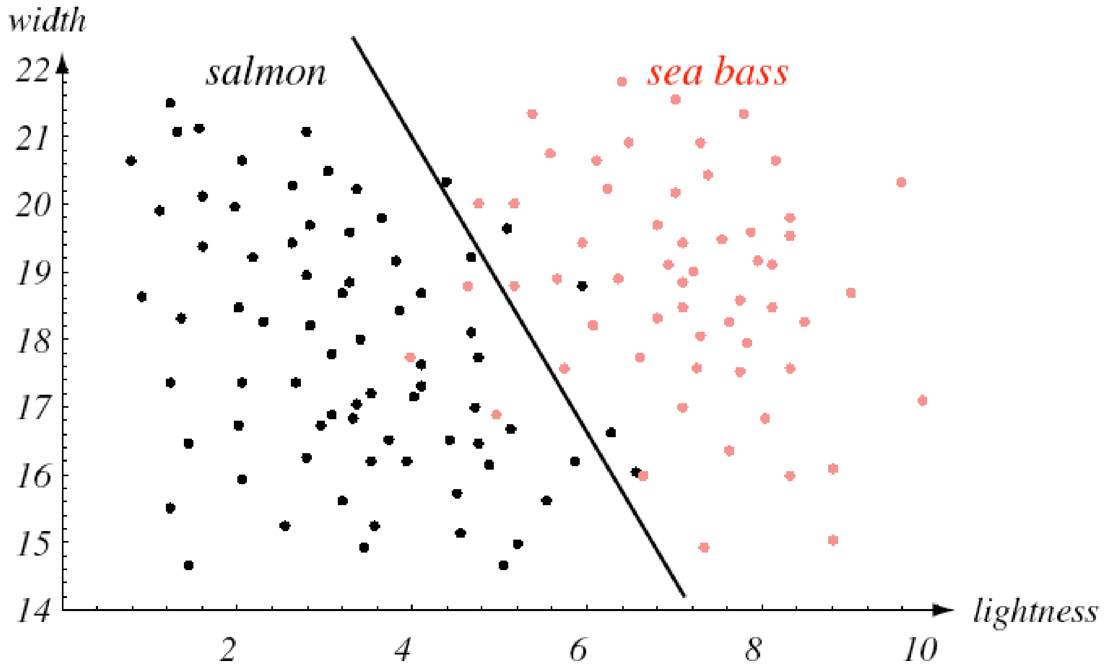
1. Sie kommen auf eine Insel, von der Sie wissen, daß die Einwohner in 2/3 aller Fälle lügen. An einer Weggabelung angekommen, wissen Sie nicht mehr weiter. Sie schätzen, daß Ihr Ziel mit 60% Wahrscheinlichkeit links liegt und fragen sicherheitshalber einen Einheimischen. Dieser sagt „rechts“. Wie entscheiden Sie sich?
2. Nun fragen Sie einen weiteren Einheimischen, ober der Erste gelogen hat. Dieser antwortet mit „Nein“. Wie wahrscheinlich ist es, daß der Erste tatsächlich gelogen hat? Ändern Sie Ihre Entscheidung nach dieser Auskunft?

# Bayessche Entscheidungstheorie

Bisherige Betrachtungen werden verallgemeinert:

- mehr als ein Merkmal
- mehr als zwei Zustände
- mehr als zwei Entscheidungsoptionen
- allgemeinere Kostenfunktionen als die Fehlerwahrscheinlichkeit

# Erweiterung auf mehr als eine Merkmal




[aus Duda et al., 2001]

Erweiterung auf d Messungen ist einfach:

Ersetze die einzelne Messung x durch einen Merkmalsvektor x aus einem d-dimensionalen euklidischen Raum, dem Merkmalsraum X.

Alle Formeln und Methoden bleiben gleich!

# Verallgemeinerung der Entscheidungssituation

2. Es kann sinnvoll, sich für mehr als nur 2 Optionen zu entscheiden, z.B
   - bei unklarer Situation sich für keine der Alternativen zu entscheiden
   - bei mehr als 2 Klassen

Allgemein erlauben wir eine Menge von $a$ möglichen Aktionen $\{\alpha_1, \alpha_2, ..., \alpha_a\}$.

3. Bei vielen Mustererkennungsproblemen gibt es mehr als 2 Klassen, d.h. wir erlauben eine Menge von $c$ möglichen Naturzuständen $\{\omega_1, \omega_2, ..., \omega_a\}$ (Mehrklassenproblem).

4. Die Kostenfunktion $\lambda(\alpha_i \mid \omega_j)$ bemißt die Kosten der Aktion $\alpha_i$, wenn die „Natur“ in Zustand $\omega_j$ ist.

# Verallgemeinerte Bayessche Entscheidungsregel

Bayes-Formel:

$$p(\omega_j \mid \mathbf{x}) = \frac{p(x \mid \omega_j)\,p(\omega_j)}{p(\mathbf{x})}$$

mit der Evidenz $$p(\mathbf{x}) = \sum_{\lambda=1}^{c} p(\mathbf{x} \mid \omega_j)\,p(\omega_j)$$

Die erwarteten Kosten für Aktion $\alpha_i$ (das bedingte Risiko von $\alpha_i$):

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j)\,p(\omega_j \mid \mathbf{x})$$

Verallgemeinerte Entscheidungsregel (minimiert das erwartete Gesamtrisiko):

„Wähle immer die Aktion $\alpha_i$, die das bedingte Risiko für die gegebene Beobachtung **x** minimiert.“

# Spezialfall: 2 Klassen, 2 Aktionen

Vereinfachte Kostenfunktion: $\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$

Bedingtes Risiko: $R(\alpha_1 \mid \mathbf{x}) = \lambda_{11} p(\omega_1 \mid \mathbf{x}) + \lambda_{12} p(\omega_2 \mid \mathbf{x})$

$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21} p(\omega_1 \mid \mathbf{x}) + \lambda_{22} p(\omega_2 \mid \mathbf{x})$$

Entscheidungsregel: Entscheide für , wenn

$$(\lambda_{21} - \lambda_{11}) p(\omega_1 \mid \mathbf{x}) > (\lambda_{12} - \lambda_{22}) p(\omega_2 \mid \mathbf{x})$$

bzw.

$$(\lambda_{21} - \lambda_{11}) p(\mathbf{x} \mid \omega_1) p(\omega_1) > (\lambda_{12} - \lambda_{22}) p(\mathbf{x} \mid \omega_2) p(\omega_2)$$

Unter der Annahme, daß $\lambda_{21} > \lambda_{11}$ , ergibt sich

$$\boxed{\frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{p(\omega_2)}{p(\omega_1)}}$$

(likelihood ratio)

# Klassifikation mit minimaler Fehlerrate

Klassifikation:

- Jeder Zustand entspricht einer Klasse
- Jede Aktion entspricht der Entscheidung, daß eine bestimmte Klasse vorliegt
- Entscheidung ist richtig, wenn die angenommene Klasse mit der tatsächlichen übereinstimmt, sonst ist sie falsch

Zugehörige Kostenfunktion:

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \qquad i, j = 1, \dots, c$$

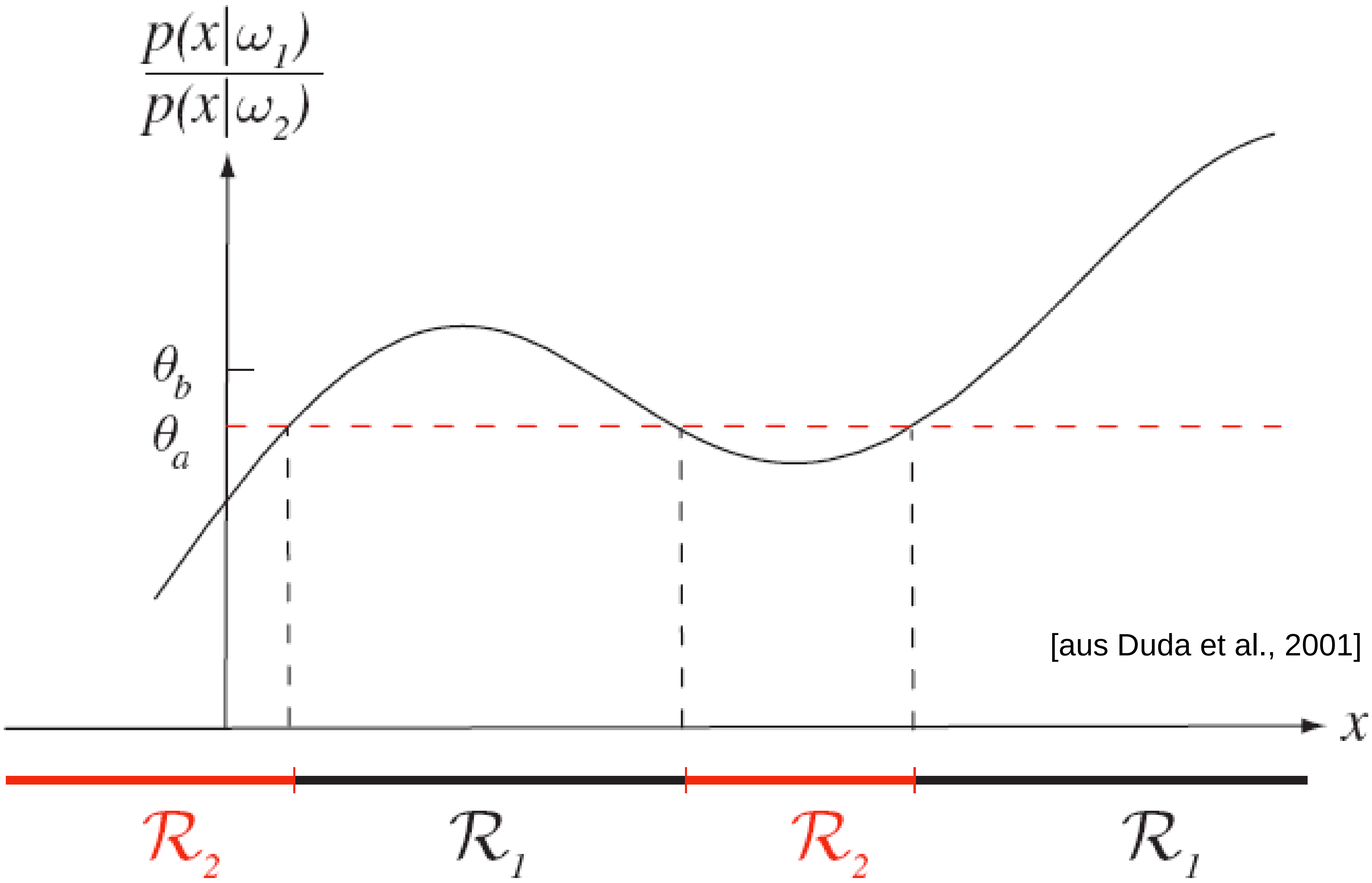
(symmetrische Kostenfunktion oder zero-one-loss)

Bedingtes Risiko: $R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) p(\omega_j \mid \mathbf{x}) = 1 - p(\omega_j \mid \mathbf{x})$

Damit ergibt sich die schon bekannte Entscheidungsregel zur Fehlerminimierung:

„Entscheide für $\omega_1$ wenn $p(\omega_1|x) > p(\omega_2|x)$, sonst für $\omega_2$“

# Likelihood ratio (Beispiel)




[aus Duda et al., 2001]