

Merkmalsequenzen in der Mustererkennung


**Dynamic Time Warping-Techniken
am Beispiel der On-line Handschrifterkennung**

Claus Bahlmann

Institut für Informatik
Albert-Ludwigs-Universität Freiburg

23. Januar 2003

Überblick


- Einführung Handschrifterkennung
- Daten- / Merkmalgewinnung in der on-line Handschrifterkennung
- Dynamic time warping (DTW) Distanz für Merkmalsequenzen
- Klassifikationstechniken für Merkmalsequenzen
 - Generative Klassifikation
 - * CSDTW — Cluster generative SDTW
 - Diskriminative Klassifikation
 - * SVM-GDTW — ein Gauss-DTW-Kern für Support-Vektor-Maschinen (SVM)
- Anwendung  *frog on hand | CSDTW*

Handschrifterkennung —Definition

Transformation einer Sprache, welche in räumlicher Form von grafischen Markierungen vorliegt, in eine symbolische Repräsentation

(Plamondon und Srihari, 2000) [8]

Handschrifterkennung — Disziplinen

	Off-line	On-line
		
<i>Erkennungszeitpunkt:</i>	nach dem Schreiben	während des Schreibens
<i>Digitalisierung:</i>	Scannen des Bildes	Abtasten der Stiftbewegung
<i>Datenrepräsentation:</i>	2D-Bild	1D-Zeitsequenz von 2D-Ortskoordinaten
<i>Beispielanwendung:</i>	Prozeß-Automatisierung bei Post, Bank, ...	„Pen-Computing“ (PDA, Tablet PC, ...)

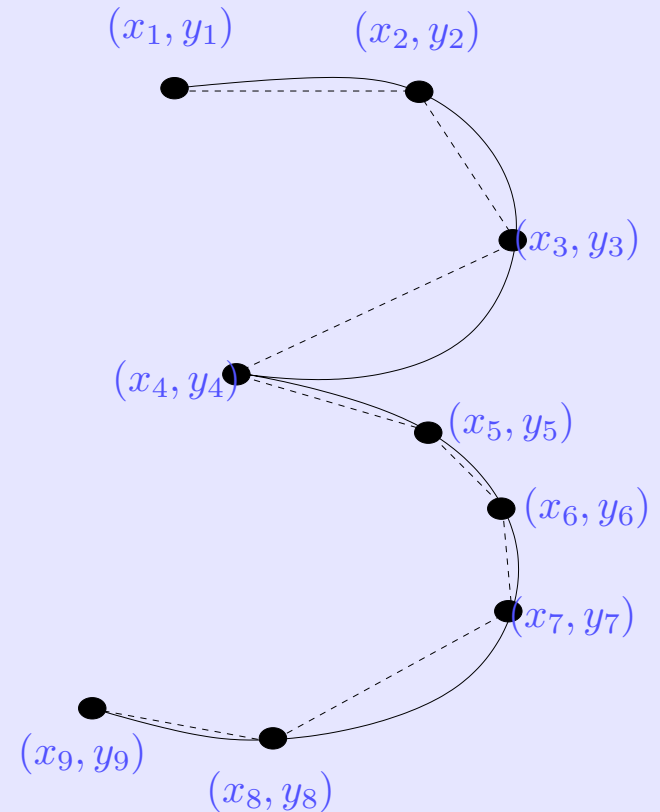
On-line Handschrifterkennung — Datengewinnung

Umgebung: Digitalisiertablett und (spezieller) Stift

Vorgehensweise: Abtastung der Ortskoordinaten eines Schriftzugs in zeitlich konstantem Intervall

—→ *Sequenz* von Koordinaten

$$\begin{aligned}\mathcal{P} &= [\mathbf{p}_1, \dots, \mathbf{p}_N] \\ &= [(x_1, y_1)^T, \dots, (x_N, y_N)^T]\end{aligned}$$



On-line Handschrifterkennung — Merkmalgewinnung

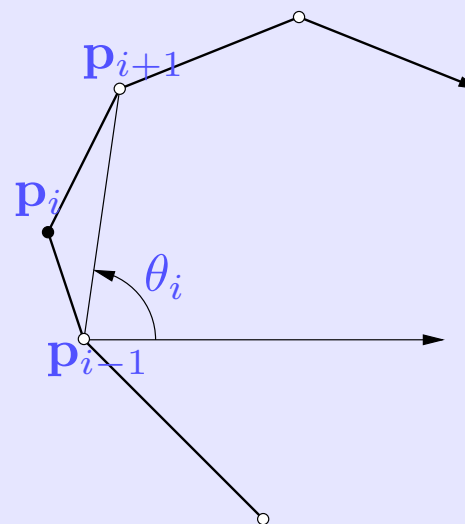
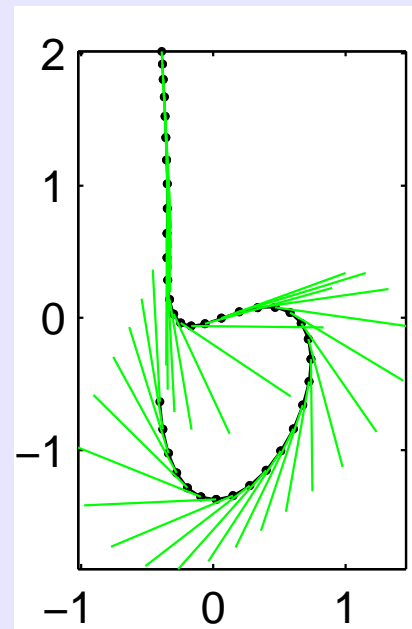
Gewinnung von lage- / skalierungsinvarianten Merkmalen der Koordinatensequenz $\mathcal{P} = [\mathbf{p}_1, \dots, \mathbf{p}_N]$:

- normierte x -Koordinate $\tilde{x}_i = \frac{x_i - \mu_x}{\sigma_x}$
- normierte y -Koordinate $\tilde{y}_i = \frac{y_i - \mu_y}{\sigma_y}$
- Tangentensteigungswinkel

$$\theta_i = \text{ang} (\Delta_i x + \sqrt{-1} \cdot \Delta_i y)$$

$$\text{Merkmalvektor } \mathbf{t}_i = (\tilde{x}_i, \tilde{y}_i, \theta_i)^\tau$$

$$\text{Merkmalvektorsequenz } \mathcal{T} = [\mathbf{t}_1, \dots, \mathbf{t}_{N_T}]$$



Dynamic Time Warping

ein Abstandsmass für Merkmalsequenzen

Probleme:

- unterschiedlich lange Sequenzen
- zeitliche Variationen in den Sequenzen

Lösungsansatz:

Einführung von

- neuer Zeitachse $\{1, \dots, N\}$
- Zuordnungsfunktion

$$\Phi : \{1, \dots, N\} \rightarrow (\{1, \dots, N_{\mathcal{T}}\} \times \{1, \dots, N_{\mathcal{R}}\})$$

definiert durch *lokale Transitionen*

$$\Phi = [\Delta\phi_1, \dots, \Delta\phi_N]$$

Dynamic Time Warping — Problemformulierung

Warping Distanz:

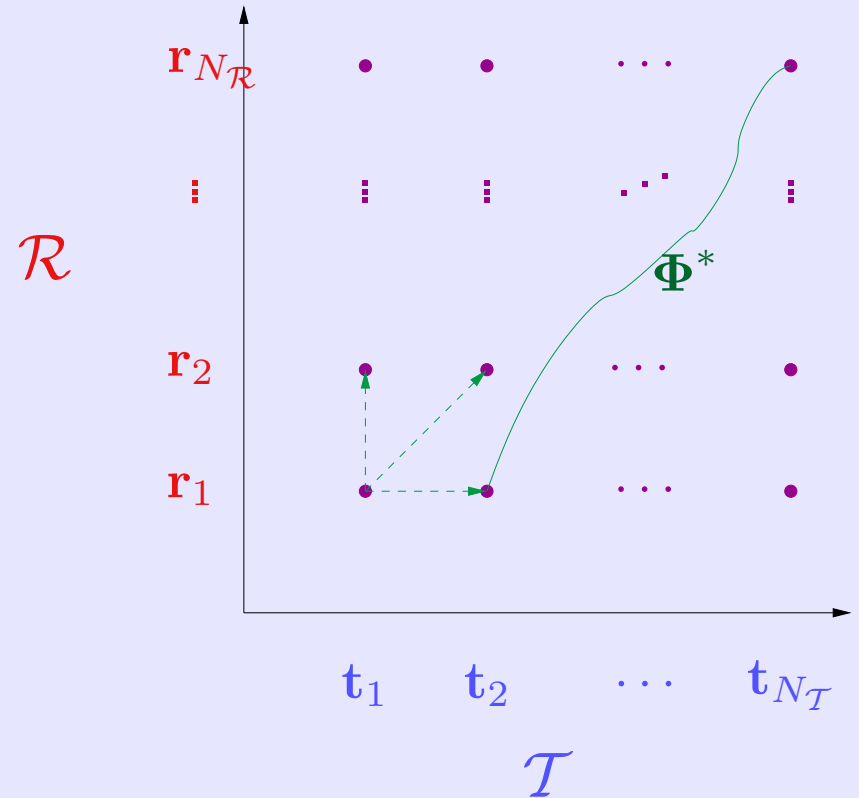
$$D_{\Phi}(\mathcal{T}, \mathcal{R}) = \sum_{n=1}^N d(\mathbf{t}_{\phi_{\mathcal{T}}(n)}, \mathbf{r}_{\phi_{\mathcal{R}}(n)})$$

DTW-Distanz := Viterbi-Distanz:

$$D^*(\mathcal{T}, \mathcal{R}) = D_{\Phi^*}(\mathcal{T}, \mathcal{R}) = \min_{\Phi} \{D_{\Phi}(\mathcal{T}, \mathcal{R})\}$$

Lokale Distanz: Quadrierte Euklidische Distanz

$$d(\mathbf{t}_i, \mathbf{r}_j) = \|\mathbf{t}_i - \mathbf{r}_j\|^2$$



Dynamic Time Warping — Algorithmen zur effizienten Berechnung

Naive Suche:

exponentiell viele Pfade !!! → scheidet daher aus

Dynamische Programmierung:

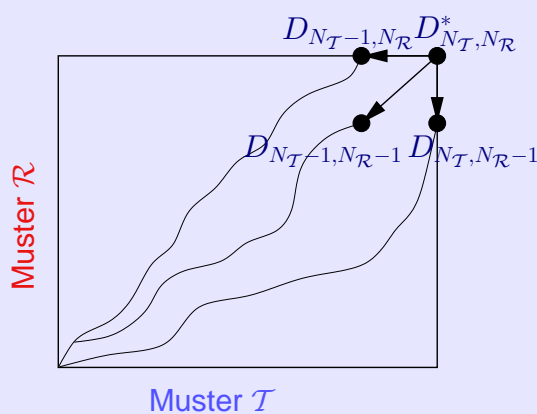
$\mathcal{O}(c \cdot N_T \cdot N_R)$ → siehe nächste Folie(n)
(c = Anzahl lokaler Transitionen)

Strahlsuche (beam search):

$\mathcal{O}(c \cdot \max\{N_T, N_R\})$ → siehe übernächste Folie

Dynamic Time Warping —Dynamische Programmierung (1)

- $D_{i,j}^*$: Viterbi-Distanz für Vergleich der Präfixe $[\mathbf{t}_1, \dots, \mathbf{t}_i]$ und $[\mathbf{r}_1, \dots, \mathbf{r}_j]$.



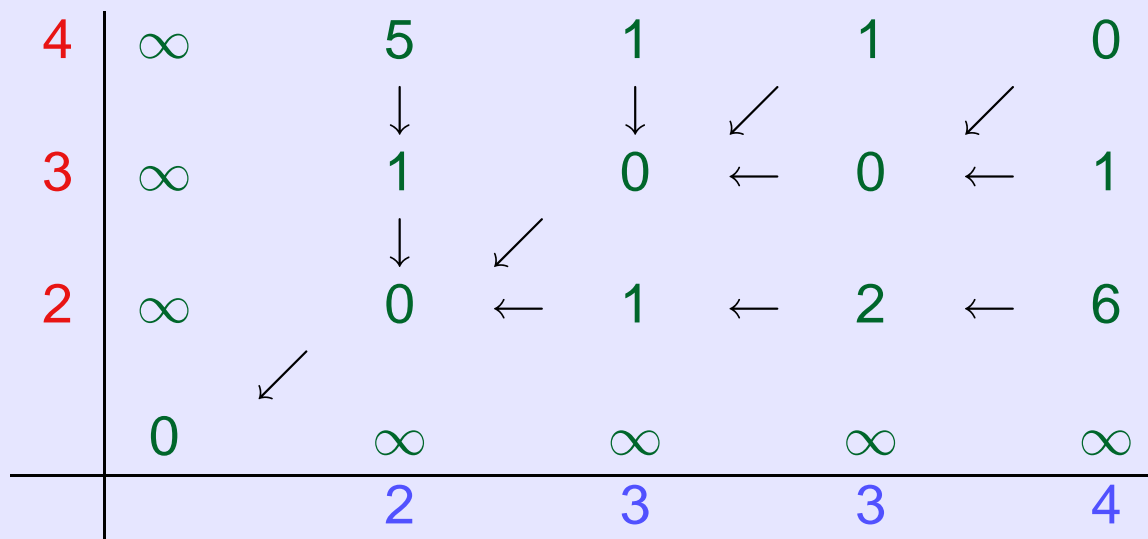
$$\begin{aligned}
 D^*(\mathcal{T}, \mathcal{R}) &= D_{N_{\mathcal{T}}, N_{\mathcal{R}}}^* \\
 &= d(\mathbf{t}_{N_{\mathcal{T}}}, \mathbf{r}_{N_{\mathcal{R}}}) + \min \left\{ \begin{array}{l} D_{N_{\mathcal{T}}-1, N_{\mathcal{R}}} \\ D_{N_{\mathcal{T}}-1, N_{\mathcal{R}}-1} \\ D_{N_{\mathcal{T}}, N_{\mathcal{R}}-1} \end{array} \right\}
 \end{aligned}$$

- Rekursiver, rückwärtsschauender Aufbau der *Viterbi-Matrix* \mathbf{D}^*
- Komplexität $\mathcal{O}(c \cdot N_{\mathcal{T}} \cdot N_{\mathcal{R}})$

Dynamic Time Warping —Dynamische Programmierung (2)

Lokale Distanzmatrix

Viterbi-Matrix

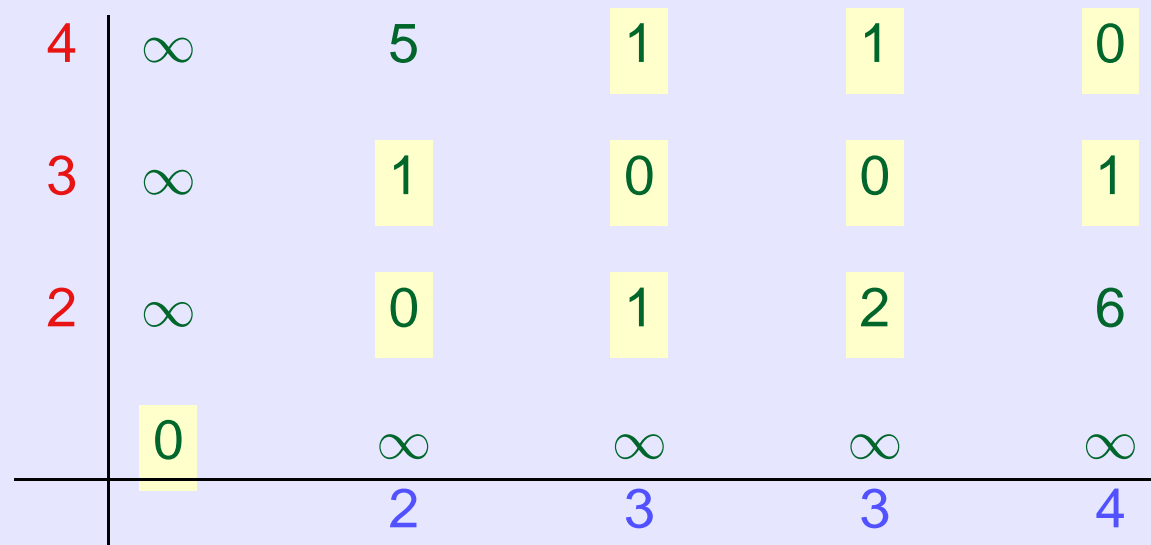


DTW in Aktion

Matlab-Demo

Dynamic Time Warping — Strahlsuche

- Iteratives Verwalten *aktiver* Pfadhypothesen
- Abbruch der Verfolgung von nichtwahrscheinlichen Pfadhypothesen
- Suboptimales Verfahren



- Komplexität $\mathcal{O}(c \cdot (N_T + N_R))$

Strahlsuche in Aktion

Matlab-Demo

Statistisches DTW — Idee

Bei bekannten Wahrscheinlichkeitsvert., z.B.

$$\mathbf{R}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\tau (\boldsymbol{\Sigma})^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Euklid. Distanz $\longrightarrow -\log(\text{klassenbedingte Wkt.})$

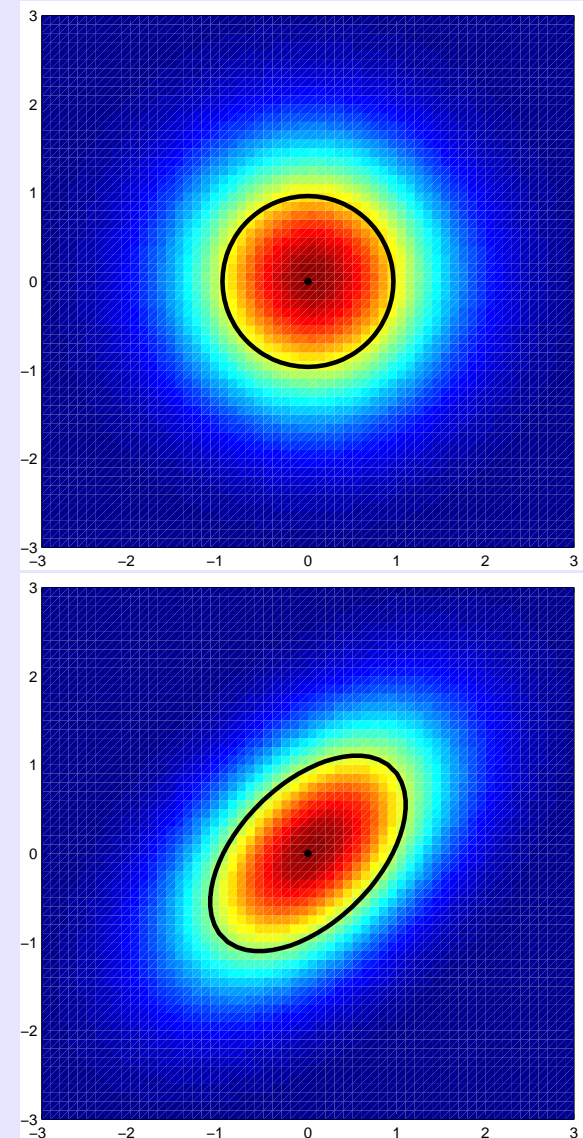
$$d(\mathbf{t}, \mathbf{r}) = \|\mathbf{t} - \mathbf{r}\|^2$$

\Downarrow

$$\begin{aligned} d(\mathbf{t}, \mathbf{R}) &= -\log P(\mathbf{t}|\mathbf{R}) \\ &= \frac{F}{2} \log(2\pi) + \frac{1}{2} \log(|\boldsymbol{\Sigma}|) \\ &\quad + \frac{1}{2} (\mathbf{t} - \boldsymbol{\mu})^\tau (\boldsymbol{\Sigma})^{-1} (\mathbf{t} - \boldsymbol{\mu}) \end{aligned}$$

Statistisches Modell

$$\mathbf{R} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$$



Statistisches DTW — Idee (2)

DTW-Distanz $\longrightarrow -\log(\text{klassenbedingte Wkt.})$

$$\begin{aligned} D_{\Phi}(\mathcal{T}, \mathcal{R}) &= \sum_{n=1}^N \left\| \mathbf{t}_{\phi_{\mathcal{T}}(n)} - \mathbf{r}_{\phi_{\mathcal{R}}(n)} \right\|^2 \\ &\Downarrow \\ D_{\Phi}(\mathcal{T}, \mathcal{R}) &= -\log P(\mathcal{T} | \mathcal{R}, \Phi) \\ &= \sum_{n=1}^N d(\mathbf{t}_{\phi_{\mathcal{T}}(n)}, \mathbf{R}_{\phi_{\mathcal{R}}(n)}) \\ &= \sum_{n=1}^N \frac{F}{2} \log(2\pi) + \frac{1}{2} \log(|\Sigma_{\phi_{\mathcal{R}}(n)}|) \\ &\quad + \frac{1}{2} \left(\mathbf{t}_{\phi_{\mathcal{T}}(n)} - \boldsymbol{\mu}_{\phi_{\mathcal{R}}(n)} \right)^{\top} \left(\Sigma_{\phi_{\mathcal{R}}(n)} \right)^{-1} \left(\mathbf{t}_{\phi_{\mathcal{T}}(n)} - \boldsymbol{\mu}_{\phi_{\mathcal{R}}(n)} \right) \\ &\quad - \log(P_{\phi_{\mathcal{R}}(n)}(\Delta\phi)) \end{aligned}$$

Statistisches Modell

$$\mathbf{R}_j = (\boldsymbol{\mu}_j, \Sigma_j, P_j(\Delta\phi))$$

Statistisches DTW — Problemformulierung

Warping Distanz:

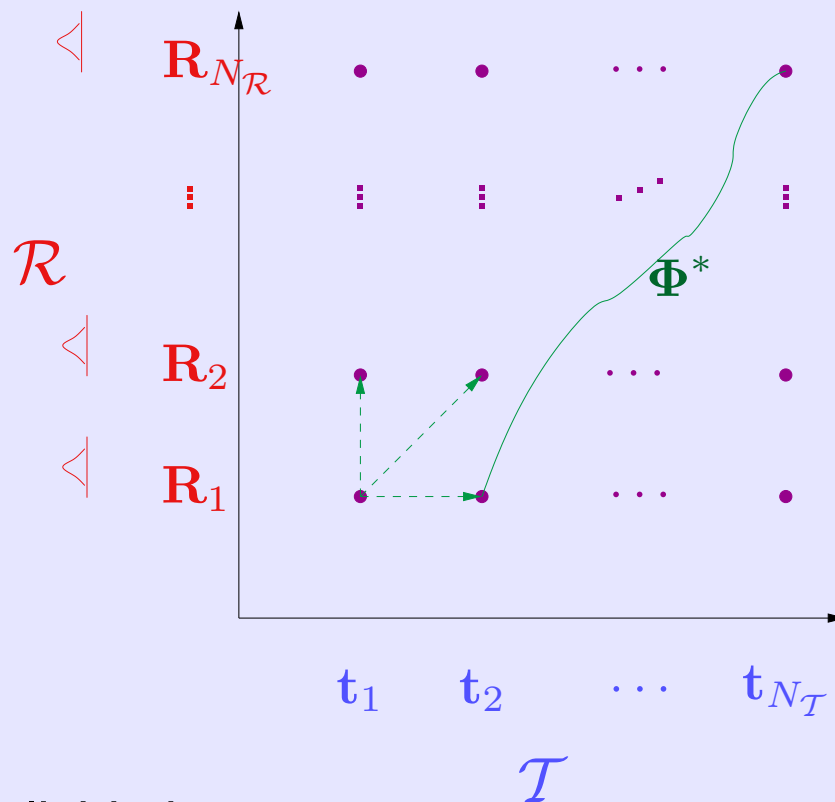
$$D_{\Phi}(\mathcal{T}, \mathcal{R}) = \sum_{n=1}^N d(\mathbf{t}_{\phi_{\mathcal{T}}(n)}, \mathbf{R}_{\phi_{\mathcal{R}}(n)})$$

SDTW-Distanz := Viterbi-Distanz:

$$D^*(\mathcal{T}, \mathcal{R}) = D_{\Phi^*}(\mathcal{T}, \mathcal{R}) = \min_{\Phi} \{D_{\Phi}(\mathcal{T}, \mathcal{R})\}$$

Lokale Distanz: $-\log$ a-posteriori Wahrscheinlichkeit

$$d(\mathbf{t}_i, \mathbf{R}_j) = -\log P(\mathbf{t}_i | \mathbf{R}_j)$$

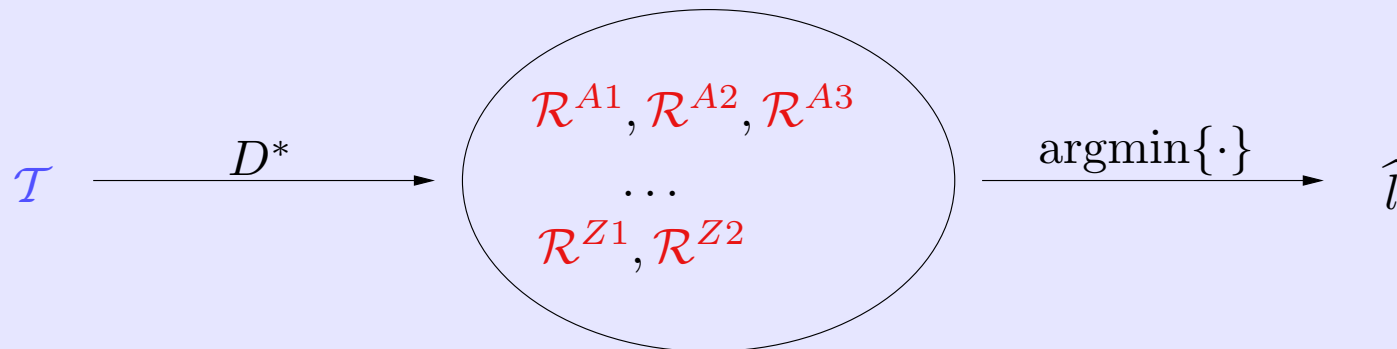
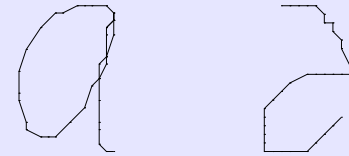


Generative Klassifikation

Cluster generative SDTW (CSDTW)

- Testmuster \mathcal{T}

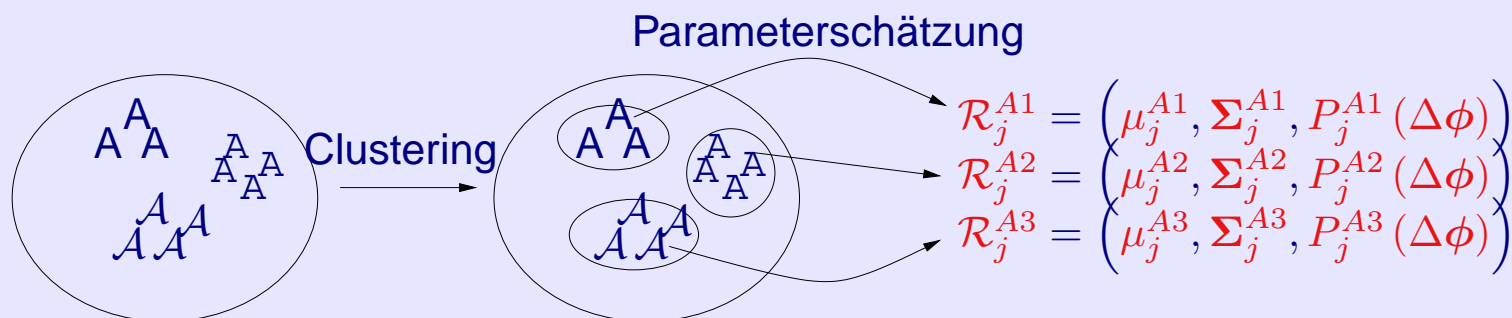
- Allographenmodelle $\mathcal{R}^{lk} = [\mathbf{R}_1^{lk}, \dots, \mathbf{R}_{N_{lk}}^{lk}]$



$$\hat{l} = \underset{l \in \{1, \dots, L\}, k \in \{1, \dots, K_l\}}{\operatorname{argmin}} \{D^*(\mathcal{T}, \mathcal{R}^{lk})\}$$

Cluster generative SDTW (CSDTW) — Training

2 Schritte für jede einzelne Klasse “A” bis “Z”;



1. Häufungspunktanalyse (Clustering) durch hierarchisches, agglomeratives Verfahren und modifiziertem DTW-Abstand
2. Schätzung der statistischen Parameter $\mathbf{R}_j^{lk} = \left(\mu_j^{lk}, \Sigma_j^{lk}, P_j^{lk}(\Delta\phi) \right)$ mit iterativem Algorithmus (Viterbi-Training):
 - (a) berechne neue Zuordnungsfunktion aller Beispiele eines Clusters zu Referenzmodell $\mathbf{R}_j^{lk}(t)$
 - (b) schätze neues Referenzmodell $\mathbf{R}_j^{lk}(t+1)$ entlang der berechneten Korrespondenzen

CSDTW in Aktion

Matlab-Demo

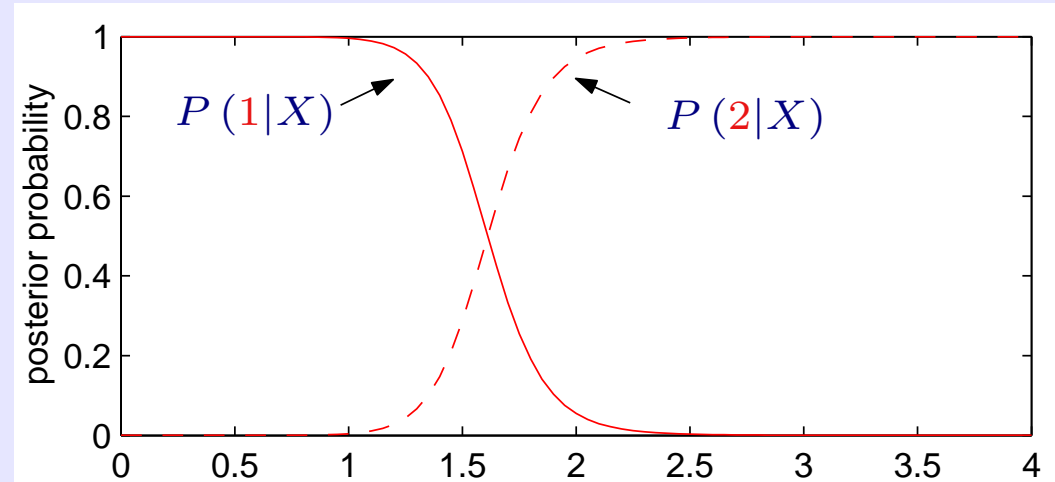
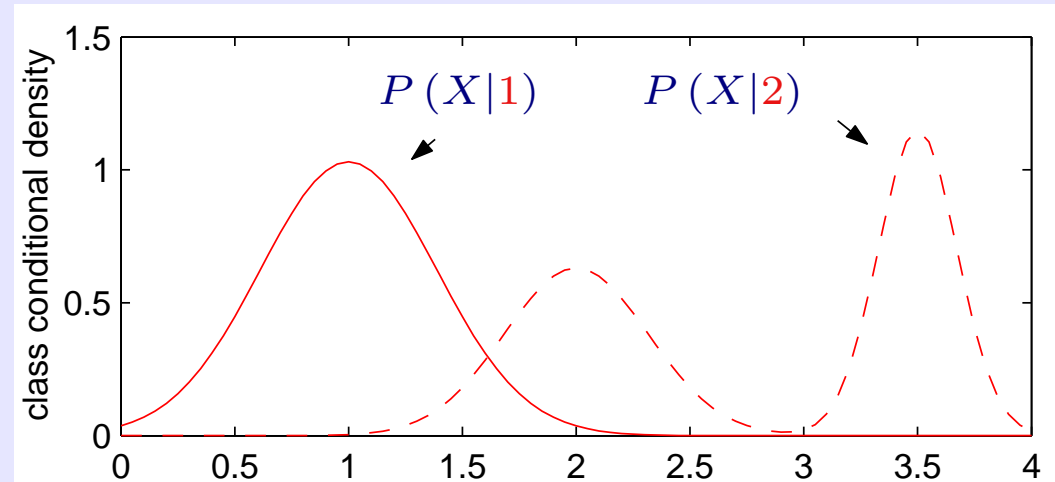
Zwischenstand

- Überblick Handschrifterkennung
- Daten- / Merkmalgewinnung
- Dynamic Time Warping (DTW)
- Statistische Erweiterung von DTW (SDTW)
- Generative Klassifikation mit *Cluster generative SDTW (CSDTW)*

Generative vs. diskriminative Klassifikation

1. Schätze klassenbedingte Wahrscheinlichkeitsdichten $P(X|l)$ für jede Klasse l
2. Wähle die Klasse mit der höchsten a-posteriori-Wahrscheinlichkeit mit Hilfe der Bayes-Regel

$$P(l|T) = \frac{P(T|l) P(l)}{P(T)}$$



Diskriminative Klassifikation

Support Vektor Maschine (SVM)

Kern:

$$K(T, P)$$

SVM Klassifikation:

$$\hat{S}(T) = \text{sign} \left(\sum_i \alpha_i S_i K(T, P_i) + b \right)$$

SVM Training: Bestimme α_i zur Maximierung der Zielfunktion

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j S_i S_j K(P_i, P_j)$$

unter den Nebenbedingungen

$$0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_i \alpha_i S_i = 0$$

Support Vektor Maschine (SVM)

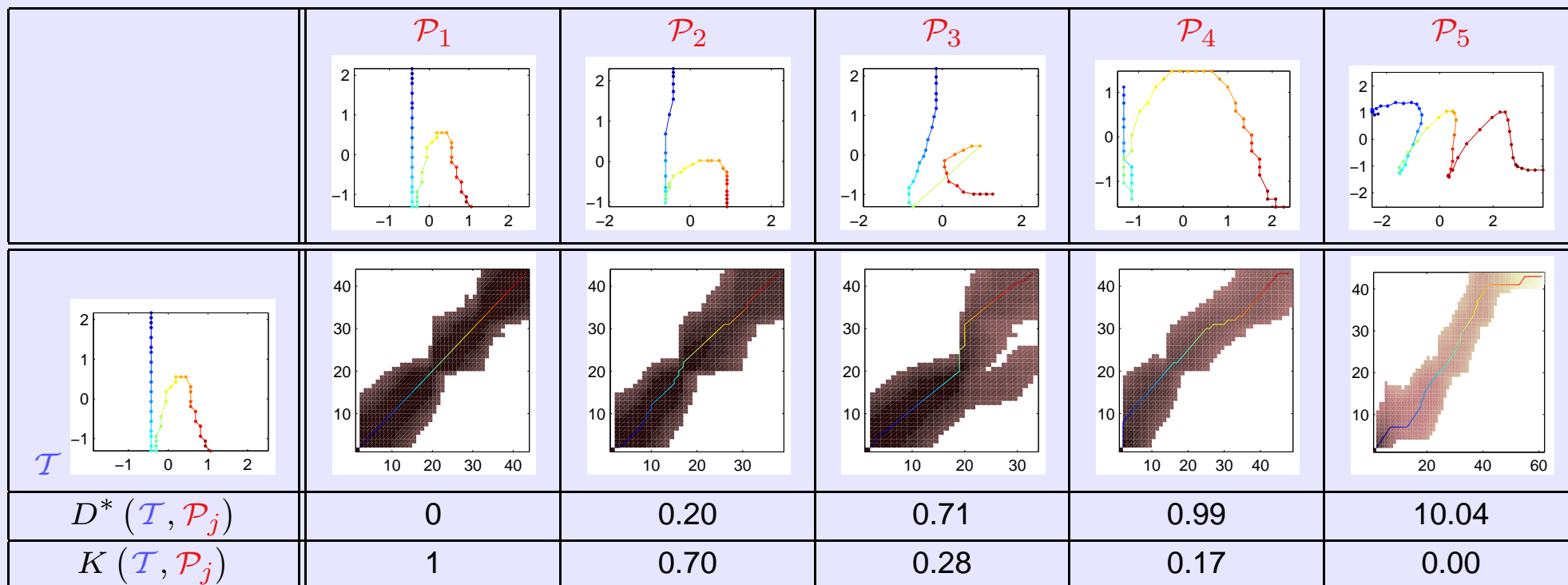
Kerne

	Vektorraumdaten	Sequenzen
Musterbeispiele	$T = (7, 5, 8)^T$ $P = (9, 3, 4)^T$	$T = [7, 5, 8]$ $P = [7, 5, 5, 8]$
Kernbeispiele $K(T, P)$	Gauss-Kern $K(T, P) = \exp(-\gamma \ T - P\ ^2)$	Gauss-DTW (GDTW)-Kern $K(T, P) = \exp(-\gamma D^*(T, P))$

SVM mit Gauss-DTW-Kern

Beispiele

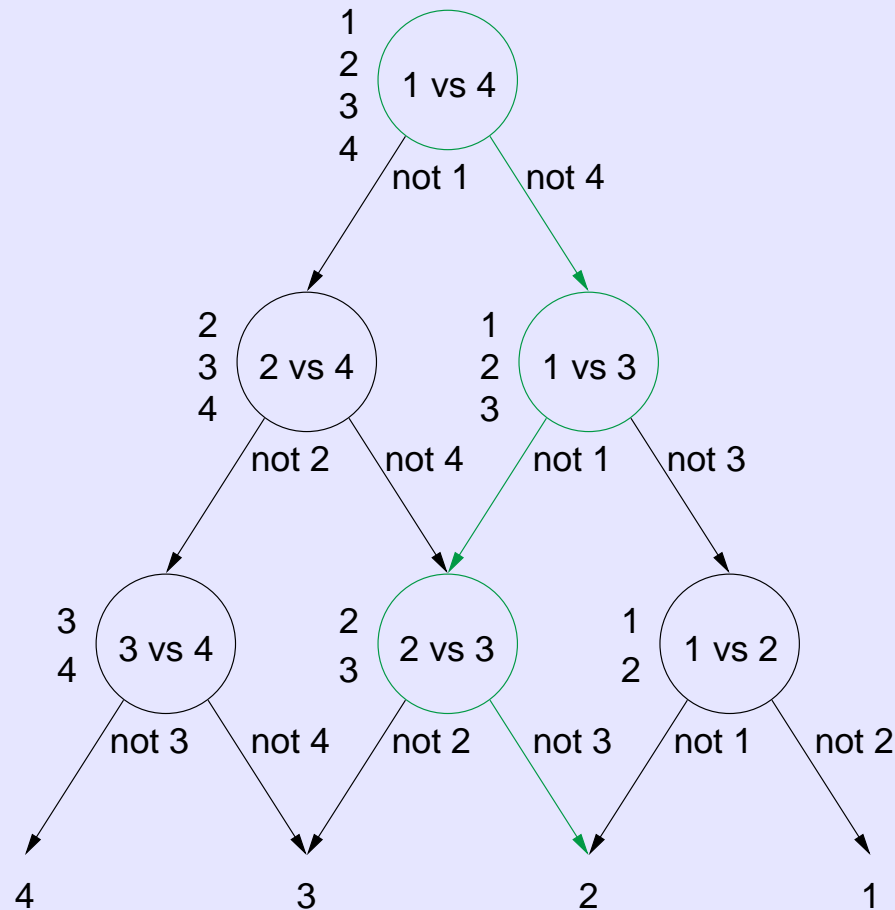
$$K(\mathcal{T}, \mathcal{P}_j) = \exp(-\gamma D^*(\mathcal{T}, \mathcal{P}_j))$$



Multi-Klassen SVM

DAG (directed acyclic graph)-SVM:

Kombiniere $K \cdot (K - 1) / 2$ Zwei-Klassen-SVMs zu **einer** K -Klassen-SVM



entnommen aus (Platt, 2000) [9]

SVM-GDTW in Aktion

Matlab-Demo

Fehlerraten

Kleinbuchstaben

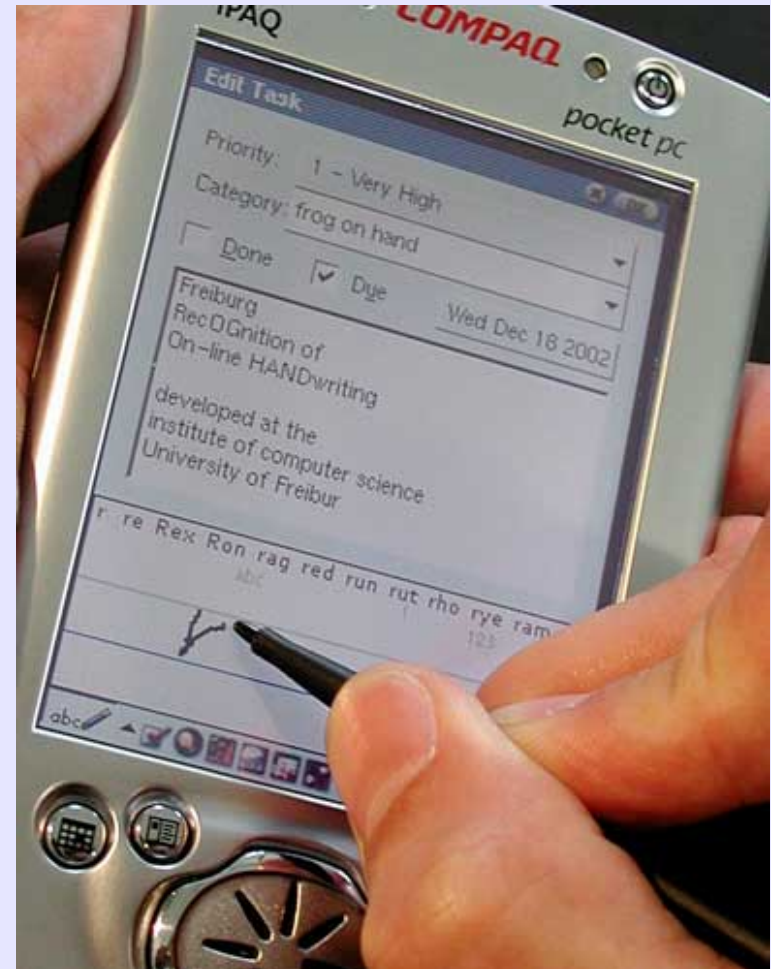
Verfahren	Fehlerrate E	UNIPEN Database Type
CSDTW [1]	9.3 %	Train-R01/V07 67 %/33 % Train/Test
DAG-SVM-GDTW [2]	12.0 %	Train-R01/V07 20 %/20 % Train/Test
MLP [7]	14.4 %	DevTest-R02/V02
HMM-NN hybrid [3]	13.2 %	Train-R01/V07
HMM [5]	14.1 %	Train-R01/V06 4 % "bad characters" removed

Anwendung



frog on hand | CSDTW

- Compaq iPAQ
 - 206 MHz Intel StrongARM Prozessor (ohne Fließkomma-Arithmetik)
 - 16 MByte Flash-Speicher ($\hat{=}$ Festplatte in PC)
 - 64 MByte RAM
- System: Embedded Linux Kernel 2.4.18 (*Familiar Distribution* [4])
- Oberfläche: *Open Palmtop Integrated Environment (OPIE)* [6]



frog on hand | CSDTW in Aktion

Demo

Zusammenfassung

- Überblick Handschrifterkennung
- Daten- / Merkmalgewinnung
- Dynamic Time Warping (DTW)
- Statistische Erweiterung von DTW (SDTW)
- Generative Klassifikation mit *Cluster generative SDTW (CSDTW)*
- Diskriminative Klassifikation mit Gauss-DTW-Kern und SVMs (*DAG-GDTW-SVM*)
- —→ DTW als vielseitig zu verwendendes Abstandsmass für Sequenzen
- *frog on hand* | *CSDTW* auf iPAQ
- Andere Anwendungen: Spracherkennung, Robotik, Gensequenzierung

Literatur

Literatur

- [1] Claus Bahlmann. A writer independent on-line handwriting recognition system with cluster-based statistical dynamic time warping. URL <mailto:bahlmann@informatik.uni-freiburg.de>. In Bearbeitung. [\(document\)](#)

- [2] Claus Bahlmann, Bernard Haasdonk, and Hans Burkhardt. On-line handwriting recognition with support vector machines—a kernel approach. In *Proc. 8th Int. Workshop Front. Handwriting Recognition (IWFHR)*, pages 49–54, Niagara-on-the-Lake, Canada, 2002. [\(document\)](#)

- [3] N. Gauthier, T. Artières, B. Dorizzi, and P. Gallinari. Strategies for combining on-line and off-line information in an on-line handwriting recognition system. In *Proc. 6th Int. Conf. Doc. Anal. Recognition (ICDAR)*, pages 412–416, Seattle, WA, 2001. [\(document\)](#)

- [4] Handhelds.org. URL <http://handhelds.org/>. Linux distribution for Compaq iPAQ. (document)
- [5] Jianying Hu, Sok Gek Lim, and Michael K. Brown. Writer independent on-line handwriting recognition using an HMM approach. *Pattern Recognition*, 33:133–147, January 2000. (document)
- [6] OPIE (Open Palmtop Integrated Environment). URL <http://www.opie.info/>. Graphical user interface for Linux Compaq iPAQ. (document)
- [7] Marc Parizeau, Alexandre Lemieux, and Christian Gagné. Character recognition experiments using UNIPEN data. In *Proc. 6th Int. Conf. Doc. Anal. Recognition (ICDAR)*, pages 481–485, Seattle, WA, 2001. (document)
- [8] Rejean Plamondon and Sargur N. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. Pattern Anal. and Mach. Intell.*, 22(1):63–84, January 2000. (document)
- [9] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin DAGS for multiclass classification. In *Advances in Neural Information Processing Systems 12*. MIT Press, 2000. (document)