

# Themen “Mathematische Methoden der Mustererkennung”

Hochschule Niederrhein, Prof. Dr. Dalitz  
Studiengang Master of Science Informatik, Wintersemester 2011/12  
(Stand: 31.01.2012)

## 1 Wiederholung Wahrscheinlichkeitsrechnung

- Ereignisse und Ereignisraum, Wahrscheinlichkeitsaxiome, Unabhängigkeit von Ereignissen
- Zufallsvariablen, Verteilungsfunktion (Cumulative Distribution Function, CDF), Wahrscheinlichkeitsdichte (Probability Density Function),
- spezielle Verteilungen: Gleichverteilung, Normalverteilung, Binomialverteilung
- Kenngrößen von Zufallsvariablen: Erwartungswert, Varianz, Standardabweichung, Median

## 2 Grundbegriffe der Mustererkennung

An einem konkreten Beispiel werden die folgenden Begriffe erläutert:

- Feature, Klasse und Classifier
- diskriminatorischer und probabalistischer Ansatz
- Trainieren und Fehlerrate

[Duda] Kap. 1

## 3 Bayessche Entscheidungstheorie

Bedingte Wahrscheinlichkeit und der Satz von Bayes, die Bayessche Entscheidungsregel, Verallgemeinerung auf stetige Featurewerte und auf mehrere Features (= mehrdimensionaler Featurevektor), die Bayessche Fehlerrate als minimal mögliche Fehlerrate, Verallgemeinerungen der Bayesschen Entscheidungsregel auf die Risikominimierung

Umformulierung der Bayesschen Entscheidungsregel über “Diskriminantenfunktionen”, Anwendung der Entscheidungsregel auf konkrete Beispiele für angenommene Wahrscheinlichkeitsverteilungen, Template Matching als Spezialfall der Bayesschen Entscheidungsregel unter stark vereinfachenden Annahmen

Bayessche Entscheidungstheorie: [Webb] Kap. 1.5, Mehrdimensionale Normalverteilung: [Duda] Kap 2.6

## 4 Dichteschätzung (“Training”)

Parametrische Verfahren (machen Annahme über Form der Verteilung):

- Likelihood-Funktion und Maximum-Likelihood Schätzer
- Erwartungstreue von Schätzern
- Schätzer für Erwartungswert und Kovarianzmatrix

Nichtparametrische Verfahren:

- Histogramm-Methoden
- Kernel-Methoden (Parzen Windows)
- kNN Methode ( $k$  Nearest Neighbor)

Die Verfahren werden an Beispielen am Computer demonstriert und ausprobiert.

Maximum-Likelihood: [Duda] Kap. 3.2, Nichtparametrische Verfahren: [Webb] Kap. 3

## 5 Metrische Räume

Die Begriffe Metrik, Norm, Skalarprodukt und deren Zusammenhang. Minkowski ( $L_p$ ) Normen (Euklidisch, Manhattan, Maximum usw.), Mahalanobis Distance, Feature Weighing als verallgemeinerte  $L_p$ -Norm, Feature-Normalisierung

## 6 Hidden-Markov Modelle

Mathematische Beschreibung von Markov-Ketten und grundlegende Eigenschaften, Hidden-Markov Modelle. Lösungsalgorithmen für die drei grundlegenden Probleme bei Hidden-Markov Modellen:

- das Evaluationsproblem (Forward Algorithmus)
- das Dekodierungsproblem (Viterbi Algorithmus)
- das Trainingsproblem (Grundidee des iterativen Schätzens)

[Rabiner] Abschnitte 2-3, Achtung: der in [Duda] Kap. 3.10 angegebene Algorithmus zur Dekodierung ist nicht der Viterbi-Algorithmus, sondern ein Algorithmus, der das Problem gar nicht löst!

## 7 Classifier Performance Estimation

Die verschiedenen Fehlerraten: Bayessche Fehlerrate  $e_B$ , tatsächliche (oder “conditional”) Fehlerrate  $e_T$  (für einen konkret verwendeten Trainingsatz), expected (oder “unconditional”) Error Rate  $e_E$  (über alle Trainingsätze fester Größe gemittelt).

Schätzer für die tatsächliche Fehlerrate  $e_T$ : das *Holdout Estimate* und seine Konfidenzintervalle (klassisch, Agresti-Coull, [HPD Intervall]). Schätzer für die expected Error Rate  $e_E$ : Resubstitution, Cross-Validation und Resampling (Bootstrap)

Vergleich der Performance von Klassifiern mittels McNemar’s Test und der Receiver Operating Characteristic (ROC) Kurve. Vergleich von Algorithmen mit dem “paired  $t$ -Test”.

[Webb] Kap 8.1-8.3, [Brown et al.], [Ditterich]

## 8 Classifier Confidence & Rejection

*Ambiguity* versus *Distance* Rejection, Reject-Option und der Satz von Chow, Konfidenzmaße basierend auf Schätzern von  $P(\omega_i|x)$  und auf Abständen im Feature-Space, Distance-Rejection basierend auf der empirischen Verteilungsfunktion der Abstände im Feature-Space

[Dalitz]

## 9 Feature Selection & Extraction

Zusammenhang zwischen Feature-*Selection*, -*Extraction* und -*Weighting*, Unterscheidung *Filter*- und *Wrapper*-Methoden, Maße für die Klassenseparation im Featurespace, Strummatrizen, Principal Component Analyse (PCA, auch “Karhunen-Loève Transformation”) und Linere Diskriminanten Analyse (LDA)

[Duda] Kap 3.8, [Webb] Kap 9.3.1 und 4.2.3, [Theodoridis] Kap 6.3 und 5.8

## 10 Clustering

Clustering als “unsupervised learning”, flaches versus hierarchisches Clustering, *k-means* Clustering: Algorithmus und Eigenschaften, Hierarchisches Clustering: Single-Link und Complete-Link Algorithmus, Graph-basiertes Clustering über den Minimum Spanning Tree

[Theodoridis] Kap 14.5.1, 15.2.1, 13 (Einleitung), [Manning et al.] Kap. 16

## Literatur

- [Theodoridis] S. Theodoridis, K. Koutroumbas: *Pattern Recognition*. Academic Press Elsevier, 4. Auflage (2009)
- [Duda] R.O. Duda, P.E. Hart, D.G. Stork: *Pattern Classification*. John Wiley and Sons (2001)
- [Webb] A. Webb: *Statistical Pattern Recognition*. Wiley-Interscience (2. Auflage, 2002)
- [Rabiner] L.R. Rabiner: *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. Proceedings of the IEEE, Vol. 77, No. 2, pp. 257-286, February 1989
- [Brown et al.] L.D. Brown, T.T. Cai, A. DasGupta: *Interval Estimation for a Binomial Proportion*. Statistical Science, Vol. 16, pp. 101-133 (2001)
- [Ditterich] T.G. Ditterich: *Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms*. Neural Computation, Vol. 10, pp. 1895-1923 (1998)
- [Dalitz] C. Dalitz: *Reject Options and Confidence Measures for kNN Classifiers*. In C. Dalitz (Ed.): “Document Image Analysis with the Gamera Framework.” Schriftenreihe des Fachbereichs Elektrotechnik und Informatik, Hochschule Niederrhein, vol. 8, pp. 16-38, Shaker Verlag (2009)
- [Manning et al.] C.D. Manning, P. Raghavan, H. Schütze: *Introduction to Information Retrieval*. Cambridge University Press (2008)