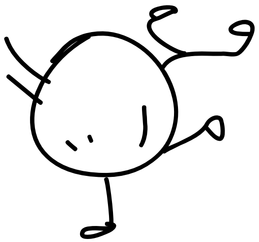


Markov Modelle

Vorlesung 186.844

07.01.2016

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

PRIP

Überblick

I. Was sind Markov Modelle?
II. Grundlagen: “Markov Models” (MM)
III. Grundlagen: “Hidden Markov Models” (HMM)
IV. Training von HMM
V. Klassifikation mit HMM
VI. Anwendungen

I. Was sind Markov Modelle?

Was sind Markov Modelle?

Markov Modelle (**Markov Models**) werden in der Muster-erkennung für die Klassifikation von Mustern eingesetzt, die sich aus einer Sequenz von „Sub-Mustern“ zusammensetzen. Sie wurden nach dem russischen Mathematiker Andrej Andrejewitsch Markov (1856-1922) benannt.

Beispiel: Angenommen das Wetter eines Tages kann als sonnig (S), bewölkt (W) oder verregnet (R) beschrieben werden.

- typische Sommerwoche: S-S-S-S-S-S-S
- typische Woche im April: R-R-R-R-R-R-R

Was sind Markov Modelle?

… weiter im Beispiel

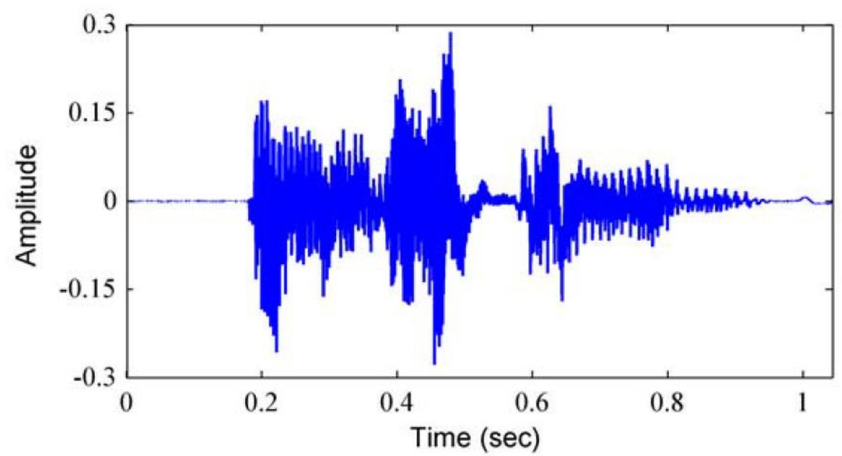
- jede Woche wird durch **Sequenz von Symbolen** (Sub-Muster oder **„States“**) beschrieben
- zwischen jeweils zwei aufeinander folgenden Symbolen gibt es einen **Übergang („Transition“)**
- Übergänge haben eine **Wahrscheinlichkeit**
  - z.B. Übergang von W → S ist wahrscheinlicher im Sommer, Übergang W → R ist wahrscheinlicher im April

Sequentielle Daten

Markov Modelle werden zur Modellierung und Klassifikation von sequentiellen Daten eingesetzt.

Sequentielle Daten ergeben sich:

- häufig aus zeitlichen Abfolgen
  - Wetter
  - Börsenkurse
  - Sprache
- aber auch in anderem Kontext
  - DNA Sequenz
  - Buchstaben eines Wortes



[Quelle: Duda et al.]

Sequentielle Daten

- zukünftige Entwicklungen sollen auf Grund der bisherigen geschätzt werden
  - z.B. Entwicklung des Börsenkurses
- es wäre unpraktisch/unrealistisch, wenn der nächste „State“ von allen bisherigen abhängt
  - Modelle wären zu komplex
- hier kommen Markov Modelle in Spiel → Schätzung des nächsten „States“ hängt nur von naher Vergangenheit ab

II. Grundlagen: "Markov Models" (MM)

Beispiel: Wettervorhersage

Ziel: Schätze wie das Wetter morgen sein wird basierend auf der Beobachtung des Wetters der Vortage.

Annahmen:

- **Vereinfachtes Modell** → Wetter hält ganzen Tag an
  d.h. ändert sich nicht von sonnig auf wolkig
- **Vereinfachte Statistik** → Wetter des aktuellen Tages w_n hängt von Wetter der vorherigen Tage $w_{n-1}, w_{n-2}, \dots$ ab
  d.h. man sammelt Daten über $P(w_n | w_{n-1}, w_{n-2}, \dots, w_1)$

Beispiel: Wettervorhersage

Wenn wir das Wetter der letzten drei Tage kennen: {sonnig, sonnig, bewölkt}, dann drückt sich die Wahrscheinlichkeit für Regen am morgigen Tag als $P(w_4 = R | w_3 = W, w_2 = S, w_1 = S)$ aus.

Problem: Umso größer n (Anzahl der Tage), umso mehr Informationen (Trainingsdaten) muss man sammeln.

Für $n = 5$ wären das $3^5 = 243$ Fälle!

Beispiel: Wettervorhersage

Lösung …

Markov Annahme (1. Ordnung):

$$P(w_n|w_{n-1}, w_{n-2}, \dots, w_1) \approx P(w_n|w_{n-1})$$

Wahrscheinlichkeit von w_n hängt nur von der Beobachtung w_{n-1}ab. Bei einer **Markov Annahme 2. Ordnung** berücksichtigt man w_{n-1}und w_{n-2}.

Diese Annahme hat einen großen Einfluss auf die nötigen Trainings-daten (Fälle). Bei einer Markov Annahme 1. Ordnung braucht man nur $3^2 = 9$.

Beispiel: Wettervorhersage

Wahrscheinlichkeiten für das morgige Wetter basierend auf dem heutigen Wetter $P(w_{\text{Morgen}}|w_{\text{Heute}})$:

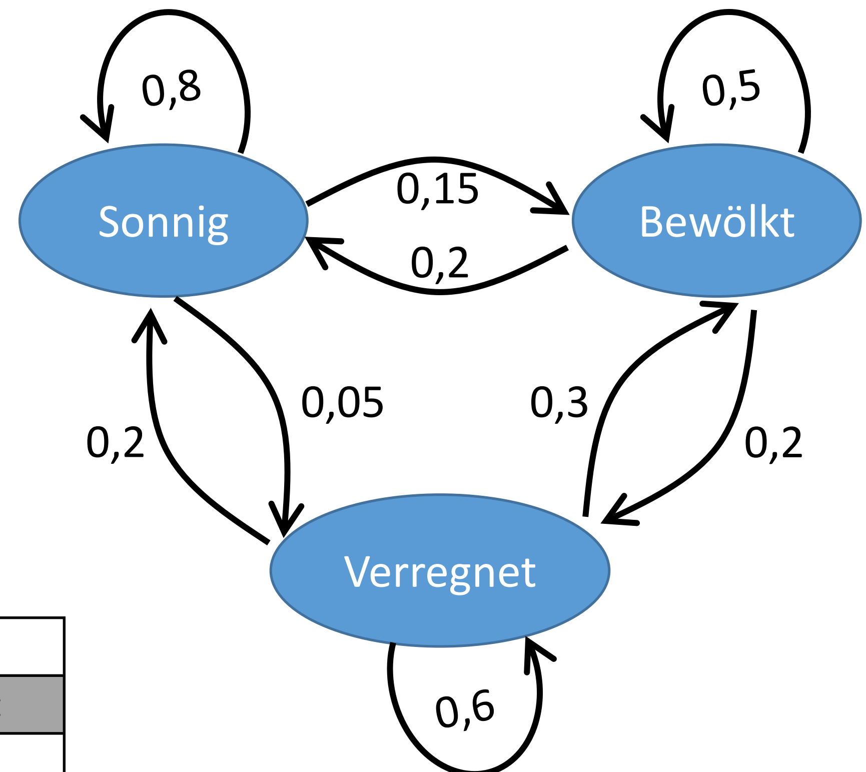
		morgiges Wetter		
		Sonnig	Verregnet	Bewölkt
heutiges Wetter	Sonnig (S)	0,8	0,05	0,15
	Verregnet (R)	0,2	0,6	0,2
	Bewölkt (W)	0,2	0,3	0,5

Die obige Tabelle zeigt die sogenannte „Übergangsmatrix“ (**State Transition Matrix**).

Beispiel: Wettervorhersage

… aus den Wahrscheinlichkeiten der Tabelle ergibt sich einem **Markov Modell 1. Ordnung**

… ein probabilistischer endlicher **Automat**



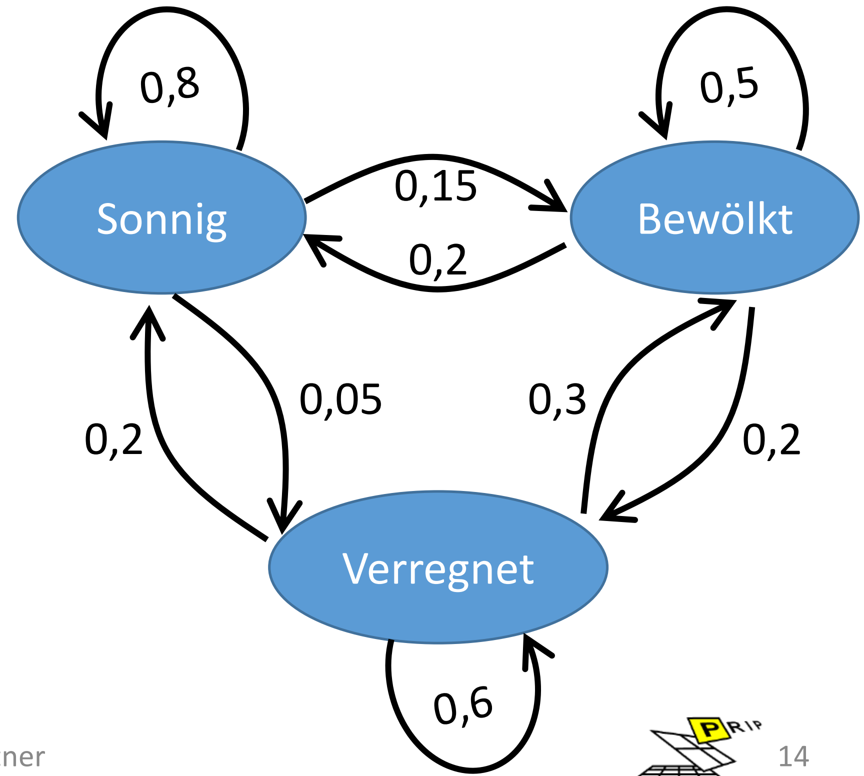
		morgiges Wetter		
		Sonnig	Verregnet	Bewölkt
heutiges Wetter	Sonnig (S)	0,8	0,05	0,15
	Verregnet (R)	0,2	0,6	0,2
	Bewölkt (W)	0,2	0,3	0,5

Fragen an das Markov Modell

… Fragen an den Wetter-Frosch

Angenommen heute ist es sonnig. Wie hoch ist die Wahrscheinlichkeit, dass es morgen sonnig bleibt und am Tag danach regnet?

$$P(w_3 = R|w_2 = S) \cdot P(w_2 = S|w_1 = S) = 0{,}05 \cdot 0{,}8 = 0{,}04$$



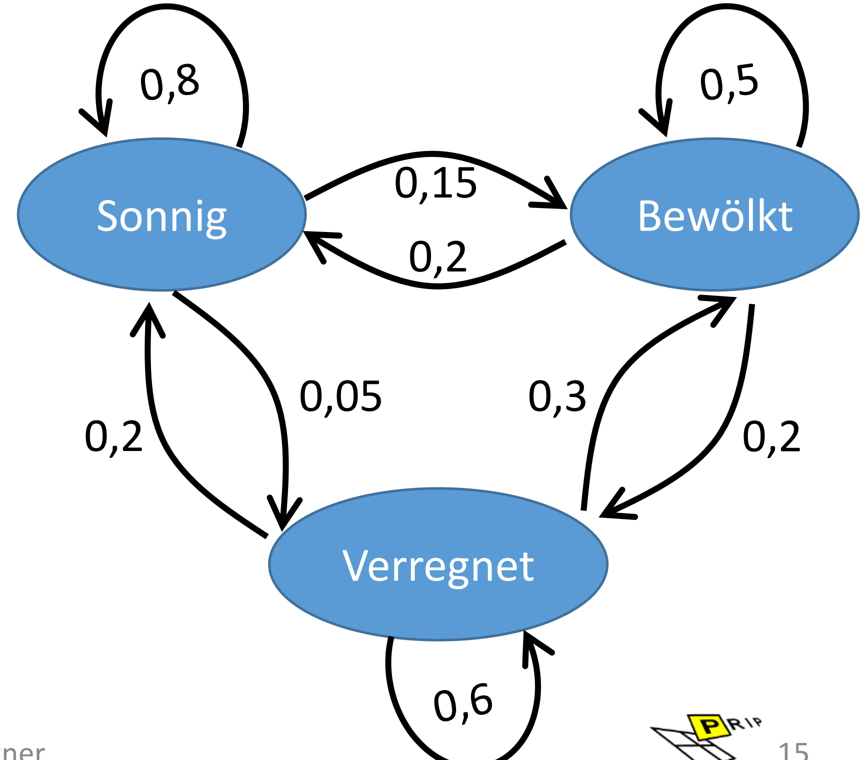
Fragen an das Markov Modell

… Fragen an den Wetter-Frosch

Angenommen heute ist es bewölkt. Wie hoch ist die Wahrscheinlichkeit, dass es in zwei Tagen regnet?

Hierfür gibt es drei mögliche Sequenzen:

1. {W,W,R}
2. {W,R,R}
3. {W,S,R}



Fragen an das Markov Modell

$P(w_3 = R | w_1 = W) =$

$P(w_3 = R | w_2 = W) P(w_2 = W | w_1 = W) +$ {W,W,R}

$P(w_3 = R | w_2 = R) P(w_2 = R | w_1 = W) +$ {W,R,R}

$P(w_3 = R | w_2 = S) P(w_2 = S | w_1 = W) =$ {W,S,R}

$0{,}3 \cdot 0{,}5 + 0{,}6 \cdot 0{,}3 + 0{,}05 \cdot 0{,}2 = 0{,}34$

III. Grundlagen: „Hidden Markov Models“ (HMM)

Hidden Markov Model

Was sind „Versteckte Markov Modelle“ oder „Hidden Markov Models“ (HMM)?

… zurück zum Beispiel der Wettervorhersage

Angenommen man wird in einen Raum ohne Fenster eingeschlossen und man wird nach dem Wetter draußen befragt. Der einzige Hinweis den man hat, ist die An- oder Abwesenheit eines Regenschirms bei der Person die täglich das Essen bringt.

	Wahrscheinlichkeit für Schirm (T)
Sonnig (S)	0,1
Verregnet (R)	0,8
Bewölkt (W)	0,3

Beispiel: Wettervorhersage

Wahrscheinlichkeit für „Wetter-Sequenz“ **im Markov Modell** („States“ sind sichtbar):

$$P(w_1, \dots, w_n) = \prod_{i=1}^{n} P(w_i | w_{i-1})$$

In HMM ist es nicht möglich das Wetter („States“) zu beobachten, deshalb benötigt man das Bayes-Theorem:

$$P(w_1, \dots, w_n | u_1, \dots, u_n) = \frac{P(u_1, \dots, u_n | w_1, \dots, w_n) P(w_1, \dots, w_n)}{P(u_1, \dots, u_n)}$$

u_i … ist „wahr“ (T) wenn Besucher Schirm dabei hat, sonst „falsch“ (F)

$P(w_1, \dots, w_n)$ … wie beim Markov Modell

$P(u_1, \dots, u_n)$ … Prior für eine bestimmte Schirm-Sequenz

$P(u_1, \dots, u_n | w_1, \dots, w_n)$ …schätzt man als $\prod_{i=1}^{n} P(u_i | w_i)$, falls w_i, u_i unabhängig von allen w_j, u_j, wobei $j \neq i$

Fragen an HMM

Annahme: $P(u = \mathrm{T}) = P(u = \mathrm{F}) = 0{,}5$

Angenommen am 1. Tag an dem man eingeschlossen wurde, war es sonnig. Am 2. Tag bringt der Besucher einen Schirm mit. Wie hoch ist die Wahrscheinlichkeit, dass es am 2. Tag regnet?

$$P(w_2 = R | w_1 = S, u_2 = T) =$$

$$\frac{P(u_2 = T | w_2 = R) P(w_2 = R | w_1 = S)}{P(u_2 = T)} = \frac{0{,}8 \cdot 0{,}05}{0{,}5} = 0{,}08$$

Fragen an HMM

Annahme: $P(u = \mathrm{T}) = P(u = \mathrm{F}) = 0{,}5$

Angenommen am 1. Tag an dem man eingeschlossen wurde, war es sonnig. Am 2. Tag bringt der Besucher einen Schirm mit am 3. aber nicht. Wie hoch ist die Wahrscheinlichkeit, dass es am 3. Tag bewölkt ist?

Laut endlichem Automaten gibt es drei Möglichkeiten:

1. {S,W,W}
2. {S,R,W}
3. {S,S,W}

Fragen an HMM

$$P(w_3 = W | w_1 = S, u_2 = T, u_3 = F) =$$

$$\frac{P(u_3 = F | w_3 = W) P(u_2 = T | w_2 = W) P(w_3 = W | w_2 = W) P(w_2 = W | w_1 = S)}{P(u_3 = F) P(u_2 = T)} +$$

$$\frac{P(u_3 = F | w_3 = W) P(u_2 = T | w_2 = R) P(w_3 = W | w_2 = R) P(w_2 = R | w_1 = S)}{P(u_3 = F) P(u_2 = T)} +$$

$$\frac{P(u_3 = F | w_3 = W) P(u_2 = T | w_2 = S) P(w_3 = W | w_2 = S) P(w_2 = S | w_1 = S)}{P(u_3 = F) P(u_2 = T)} =$$

Fragen an HMM

$P(w_3 = W | w_1 = S, u_2 = T, u_3 = F) =$

$$\frac{0{,}7 \cdot 0{,}3 \cdot 0{,}5 \cdot 0{,}15}{0{,}5 \cdot 0{,}5} + \frac{0{,}7 \cdot 0{,}8 \cdot 0{,}2 \cdot 0{,}05}{0{,}5 \cdot 0{,}5} + \frac{0{,}7 \cdot 0{,}1 \cdot 0{,}15 \cdot 0{,}8}{0{,}5 \cdot 0{,}5} = 0{,}119$$

IV. Training von HMM

Training eines HMM

- „States“ (Sub-Muster) sind versteckt (nicht beobachtbar)
- Beobachtungen die auf die versteckten Sub-Muster schließen lassen können verrauscht sein
- Trainingsdaten sind „Beobachtungssequenzen“
- meistens ist die Sequenz der Sub-Muster (States) für jede Klasse im vorhinein bekannt
  - z.B.: Buchstabensequenzen von Wörtern

Bestandteile eines HMM

N … Anzahl der „States“ (Sub-Muster) im Model

- Wetterstatus: sonnig, verregnet und bewölkt
- Reihen der Zahlenmatrix (siehe rechts)

0	0	1
0	0	1
0	0	1

M … Anzahl der möglichen Beobachtungen (Ausprägungen)

- Schirm: „True“ oder „False“
- Zahlen: {000}, {001}, {100}, {010}, {101}, {011}, {110}, {111},

L … Länge der Beobachtungssequenz

- Anzahl der Beobachtungen
- $\mathcal{O} = (O_1, O_2, \dots, O_L)$

Parameter eines HMM

… mit Hilfe der Trainingsdaten wird ein HMM $\lambda = (A, B, \pi)$ je Klasse ermittelt

Übergangsmatrix (State Transition Matrix):

$$A = [a_{ij}]; a_{ij} = P(S_{t+1} = j | S_t = i)$$

Wahrscheinlichkeit für den Übergang von State i zu State j zum Zeitpunkt t.

Parameter eines HMM

… mit Hilfe der Trainingsdaten wird ein HMM $\lambda = (A, B, \pi)$ je Klasse ermittelt

Beobachtungswahrscheinlichkeit (Observation Probability):

$$b_j(k) = P(O_t = k | S_t = j), 1 \leq k \leq M$$

Wahrscheinlichkeit die Beobachtung k zum Zeitpunkt t zu machen, wenn der „State“ j ist.

Hinweis: In dieser VO arbeiten wir mit diskreten HMM. Daher gibt es auch eine endlich Anzahl an Ausprägungen/Beobachtungen.

Parameter eines HMM

… mit Hilfe der Trainingsdaten wird ein HMM $\lambda = (A, B, \pi)$ je Klasse ermittelt

Anfangs-State-Wahrscheinlichkeit (Initial State Probability):

$$\pi_i = P\left(S_{t_1} = i\right), 1 \leq i \leq N$$

Wahrscheinlichkeit das Sequenz mit State i beginnt.

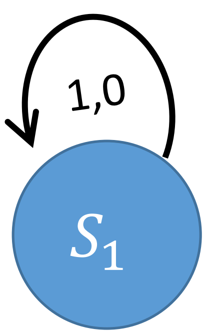
Beispiel: Zahlen

- 3x3 Muster für Zahlen 1 und 7
- jede Zeile im Muster ist ein „State“
- {001} ist State S_1 und {111} ist State S_2

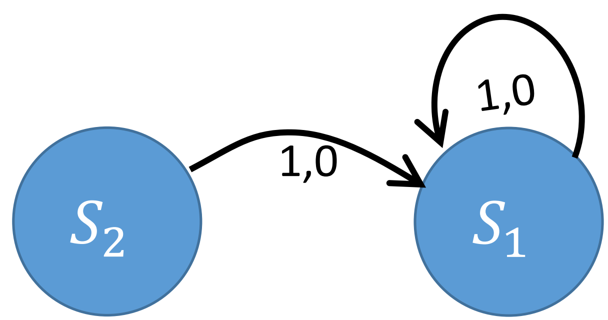
0	0	1
0	0	1
0	0	1

1	1	1
0	0	1
0	0	1

Zahl 1: S_1, S_1, S_1



Zahl 7: S_2, S_1, S_1



Beispiel: Zahlen

- Trainingsdaten für Zahl „1“ mit Noise

0	0	1
0	0	1
0	0	1

1	0	1
0	0	1
0	0	1

0	1	1
0	0	1
0	0	1

0	0	0
0	0	1
0	0	1

0	0	1
0	0	1
0	0	1

0	0	1
1	0	1
0	0	1

0	0	1
0	1	1
0	0	1

0	0	1
0	0	0
0	0	1

0	0	1
0	0	1
0	0	1

0	0	1
0	0	1
1	0	1

0	0	1
0	0	1
0	1	1

0	0	1
0	0	1
0	0	0

Beispiel: Zahlen

Training für Zahl (Klasse) „1“

- 12 Trainingsmuster
  - je 3 Beobachtungen O
  - ein Trainingsmuster ist eine Sequenz $\mathcal{O} = (O_1, O_2, O_3)$
  - 36 Beobachtungen von zwei möglichen „States“ S_1 und S_2
- 2 Cluster
  - Repräsentant von Cluster 1: {0,0,1} S_1
  - Repräsentant von Cluster 2: {1,1,1} S_2

Beispiel: Zahlen

1. Clustering der Trainingsdaten:

- jede Beobachtung O wird einem Cluster zugeordnet
- abhängig von der Distanz zu den jeweiligen Repräsentanten

im Beispiel:

- Hamming-Distanz: zählt die nicht übereinstimmenden Bits
- 30/36 Beobachtungen werden Cluster 1 zugeordnet
- 6 Beobachtungen sind von beiden Clustern gleich weit entfernt → werden in diesem Beispiel Cluster 2 zugewiesen

Beispiel: Zahlen

Berechnung der Parameter des HMM für Zahl „1“

2. Anfangs-State-Wahrscheinlichkeit:

$$\pi_i = \frac{\text{Anzahl der Fälle in denen } O_1 \text{ dem Cluster } i \text{ zugeordnet wird}}{\text{Gesamtanzahl der Muster}}$$

<u>im Beispiel:</u>

$$\pi_1 = \frac{10}{12} = \frac{5}{6} \text{ und } \pi_2 = \frac{2}{12} = \frac{1}{6}$$

Beispiel: Zahlen

Berechnung der Parameter des HMM für Zahl „1“

3. Übergangsmatrix:

$$a_{ij} = \frac{\text{Anzahl der Fälle in denen } O_t \text{ in Cluster } i \text{ liegt und } O_{t+1} \text{ in Cluster } j}{\text{Anzahl der Fälle in denen } O_t \text{ in Cluster } i \text{ liegt}}$$

im Beispiel:

$$a_{11} = \frac{16}{20} = \frac{4}{5} \text{ und } a_{12} = \frac{4}{20} = \frac{1}{5}$$

$$a_{21} = \frac{4}{4} = 1 \text{ und } a_{22} = \frac{0}{4} = 0$$

Beispiel: Zahlen

Berechnung der Parameter des HMM für Zahl „1“

4. Beobachtungswahrscheinlichkeit

Schritt 1: Berechne die Distanzen $d_1, d_2, \dots, d_N$ zwischen „Beobachtungsvektor“ k und jedem der N „States“.

$$b_j(k) = 1 - \frac{d_j}{\text{maximal mögliche Distanz}}$$

Schritt 2: Normalisieren der Distanzen, sodass $\sum_j b_j(k) = 1$.

Beispiel: Zahlen

$M = 4$ … Beobachtungen für Zahl (Klasse) „1“

0	0	1
0	0	1
0	0	1

1	0	1
0	0	1
0	0	1

0	1	1
0	0	1
0	0	1

0	0	0
0	0	1
0	0	1

0	0	1
0	0	1
0	0	1

0	0	1
1	0	1
0	0	1

0	0	1
0	1	1
0	0	1

0	0	1
0	0	0
0	0	1

0	0	1
0	0	1
0	0	1

0	0	1
0	0	1
1	0	1

0	0	1
0	0	1
0	1	1

0	0	1
0	0	1
0	0	0

Beispiel: Zahlen

- mögliche Beobachtungen für Zahl „1“: {001}, {000}, {101},{011}
- maximale Distanz = 3; $S_1 = \{001\}$ und $S_2 = \{111\}$

Schritt 1: $b_1(001) = 1 - \frac{0}{3} = 1$ und $b_2(001) = 1 - \frac{2}{3} = \frac{1}{3}$

Schritt 2: $b_1(001) = \frac{1}{\frac{4}{3}} = \frac{3}{4}$ und $b_2(001) = \frac{\frac{1}{3}}{\frac{4}{3}} = \frac{1}{4}$

alle anderen Ergebnisse:

$$b_1(000) = 1 \text{ und } b_2(000) = 0$$

$$b_1(101) = \frac{1}{2} \text{ und } b_2(101) = \frac{1}{2}$$

$$b_1(011) = \frac{1}{2} \text{ und } b_2(011) = \frac{1}{2}$$

Beispiel: Zahlen

$M = 6$ … Beobachtungen für Zahl (Klasse) „7“

1	1	1
0	0	1
0	0	1

0	1	1
0	0	1
0	0	1

1	0	1
0	0	1
0	0	1

1	1	0
0	0	1
0	0	1

1	1	1
0	0	1
0	0	1

1	1	1
1	0	1
0	0	1

1	1	1
0	1	1
0	0	1

1	1	1
0	0	0
0	0	1

1	1	1
0	0	1
0	0	1

1	1	1
0	0	1
1	0	1

1	1	1
0	0	1
0	1	1

1	1	1
0	0	1
0	0	0

Beispiel: Zahlen

HMM für Zahl (Klasse) „7“

$\pi_1 = 0$	$\pi_2 = 1$
$a_{11} = \frac{4}{5}$ und $a_{12} = \frac{1}{5}$	$a_{21} = \frac{6}{7}$ und $a_{22} = \frac{1}{7}$
$b_1(111) = \frac{1}{4}$ und $b_2(111) = \frac{3}{4}$ $b_1(011) = \frac{1}{2}$ und $b_2(011) = \frac{1}{2}$ $b_1(110) = 0$ und $b_2(110) = 1$ $b_1(001) = \frac{3}{4}$ und $b_2(001) = \frac{1}{4}$ $b_1(101) = \frac{1}{2}$ und $b_2(101) = \frac{1}{2}$ $b_1(000) = 1$ und $b_2(000) = 0$	

Hinweis: Beobachtungswahrscheinlichkeiten sind unabhängig von der Klasse.

V. Klassifikation mit HMM

Klassifikation

Zur Klassifikation wird die **Wahrscheinlichkeit jedes HMMs** für ein Muster $m = (m_1, \dots, m_L)$ ermittelt:

$$P(m|\lambda) = \pi \cdot b(m_1) \cdot a \cdot b(m_2), \dots, a \cdot b(m_L)$$

π … Anfangs-State-Wahrscheinlichkeit für m_1

b … Beobachtungswahrscheinlichkeit

a … Wahrscheinlichkeit für Übergang (State Transition)

Beispiel: Zahlen

$m = (001{,}001{,}001)$

$P(m|\lambda_1) = \pi_1 \cdot b_1(001) \cdot a_{11} \cdot$
$b_1(001) \cdot a_{11} \cdot b_1(001) =$

$\frac{5}{6} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} = \mathbf{0{,}225}$

HMM λ_1	
$\pi_1 = \frac{5}{6}$	$\pi_2 = \frac{1}{6}$
$a_{11} = \frac{4}{5}$ $a_{12} = \frac{1}{5}$	$a_{21} = 1$ $a_{22} = 0$
$b_1(011) = \frac{1}{2}$ und $b_2(011) = \frac{1}{2}$ $b_1(001) = \frac{3}{4}$ und $b_2(001) = \frac{1}{4}$ $b_1(101) = \frac{1}{2}$ und $b_2(101) = \frac{1}{2}$ $b_1(000) = 1$ und $b_2(000) = 0$	

Beispiel: Zahlen

$m = (001{,}001{,}001)$

$$P(m|\lambda_7) = \pi_1 \cdot b_1(001) \cdot a_{11} \cdot b_1(001) \cdot a_{11} \cdot b_1(001) =$$

$$0 \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} = 0$$

HMM λ_7	
$\pi_1 = 0$	$\pi_2 = 1$
$a_{11} = \frac{4}{5}$ $a_{12} = \frac{1}{5}$	$a_{21} = \frac{6}{7}$ $a_{22} = \frac{1}{7}$
$b_1(111) = \frac{1}{4}$ und $b_2(111) = \frac{3}{4}$ $b_1(011) = \frac{1}{2}$ und $b_2(011) = \frac{1}{2}$ $b_1(110) = 0$ und $b_2(110) = 1$ $b_1(001) = \frac{3}{4}$ und $b_2(001) = \frac{1}{4}$ $b_1(101) = \frac{1}{2}$ und $b_2(101) = \frac{1}{2}$ $b_1(000) = 1$ und $b_2(000) = 0$	

Beispiel: Zahlen

$m = (101,001,001)$

$P(m|\lambda_1) = \pi_2 \cdot b_2(101) \cdot a_{21} \cdot$

$b_1(001) \cdot a_{11} \cdot b_1(001) =$

$\frac{1}{6} \cdot \frac{1}{2} \cdot 1 \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} = 0{,}0375$

HMM λ_1	
$\pi_1 = \frac{5}{6}$	$\pi_2 = \frac{1}{6}$
$a_{11} = \frac{4}{5}$ $a_{12} = \frac{1}{5}$	$a_{21} = 1$ $a_{22} = 0$
$b_1(011) = \frac{1}{2}$ und $b_2(011) = \frac{1}{2}$ $b_1(001) = \frac{3}{4}$ und $b_2(001) = \frac{1}{4}$ $b_1(101) = \frac{1}{2}$ und $b_2(101) = \frac{1}{2}$ $b_1(000) = 1$ und $b_2(000) = 0$	

Beispiel: Zahlen

$m = (101,001,001)$

$P(m|\lambda_7) = \pi_2 \cdot b_2(101) \cdot a_{21} \cdot b_1(001) \cdot a_{11} \cdot b_1(001) =$

$1 \cdot \frac{1}{2} \cdot \frac{6}{7} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{3}{4} =$ **0,193**

HMM λ_7	
$\pi_1 = 0$	$\pi_2 = 1$
$a_{11} = \frac{4}{5}$ $a_{12} = \frac{1}{5}$	$a_{21} = \frac{6}{7}$ $a_{22} = \frac{1}{7}$
$b_1(111) = \frac{1}{4}$ und $b_2(111) = \frac{3}{4}$ $b_1(011) = \frac{1}{2}$ und $b_2(011) = \frac{1}{2}$ $b_1(110) = 0$ und $b_2(110) = 1$ $b_1(001) = \frac{3}{4}$ und $b_2(001) = \frac{1}{4}$ $b_1(101) = \frac{1}{2}$ und $b_2(101) = \frac{1}{2}$ $b_1(000) = 1$ und $b_2(000) = 0$	

VI. Anwendungen

Einleitung

- HMMs werden für die Verarbeitung von sequentiellen Daten eingesetzt
  - z.B. Spracherkennung, Handschrifterkennung, etc.
- HMMs werden konstruiert, sodass ihre verborgenen „States“ semantischen Einheiten entsprechen
  - z.B. Phoneme, Buchstaben, etc.
- Ziel: „States“ und Sequenzen in sequentiellen Daten (z.B. Sprachsignale) zu erkennen

Spracherkennung

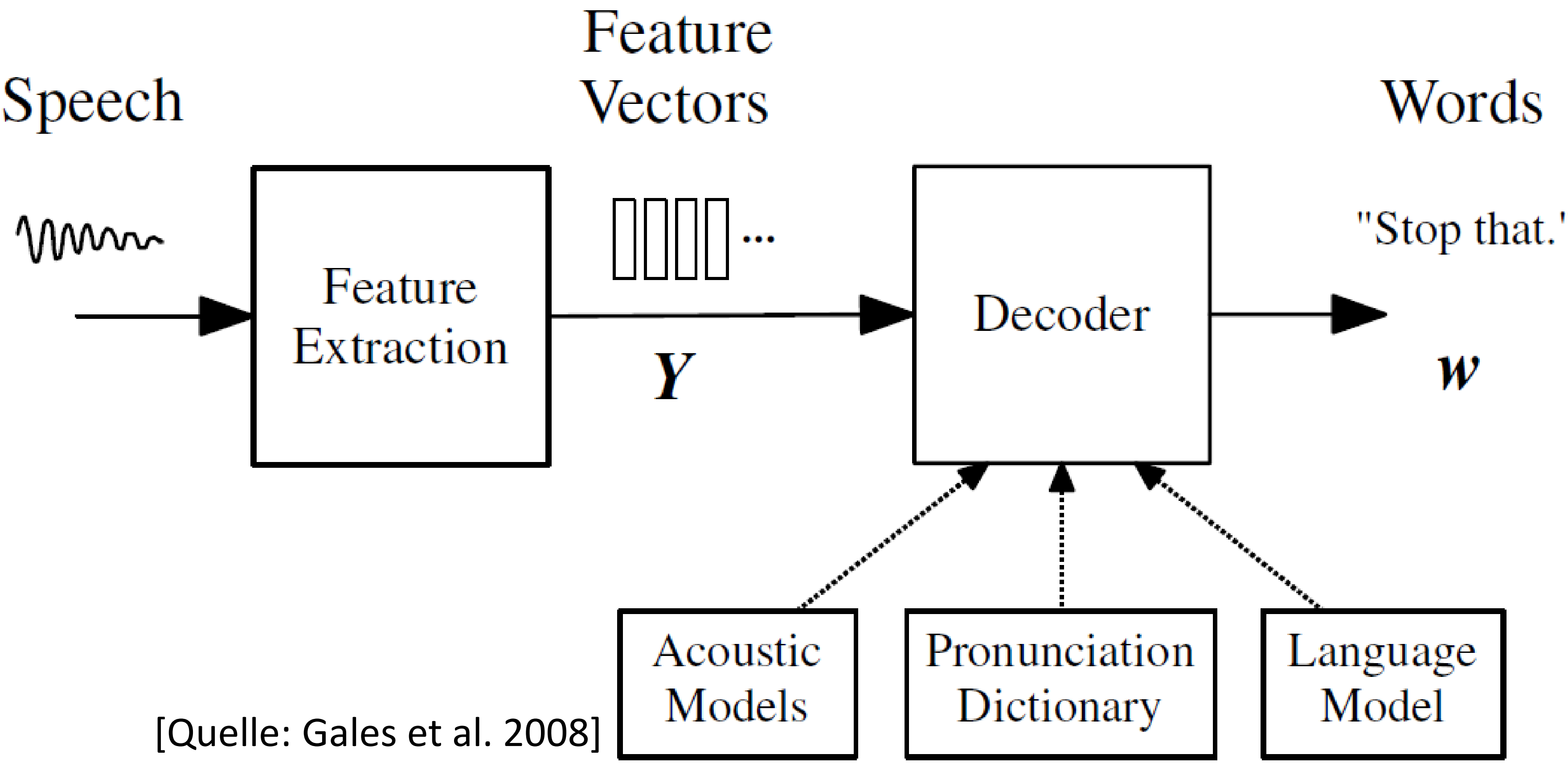
- bereits als kommerzielle Software verfügbar
- auf unseren Smartphones
- in einer kontrollierten Umgebung können ausgezeichnete Ergebnisse erzielt werden
- ABER: Problem der Spracherkennung nicht gelöst
- existierende Lösungen können nicht mit stark verrauschte Daten umgehen
  - z.B. Sprachaufnahme einer Präsentation auf einer Konferenz

Spracherkennung

- Tonspur wird in sogenannte „Frames“ mit fixer Länge in ms zerlegt
- Frames überlappen
- die meisten Systeme extrahieren daraus Mel-Frequenz-Cepstrum-Koeffizienten als Features
  - erlauben kompakte Darstellung des Frequenzspektrums
- akustische Events werden durch kontinuierliche HMMs modelliert

Sprachenkennung

… so könnte eine System aufgebaut sein



[Quelle: Gales et al. 2008]

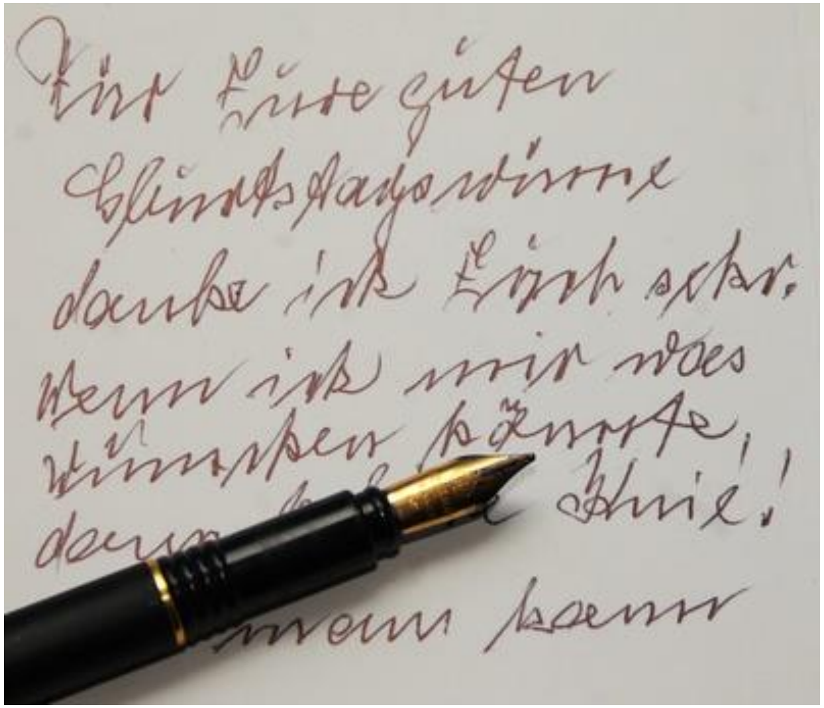
Handschrifterkennung

- im Vergleich zur Spracherkennung ist die Segmentierung in einzelnen Wörter einfacher
- Aufgaben:
  - optische Buchstabenerkennung (OCR)
  - automatische Analyse von Formularen
  - Worterkennung

Handschrifterkennung

Vorverarbeitung:

- optische Erfassung der Handschrift (Scann)
- Entzerrung
- Kontrastverbesserung
- Segmentierung des Bildes in Textzeilen



Handschrifterkennung

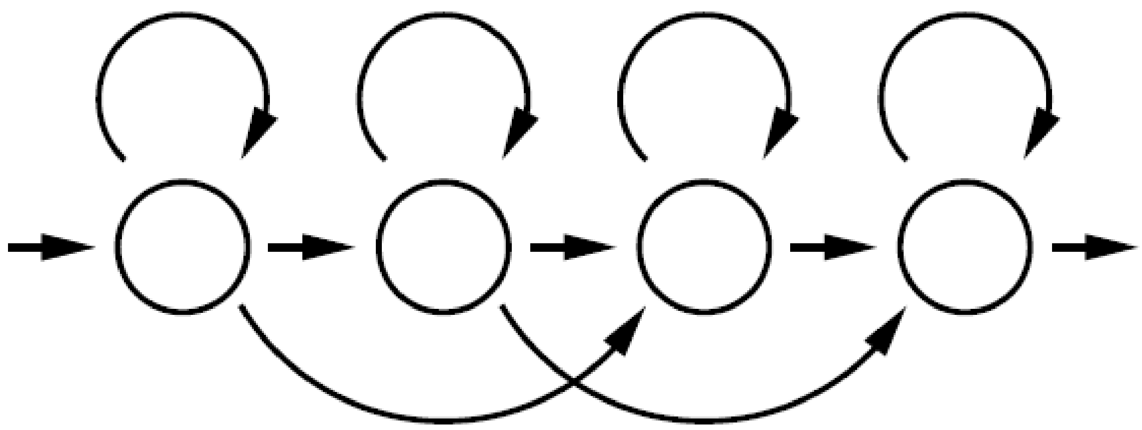
Feature Extraktion:

- häufig wird die „Sliding-Window“ Methode verwendet
- dabei wird Textzeile in überlappende kleine Bereiche unterteilt (z.B. 1/15 der Zeile)
- aus diesen Bereichen wird je ein Feature-Vektor extrahiert
- oft wird LDA (=Linear Discriminant Analysis) verwendet um die Dimensionalität zu reduzieren

Handschrifterkennung

Zeichenmodellierung:

- mit Hilfe vom HMMs
- Feature-Vektoren sind die Beobachtungen mit denen man auf die „hidden“ States schließt
- oft wird ein Bakis-Modell eingesetzt
  - diese Modelle erlauben es „States“ zu über springen



Handschrifterkennung

Sprachmodellierung:

- mit Hilfe von n-Gramm-Modellen
- statistisches Sprachmodell
- Sequenzen von n Elementen aus einem Text
- Elemente können Sprachlaute, Silben, Buchstaben oder Wörter sein
- ermöglicht die Schätzung des nächsten Elements
- z.B.: 3-Gramm für „Guten Morgen"
  - {Gut} {ute} {ten} {en } {n M} { Mo} {Mor} {org} {rge} {gen}

Letzte VO
Do, 14.01.2016