

Syntaktische und Statistische Mustererkennung

VO 1.0 840.040
(UE 1.0 840.041)

Bernhard Jung

bernhard@jung.name
<http://bernhard.jung.name/VUSSME/>

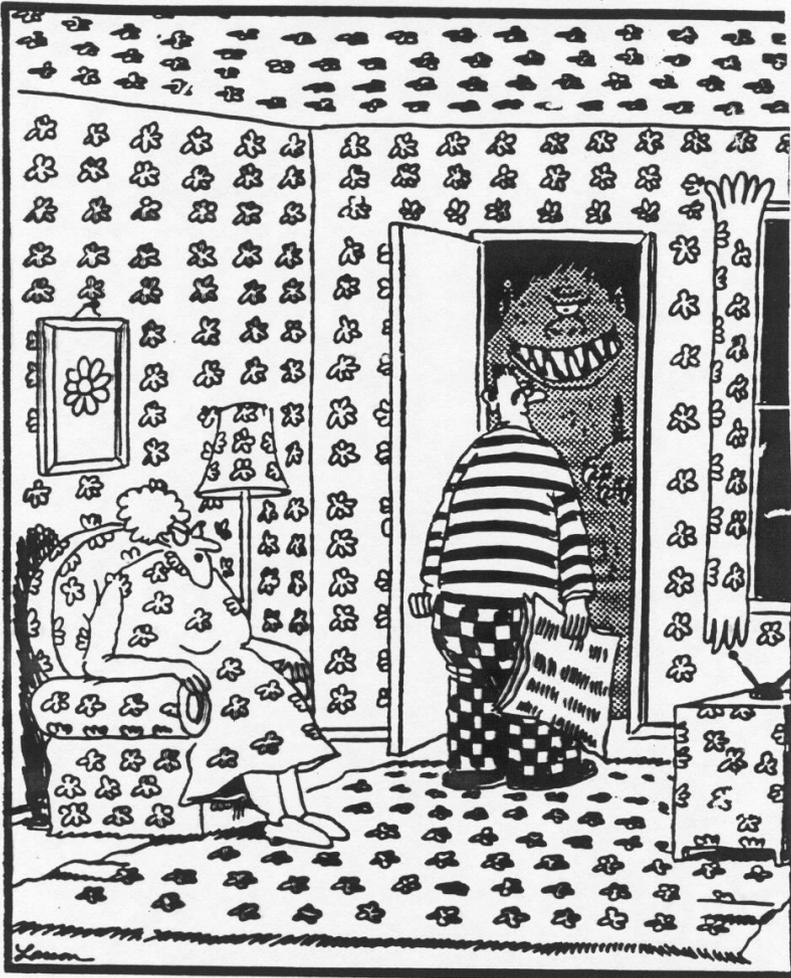
Organisatorisches

Donnerstags, 13-15 (c.t.), Seminarraum des IAI

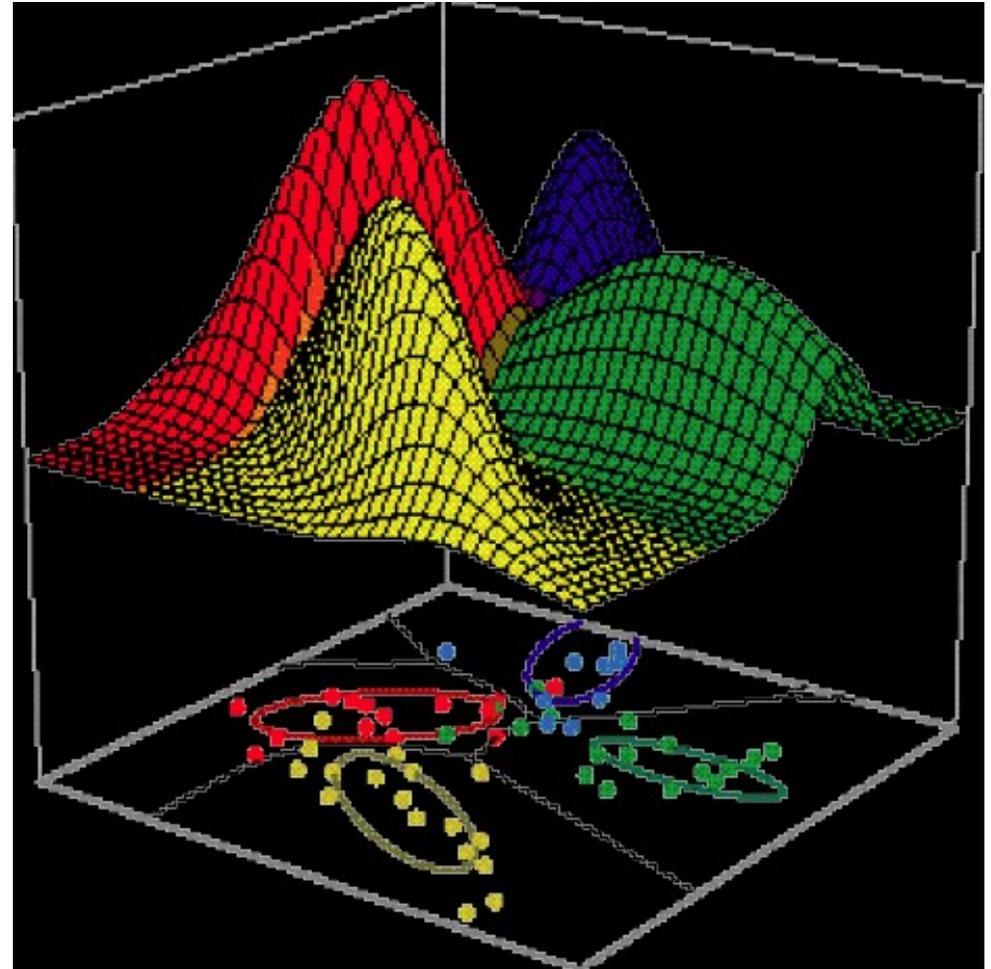
| | | |
|----------|----|--|
| 13.10.16 | VO | 1. Organisatorisches, Einführung, der Mustererkennungsprozess |
| 20.10.16 | VO | 2. Abstandsmaße; Entscheidungstabellen/-bäume |
| 27.10.16 | | |
| 3.11.16 | VO | 3. Entscheidungstheorie, Lineare Klassifikation |
| 10.11.16 | | |
| 17.11.16 | UE | Übung 1 |
| 24.11.16 | VO | 4. Nichtlineare Klassifikation, Clustering |
| 1.12.16 | VO | 5. Graphical models |
| 8.12.16 | | FEIERTAG |
| 15.12.16 | VO | 6. Syntaktische Mustererkennung: Merkmale, Repräsentation, Grammatiken: Parsing und Lernen |
| 22.12.16 | UE | Übung 2 |
| 29.12.16 | | FERIEN |
| 5.1.17 | | FERIEN |
| 12.1.17 | VO | 7. Part-based models, Kombinationen struktureller und statistischer Erkennung; Zusammenfassung |
| 19.1.17 | UE | Übung 3 |
| 26.1.17 | VO | Prüfung |

Mustererkennung

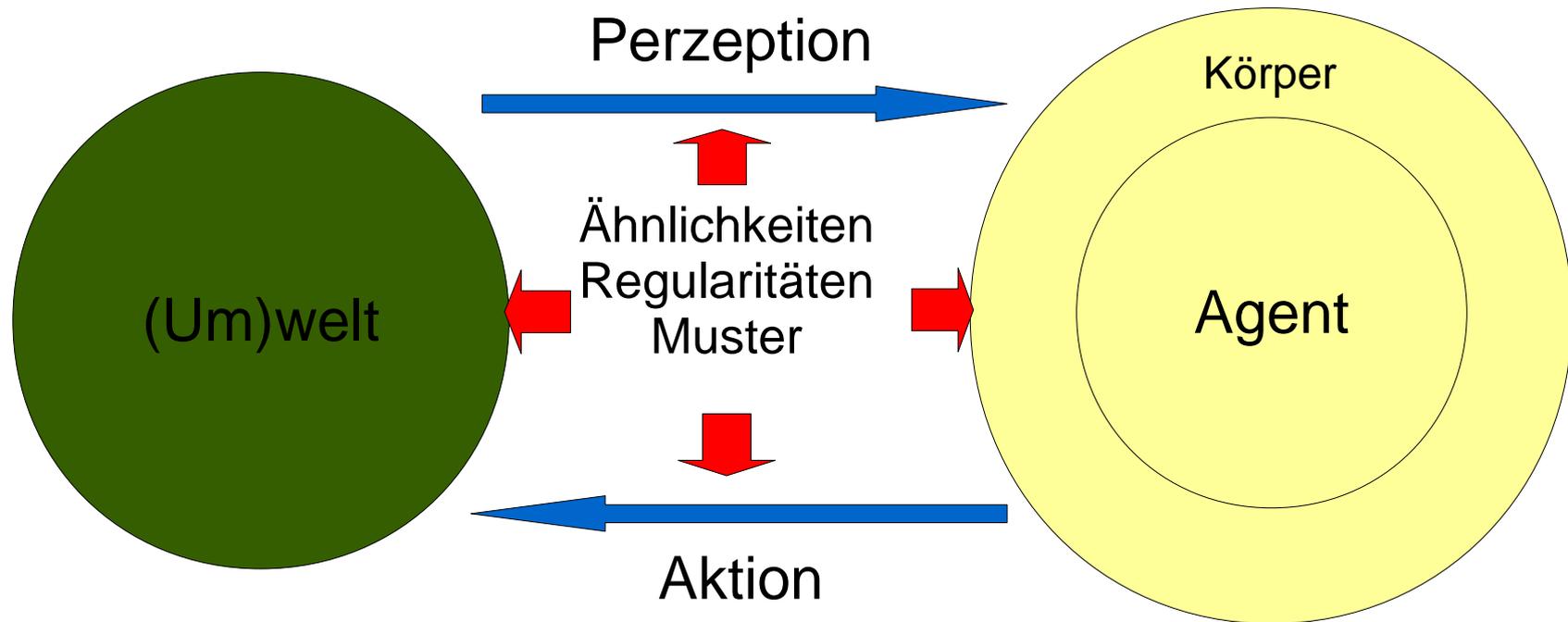
Mustererkennung



When the monster came, Lola, like the peppered moth and the arctic hare, remained motionless and undetected. Harold, of course, was immediately devoured.



Agent-Umwelt-Interaktion



Was ist ein Muster?

Anordnung aus **einfachen Bestandteilen**, die in ihrer **Anordnung** und ihrer Beziehung einen bestimmten, **anwendungsabhängigen** Sinn ergeben. Muster weisen Gesetzmäßigkeiten auf, haben **Struktur**.

Beispiele:

Buchstaben, Worte, Sätze

Bilder, Gesichter

EKG-Signale

Patientenakten, Krankheitsverläufe

Mustererkennung

Mathematisch-technische Aspekte der automatischen Verarbeitung und Auswertung von Mustern
(ohne notwendigerweise biologische Mechanismen zu berücksichtigen)

Die **automatische Berechnung** von Aussagen über Tatsachen, Sachverhalte, Teile der Welt

Aufgabenspezifische Beschreibung von Sensordaten

Forschungsfragestellungen

Optimalität von Lösungen

Theoretische Betrachtung von Problemklassen und derer Charakteristika

Konkret (ingenieurmäßig) realisierbare Lösungen

Wie lassen sich wirtschaftlich, medizinisch, ... relevante Probleme effizient, effektiv, optimal lösen?

Beispiele für Anwendungen von Mustererkennung

Schrift

Postanschriftleser

Belegleser

OCR Systeme

Sprache

Maschinelle Übersetzung

Spracherkennung

Fernerkundung

Überwachung

Proteomics

Biometrie

Gesichtserkennung

Fingerabdruck

Medizin

EEG, EKG

Klassifikation & Lokalisation in Bildern

Wirtschaft

Trends

(Industrielle) Automatisierung

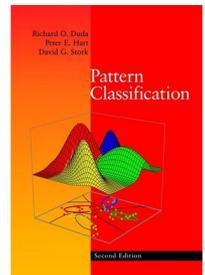
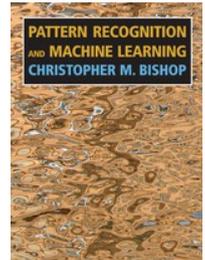
Verwandte Gebiete?

- Pattern Recognition
 - Methoden und Algorithmen zur Klassifikation
 - Verankert in Statistik und Entscheidungstheorie
 - Design von praktischen Systemen
- Machine Learning
 - Lernen aus Daten statt Programmierung
- Data Mining
 - Suche nach relevanter Information in strukturierten Datenbanken

Weiterführende Information

- Literatur

- Bishop C., Pattern Recognition and Machine Learning, 2006.
- Duda R. et al., Pattern Classification, 2nd Edition, 2000.



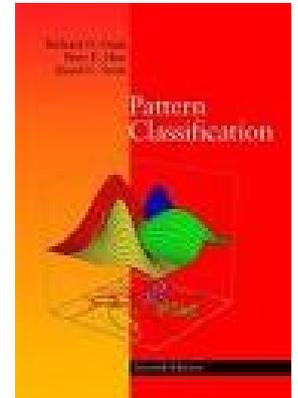
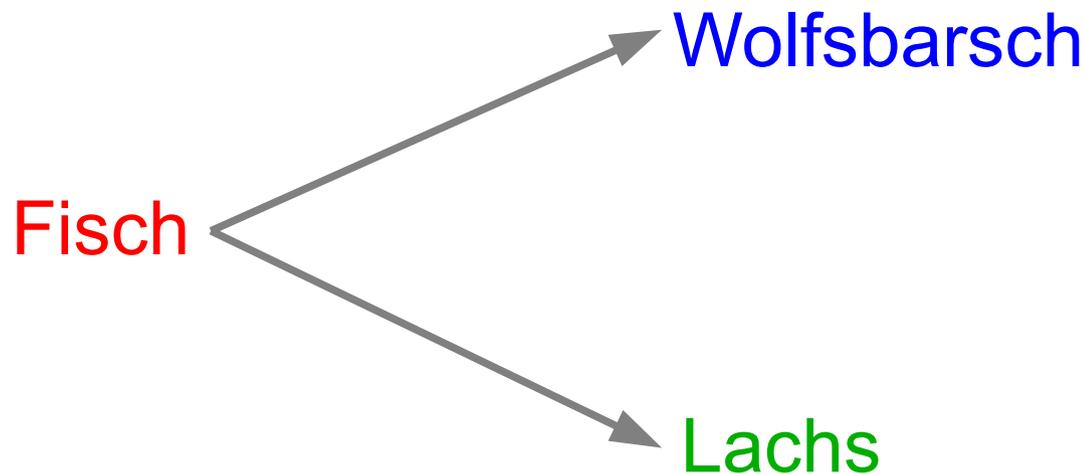
Weiterführende Information

- (Free) Online Courses (Coursera, Udacity, ...)
 - Geoffrey Hinton, Neural Networks for Machine Learning,
<https://www.coursera.org/course/neuralnets>
 - Andrew Ng, Machine Learning
<https://www.coursera.org/learn/machine-learning>
 - Peter Norvig, Sebastian Thrun: Intro to Artificial Intelligence, Learn the Fundamentals of AI
<https://www.udacity.com/course/intro-to-artificial-intelligence--cs271>
 - Charles Isbell, Michael Littman: Machine Learning 1/2/3: Supervised Learning, Unsupervised Learning, Reinforcement Learning
<https://www.udacity.com/course/machine-learning-supervised-learning--ud675>

Ein typischer Mustererkennungsprozess

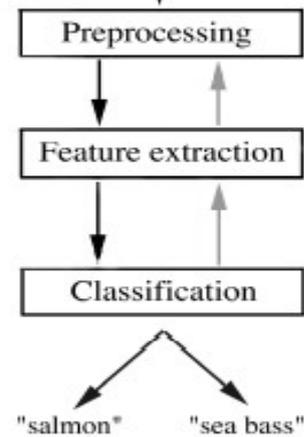
Wolfsbarsch oder Lachs?

Aufgabe: Sortierung von Fischen
auf einem Förderband
nach der Spezies
mit Hilfe optischer Sensoren



aus: Duda, Hart, Stork:
Pattern Classification.
Wiley & Sons. 2000.

Wolfsbarsch oder Lachs?



Problemanalyse

Aufnahme von Beispielbildern und Extraktion von Merkmalen:

- Länge
- Helligkeit
- Breite
- Anzahl und Form der Flossen
- Position des Mauls, etc.

Menge möglicher Merkmale für die Verwendung zur Klassifikation!

Vorverarbeitung

Segmentationsverfahren

- Zur Isolation einzelner Fische im Bild von anderen Fischen und dem Hintergrund

Merkmalsextraktion

- Für jeden einzelnen Fisch

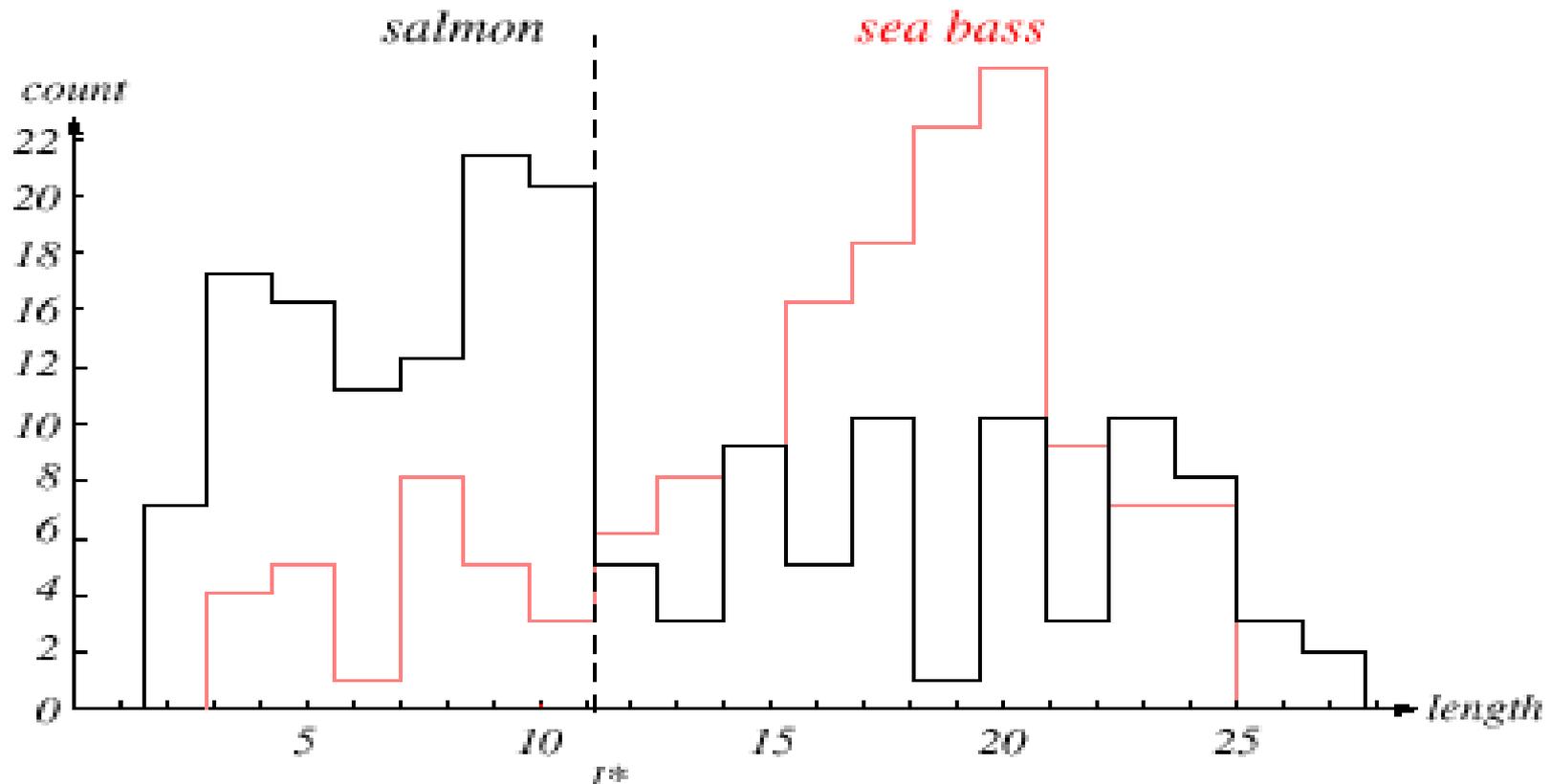
Reduktion der Daten durch Messung bestimmter Merkmale

Klassifikation

- Basierend auf den extrahierten Merkmalen

Klassifikation: Länge

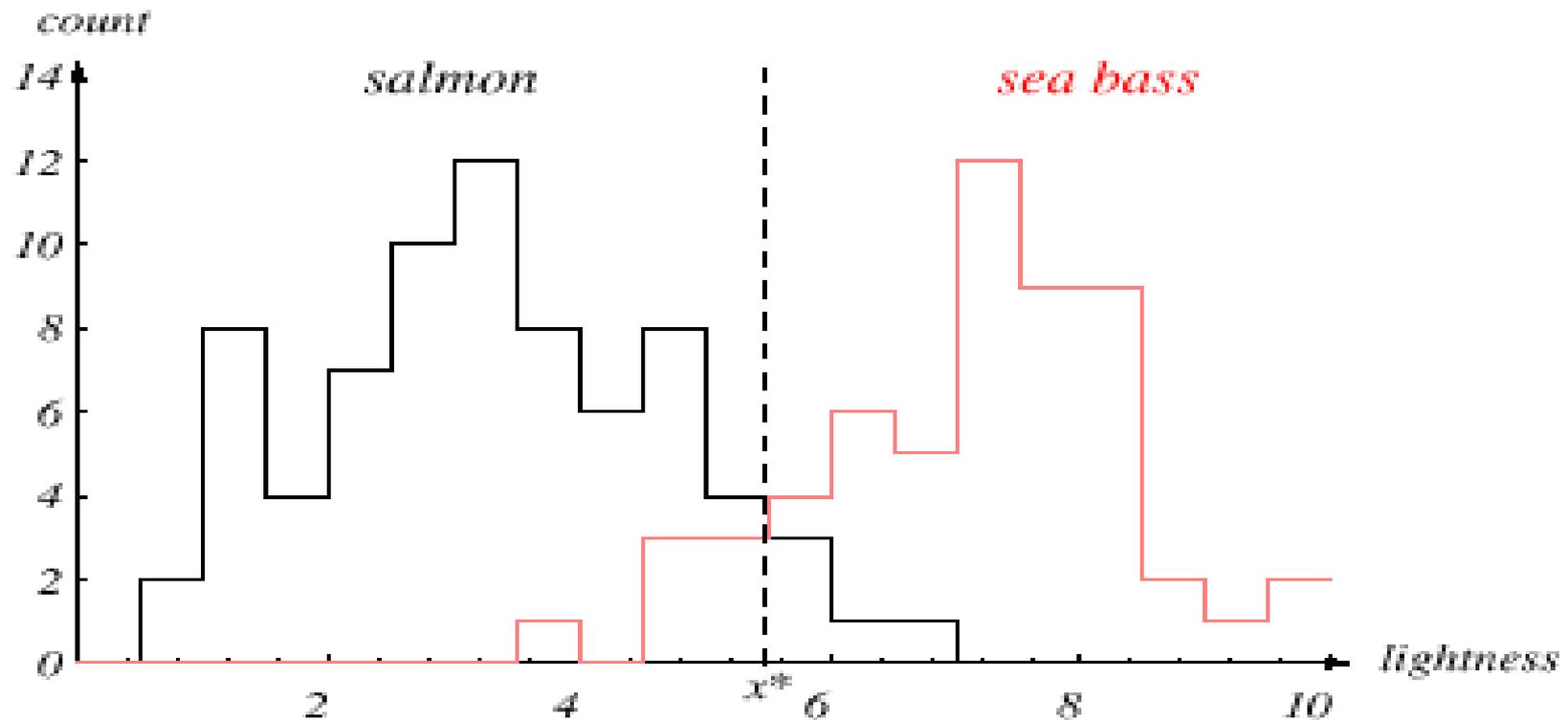
Anhand des Merkmals „Länge“ als mögliches Unterscheidungsmerkmal



Klassifikation: Helligkeit

Erkenntnis: Die Länge alleine ist ein schlechtes Merkmal!

Mögliches anderes Merkmal: Helligkeit



Entscheidungsgrenze und Kosten

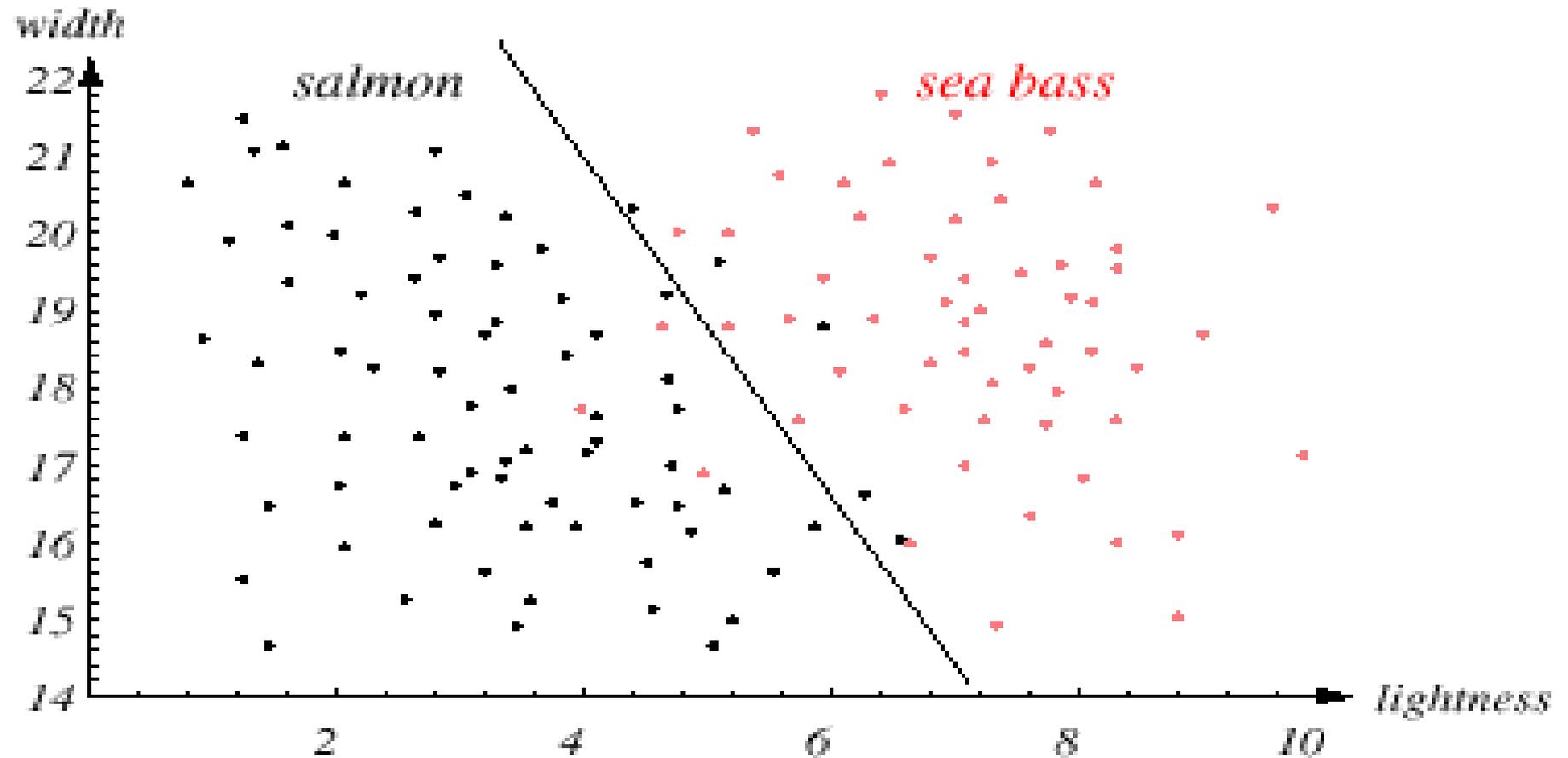
Beziehung zwischen

- Entscheidungsgrenze (Decision Boundary)
- Kosten

Verschiebung der Entscheidungsgrenze in Richtung kleinerer Werte von Helligkeit reduziert die Kosten (=Anzahl Wolfsbarsche klassifiziert als Lachs!)

→ Aufgabe der Entscheidungstheorie (Decision Theory)

Länge und Helligkeit

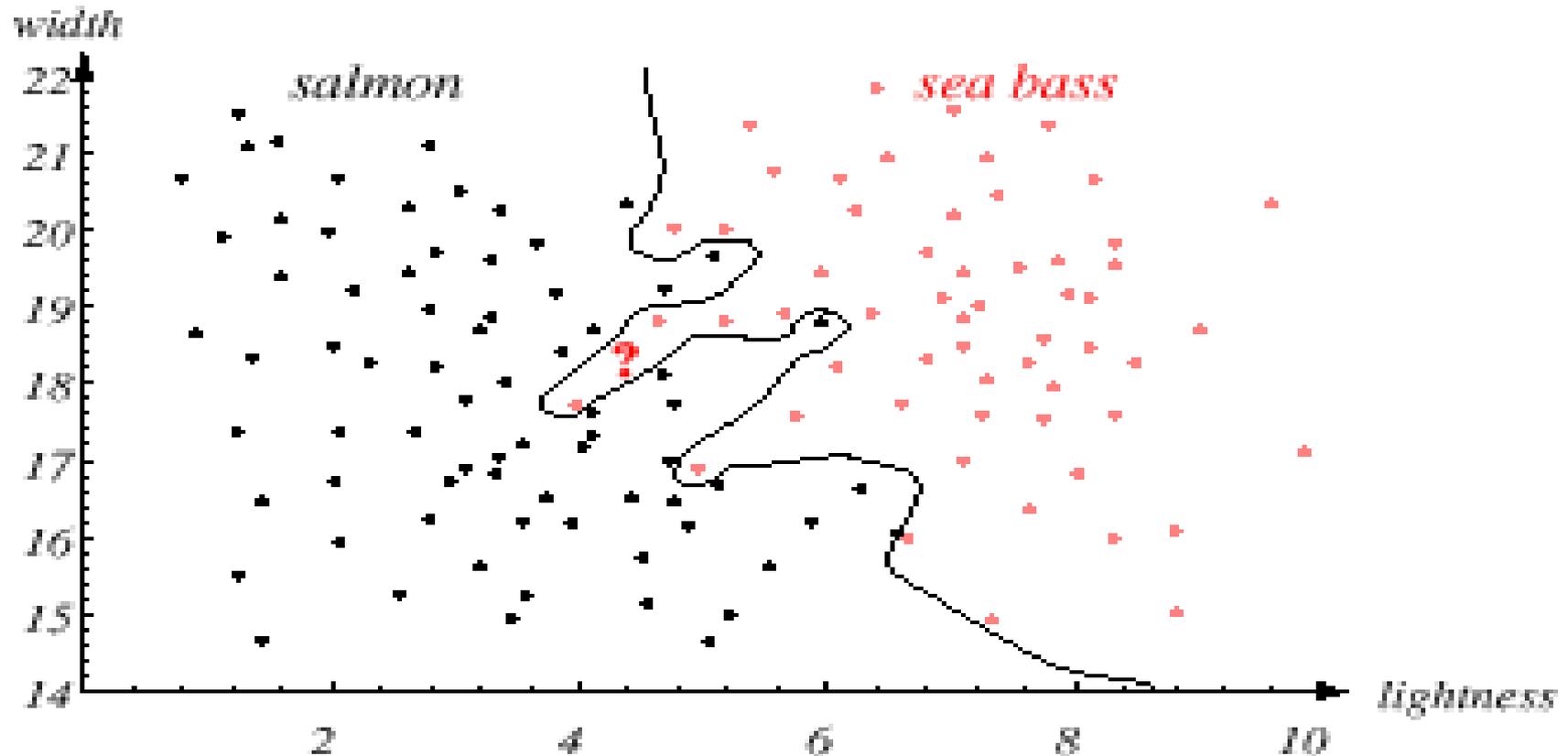


Weitere Merkmale?

Sofern sie zusätzliche Information zur Klassifikation liefern ...

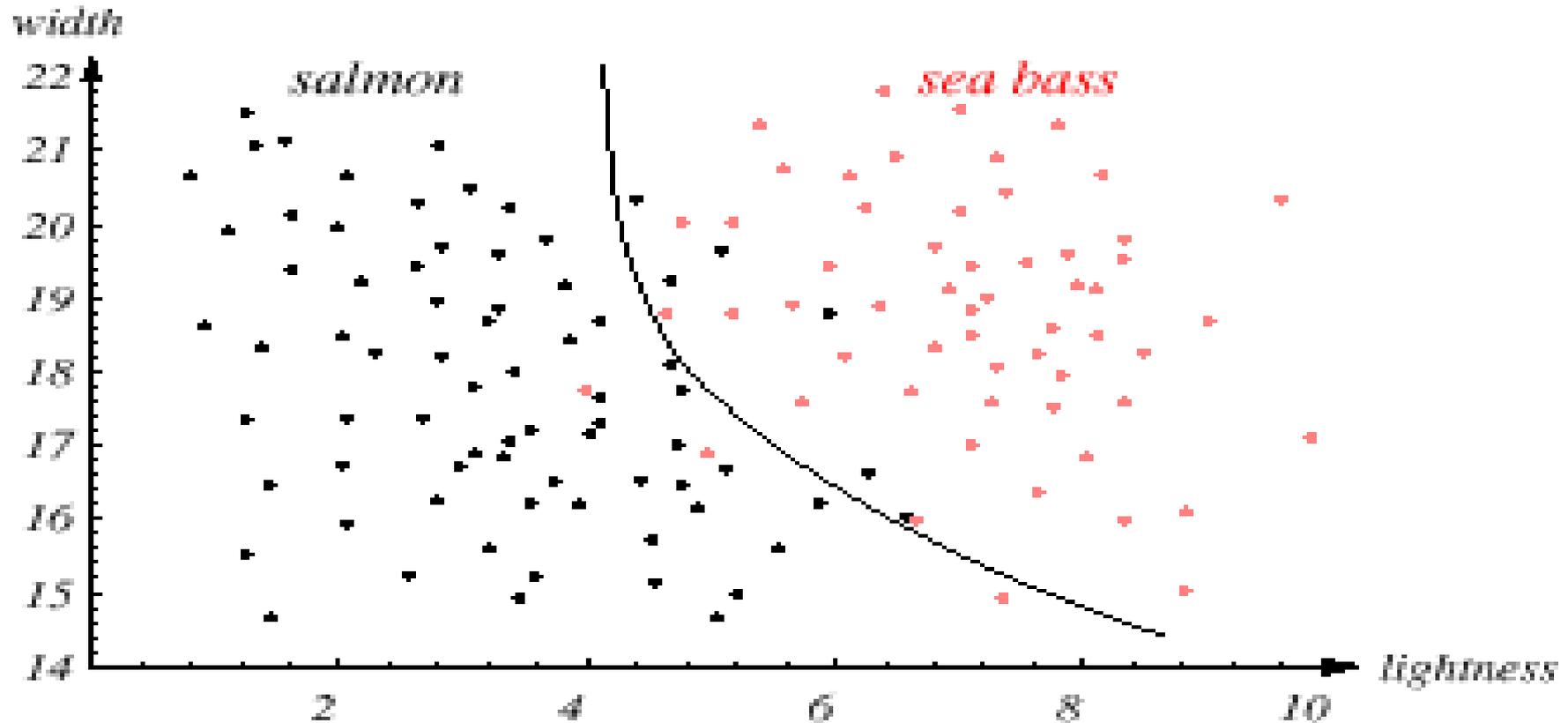
- Merkmale, die wenig bzw. unkorreliert mit bisherigen sind, hingegen ...
- Merkmale, die verrauscht sind, können die Performanz reduzieren

Optimale Entscheidungsgrenze?



Optimalität auf Trainingsdaten!
Generalisierungsfähig?

Generalisierungsfähigkeit



Wahl der Entscheidungsgrenze!
“Optimalität” auf “echten” (Test-)Daten!

Aufgabenstellungen im Mustererkennungsprozess

Klassifikation

Feststellen der Klassenzugehörigkeit

- Zusammenfassen von Objekten in Untermengen, die gemeinsame Eigenschaften aufweisen (Ähnlichkeit).
- Vernachlässigung von „feinen“ Unterschieden bei einzelnen Elementen

Zuordnung der Merkmale X zu einer Klasse Y .

Beispiele:

- Apfel oder Birne, Wolfsbarsch oder Lachs
- OCR (Optical Character Recognition): Ziffern 0..9
- Ein Patient hat das Symptommuster x . Welcher Diagnose lässt sich dieses am besten zuordnen?

Regression

Einfluss der Merkmale X auf eine Zielgröße Y .

- Wie ist die Dosierung eines Medikaments zu wählen, um eine Lungenentzündung möglichst gut zu behandeln?
- Vorhersage der Lebenserwartung anhand von diagnostischen Merkmalen

Clustering

Annahme: beobachteten Objekte stammen aus K verschiedenen Klassen.
Zerlegung der Menge der Objekte derart, dass

- Objekte einer bestimmten Untermenge möglichst ähnlich und
- Objekte aus verschiedenen Untermengen möglichst „unähnlich“ sind.

Beispiele:

- Patient weist das Symptommuster X auf. Enthalten die Daten einen Fehler?
- Lässt sich eine bestimmte Krankheit in Unterklassen einteilen?

Musteranalyse

Strukturelle Beschreibung gesucht

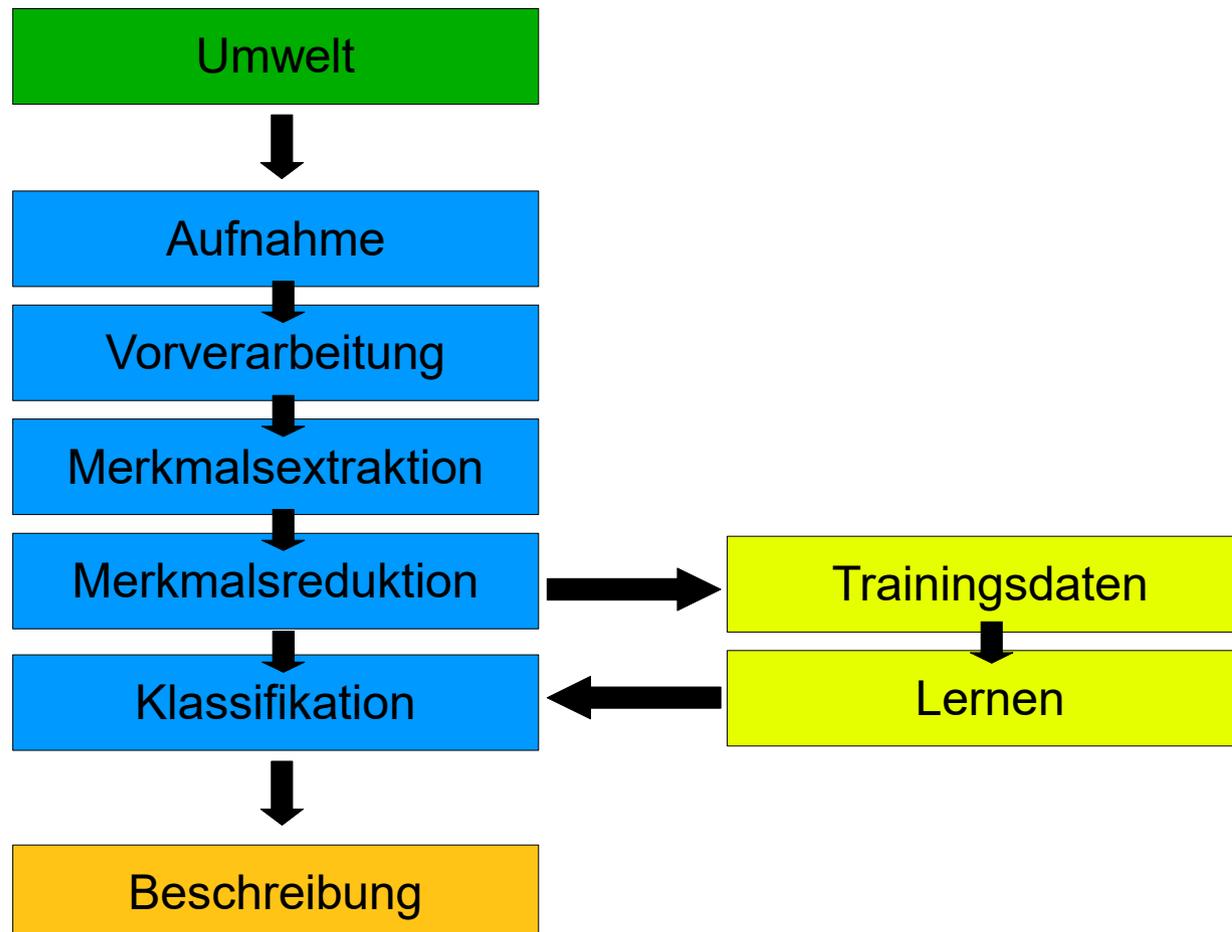
- Zerlegung in einfachere Teilmuster
- Analyse der Beziehungen zwischen Teilmustern

Beispiele:

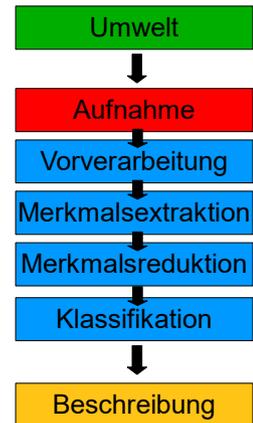
- Buchstabe: Strichanalyse, Schablonenmatching
- Gesichter: Mund, Augen, Nase, Ohren
- Menschliche Bewegung: Rumpf, Arme, Beine, ...
- Sprache: Grammatikalische Analyse, Satzstruktur (Parsing)

Schritte des Mustererkennungsprozesses

Schritte der Mustererkennung



Aufnahme (Datenerfassung)



Aufnahmegerät (Sensor)

- Beschreibung der aufgenommenen Information durch multivariate Funktionen:

$$\vec{f}(\vec{x}) = \begin{bmatrix} f_1(\chi_1, \chi_2, \dots, \chi_n) \\ f_2(\chi_1, \chi_2, \dots, \chi_n) \\ \cdot \\ \cdot \\ \cdot \\ f_m(\chi_1, \chi_2, \dots, \chi_n) \end{bmatrix}$$

- Implizite Vorverarbeitung
- sensorspezifische Signalveränderung
- Informationsreduktion
- Aufgenommenes Muster (Ausschnitt der Welt)
- Diskretisierung

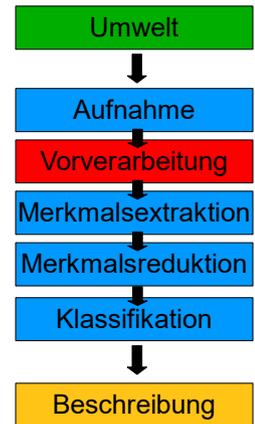
Vorverarbeitung

Zur Datenreduktion und
Qualitätsverbesserung

Methoden:

- Restauration
- Entstörung, Entzerrung
- Vignettierung
- Rauschen
- Ausreißer

- Normierung
- Größe
- Skalenstandardisierung
- Segmentierung
- Skelette
- Konturen
- Codierung
- Thresholding
- Komprimierung



Merkmalsextraktion

Transformation der Daten in den Merkmalsraum
(=problemspezifische Repräsentationsform)

- Heuristische Extraktion
- Form, Farbe, Größe
- Kanten, Ecken, Krümmung (curvature)
- SIFT (scale-invariant feature transform)
- Hough Transformation

- Konturen (snakes)
- Bewegung, Optical Flow
- Spektren: Diskrete Fourier-Transformation (DFT), Wavelets
- Momente
- Schwerpunkt, Trägheitsmoment, etc.



Merkmalsreduktion



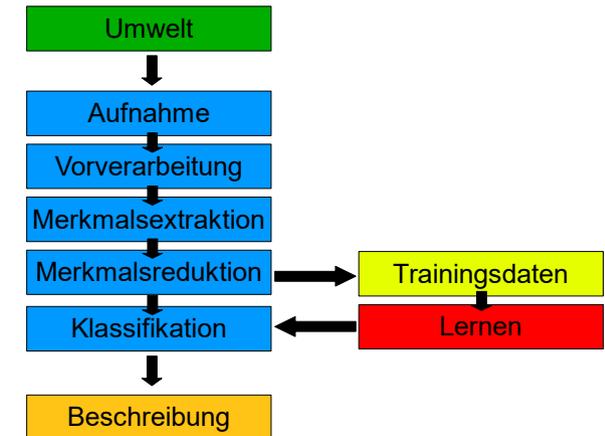
Repräsentation

- Vektor
- Symbolkette (String)
- Baum
- Graph
- ...

Reduktion

- Filtern
- Principal/Independent components analysis
- Isomaps
- Latent semantic analysis
- Kernel-trick
- ...

Lernen



Auswahl von
Klassifikationsverfahren und
Methodik zur Modellauswahl

- Festlegen von Parametern der Klassifikatoren
- Evaluierung der Modelle

Arten von Lernen

- Überwachtes
- Unüberwachtes
- Re-inforcement Learning

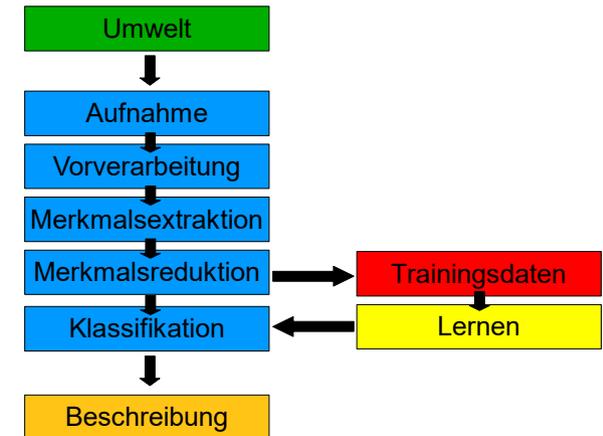
Trainingsdaten

Cross-validation: zufällige Aufteilung der Lerndaten in

- Training set
- Validation set
- Test set

Arten der Kreuzvalidierung

- Holdout validation
- K-fold cross validation
- Leave-one-out cross-validation
- Wozu? - Modellselektion und classification error estimation



Klassifikation

Disjunkte Musterklassen

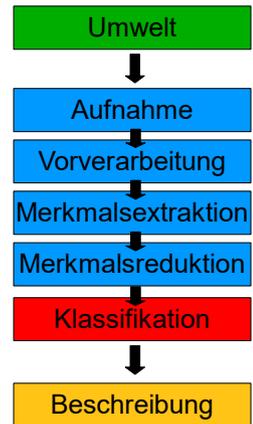
Zurückweisung: keine Klasse!

oder auch Regression ...

Benötigt Definition der
Ähnlichkeit

Arten der Klassifikation:

- Linear vs. nicht-linear
- Numerisch vs. nicht-numerisch
- Fest dimensioniert vs. variabel
- Parametrisch vs. nicht parametrisch



Beschreibung

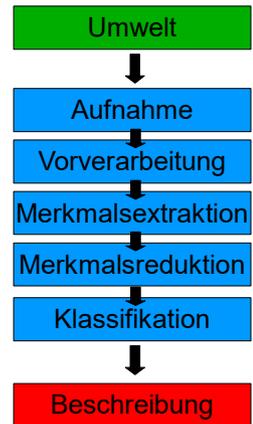
Zerlegung eines komplexen Musters in Bestandteile

Analyse der Beziehungen

Beschreibung

- Formal
- Natürlichsprachig

Bedarf eines problemspezifischen Wissensrepräsentation



Ausblick

Nächste Vorlesung:

Donnerstag, 20.10.2016 13-15 (c.t.)

Abstandsmaße; Entscheidungstabellen/-bäume