6.4 Hidden – Markow - Modelle

6.4.1 Definition

Idee

- Ein Referenzmuster wird durch einen stochastischen Prozess, d.h. durch eine Zufallsgröße $\xi(t)$, die sich im Verlauf der Zeit t ändern kann, beschrieben.
- insbesondere betrachtet man hier Prozesse mit einer diskreten Zeitvariablen

Definition
Hidden – Markov – Modell (HMM)

$$\boxed{\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)}$$

$S = \left\{S^1, S^2, \ldots, S^N\right\}$ *endliche Menge von* Zuständen

$S = \left\{s : 1 \leq s \leq N\right\}$ werden oft nur mit ihren Indizes bezeichnet

manchmal auch $S = \left\{S_1, S_2, \ldots, S_N\right\}$

$S_t \in S \qquad t = 1,2,3,\ldots$ Zustand zum Zeitpunkt t

$$P(S_t \mid S_1, S_2, \ldots, S_{t-1}) = P(S_t \mid S_{t-1})$$

Definition
Hidden – Markov – Modell (HMM)

$$\boxed{\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)}$$

$$M = \left\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^L\right\}$$

endliche Menge von Merkmalsvektoren
(Beobachtungen, akustische Ereignisse)
(diskrete HMM – DDHMM)

$$M = \left\{\mathbf{x} \in R^n\right\}$$

stetige Zufallsvariable (kontinuierliche HMM – CDHMM)

$$O_t \in M$$

Beobachtung zum Zeitpunkt t

$$P(O_t \mid O_1, O_2, \ldots, O_{t-1}, S_1, S_2, \ldots, S_t) = P(O_t \mid S_t)$$

Beobachtungsfolge: $o = o_1, o_2, \ldots, o_T \qquad 1 \leq o_i \leq L$

Definition

$$\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)$$

Vektor der Anfangswahrscheinlichkeiten

$$\mathbf{e} = \left(e_1, e_2, \ldots, e_N\right) \qquad e_i = P\left(S_1 = i\right)$$

Wahrscheinlichkeit, dass sich der Prozess
zum Zeitpunkt t = 1 *im Zustand i befindet*

$$e_i \geq 0 \qquad \sum_{i=1}^{N} e_i = 1$$

Definition

$$\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)$$

Matrix der Übergangswahrscheinlichkeiten:

$$\mathbf{A} = (a_{ij})$$

$$i, j = 1, \ldots, N$$

$$\boxed{a_{ij} = P(S_t = j \mid S_{t-1} = i)}$$

(unabhängig von t)

$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad 1 \leq \mathrm{i} \leq \mathrm{N}$$

(Ausnahme: Endzustand – Summe = 0)

Definition – diskrete HMM

$$\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)$$

Matrix der Auftrittswahrscheinlichkeiten

$$\mathbf{B} = (b_{jk}) \qquad b_{jk} = P(O_t = \mathbf{x}^k \mid S_t = j)$$

$$j = 1, \ldots, N \quad k = 1, \ldots, L \qquad b_{jk} = P(o_t = k \mid S_t = j)$$

$$b_{jk} \geq 0 \qquad \sum_{k=1}^{L} b_{jk} = 1 \quad 1 \leq \mathrm{j} \leq \mathrm{N}$$

Definition – kontinuierliche HMM

$$\lambda = \left(S, M, \mathbf{e}, \mathbf{A}, \mathbf{B}\right)$$

$$b_j(\mathbf{x}) = p(\mathbf{x} \mid S_t = j)$$

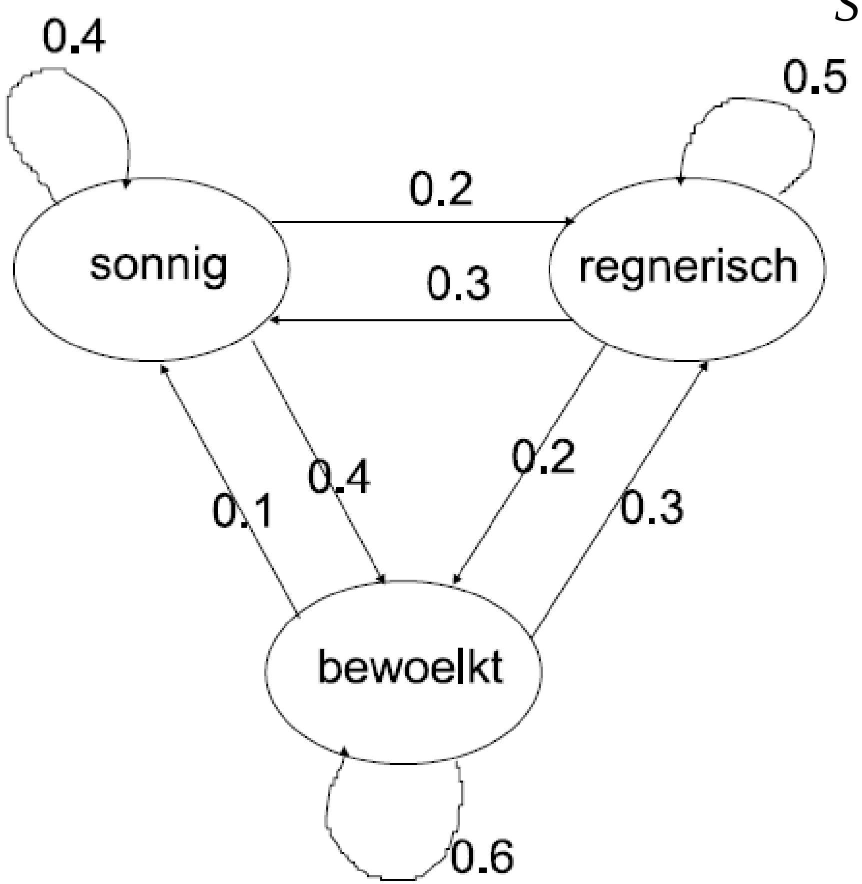
Dichtefunktion

Bemerkungen

- HMM beschreiben einen zweistufigen stochastischen Prozess
  - In der ersten Stufe handelt es sich um einen diskreten stochastischen Prozess. Die betrachtete Zustandsmenge ist dabei endlich. Der Prozess beschreibt probabilistisch die Zustandsübergänge in diesem diskreten, endlichen Zustandsraum.
  - In der zweiten Stufe wird dann zusätzlich zu jedem Zeitpunkt t eine Ausgabe (Beobachtung, Emission) generiert.
- Diese Folge von Ausgaben ist als Einziges vom Verhalten des Modells beobachtbar, die eingenommene Zustandsfolge dagegen nicht. Diese ist versteckt (hidden).
- Im Bereich der automatischen Spracherkennung haben HMM alle konkurrierenden Ansätze verdrängt.
- Aktuelle Anwendungen von HMMs für Signalanalyseaufgaben verwenden meist kontinuierliche HMMs.

Beispiel

Ein Zeitungsleser in Australien bekommt jeden Tag die Temperaturwerte in Deutschland geliefert, aber keine sonstige Information über das Wetter, welches nur abhängig vom Wetter des Vortages ist.



$$S = \left\{S^1(\text{sonnig}), S^2(\text{bewoelkt}), S^3(\text{regnerisch})\right\}$$

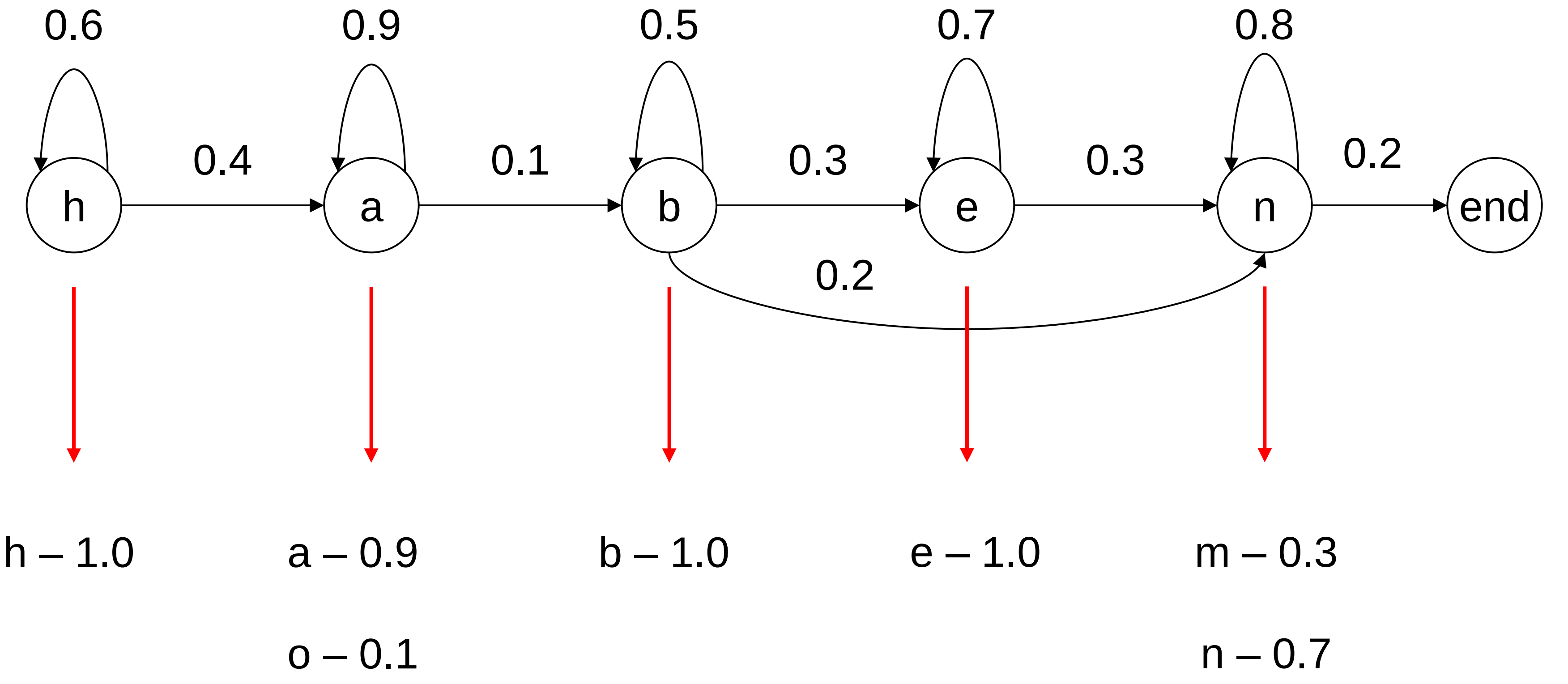
$$M = \left\{0°, 10°, 20°, 30°\right\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

Beispiel - haben



Beispiel – haben

- begonnen wird im Zustand 1
- Begonnen wird also mit dem Phonem [h]. Im Zustand 1 wird nur dieses Phonem ausgegeben. Mit einer Wahrscheinlichkeit von 0.6 dauert das [h] noch ein weiteres Zeitintervall (10 ms), mit einer Wahrscheinlichkeit von 0.4 erfolgt der Übergang zum Zustand 2.
- Im Zustand 2 können 2 Phoneme ([a], [o]) ausgegeben werden. Mit Wahrscheinlichkeit 0.9 aber das [a]. Es ist sehr wahrscheinlich, dass das [a] länger ausgesprochen wird, lediglich mit 0.1 erfolgt der Wechsel zum Zustand 3.
- Vom Zustand 3 (Phonem [b]) aus gibt es eine weitere Wahlmöglichkeit. Das [e] (Zustand 4) kann verschluckt werden und es wird direkt zum vorletzten Zustand gesprungen.
- Eine mit diesem Modell erzeugte Phonemfolge könnte z.B.

  [h][h][a][a][a][a][a][a][a][b][b][m][m][m]

  sein.

Arten von HMM

- ergodisch, wenn alle Zustände paarweise miteinander verbunden sind (alle Übergangswahrscheinlichkeiten sind größer 0)
- Links – Rechts – Modell
- Lineares HMM

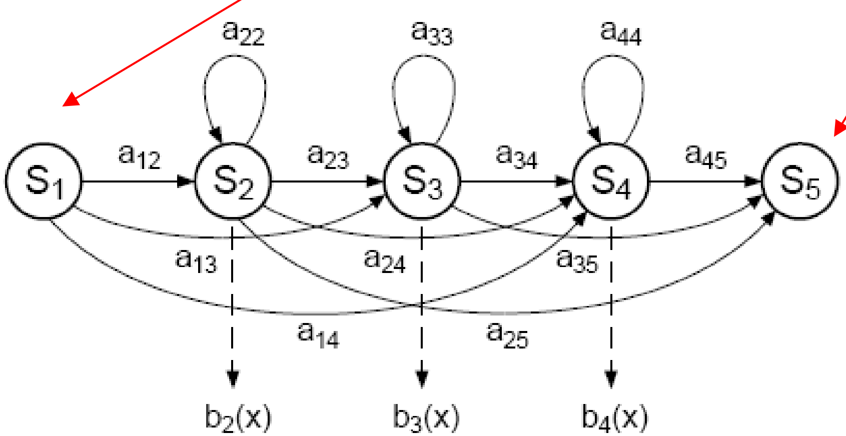
Links – Rechts – HMM

ein Zustand, der einmal verlassen wurde, kann zu keinem späteren Zeitpunkt nochmals besucht werden

$$a_{ij} = 0 \quad j < i$$

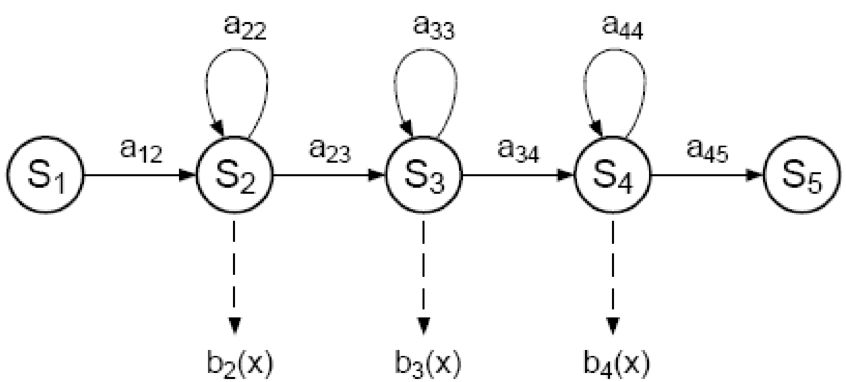
es existiert ein ausgezeichneter Anfangszustand S_1

es existiert ein ausgezeichneter Endzustand



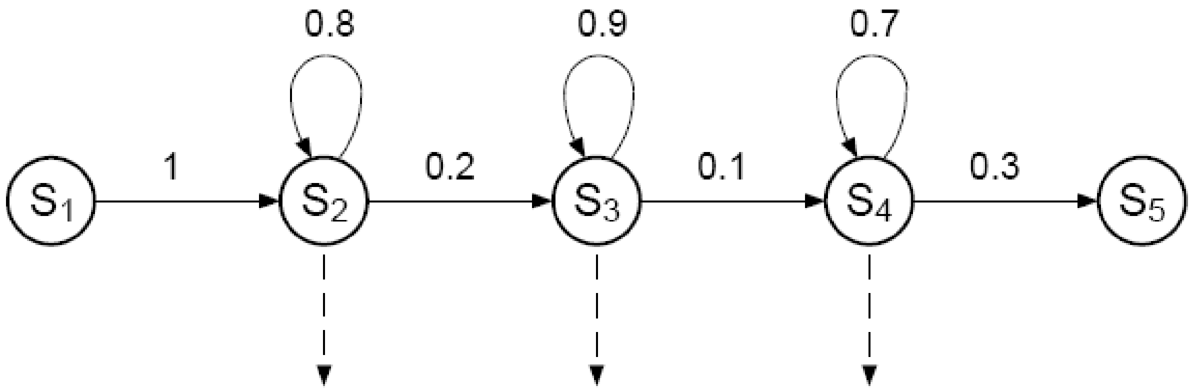
$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} & 0 \\ 0 & a_{22} & a_{23} & a_{24} & a_{25} \\ 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Lineares HMM



$$A = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Lineares Modell

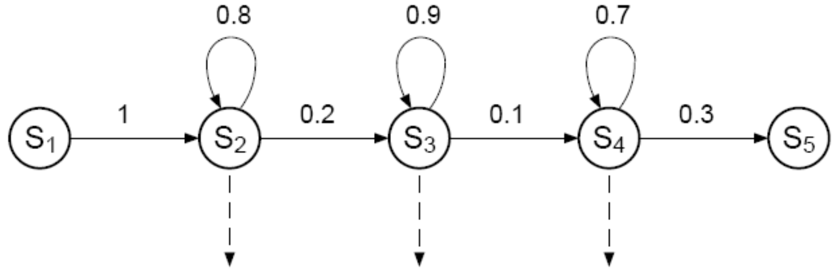


x	$P(x\|S_2)$
1	0.04
2	0.12
3	0.08
4	0.76

x	$P(x\|S_3)$
1	0.03
2	0.91
3	0.04
4	0.02

x	$P(x\|S_4)$
1	0.68
2	0.11
3	0.05
4	0.16

Lineares Modell



x	$P(x \mid S_2)$
1	0.04
2	0.12
3	0.08
4	0.76

x	$P(x \mid S_3)$
1	0.03
2	0.91
3	0.04
4	0.02

x	$P(x \mid S_4)$
1	0.68
2	0.11
3	0.05
4	0.16

Wahrscheinlichkeit, dass das HMM die Zustandsfolge s=1223445 durchläuft:

$$P(s) = a_{12}a_{22}a_{23}a_{34}a_{44}a_{45} = 1 \cdot 0.8 \cdot 0.2 \cdot 0.1 \cdot 0.7 \cdot 0.3 = 0.00336$$

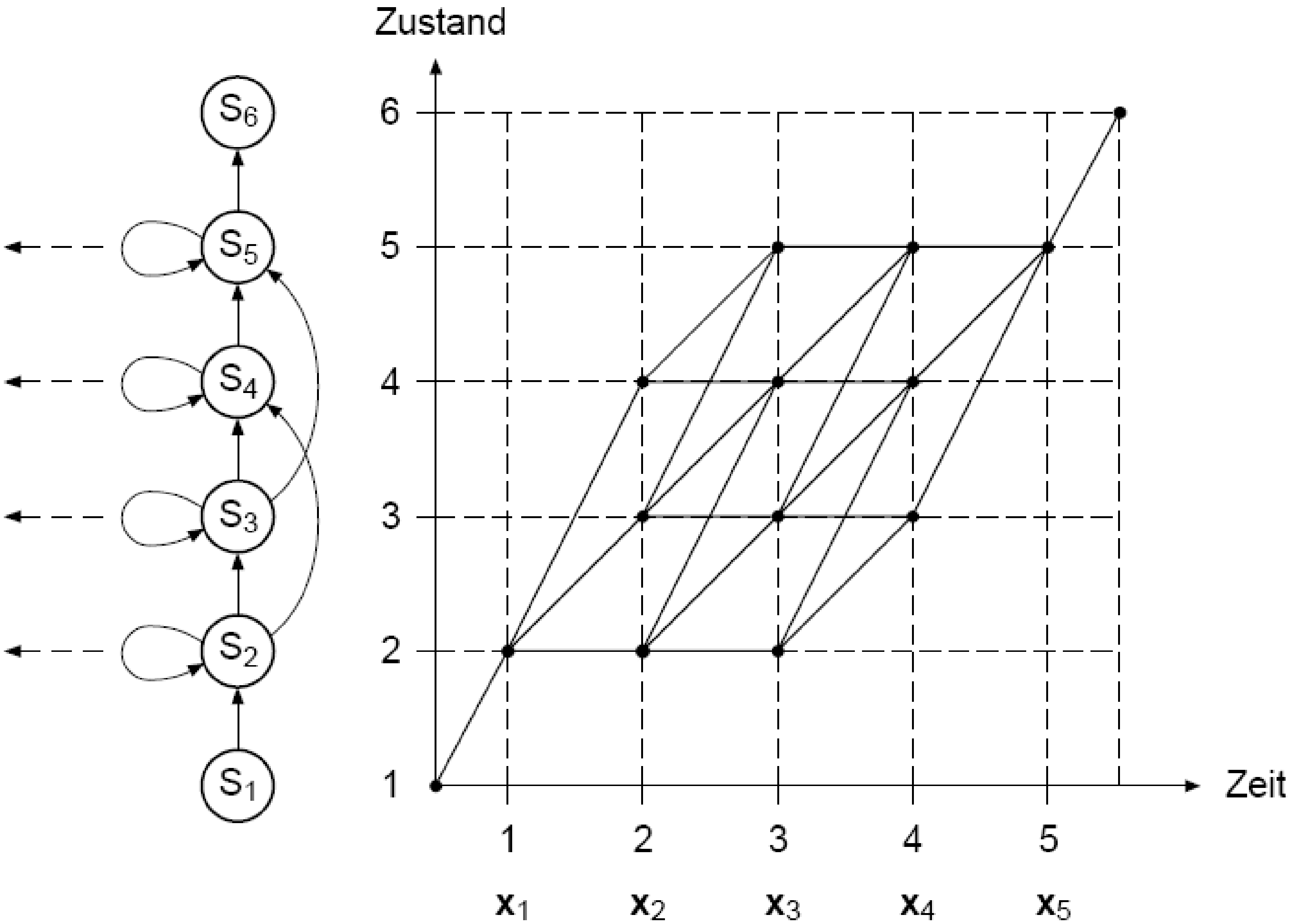
Wahrscheinlichkeit, dass das HMM die Beobachtungsfolge o=24214 erzeugt, wenn es s durchläuft:

$$P(o \mid s) = b_{22} \cdot b_{24} \cdot b_{32} \cdot b_{41} \cdot b_{44} = 0.12 \cdot 0.76 \cdot 0.91 \cdot 0.68 \cdot 0.16 \approx 0.00903$$

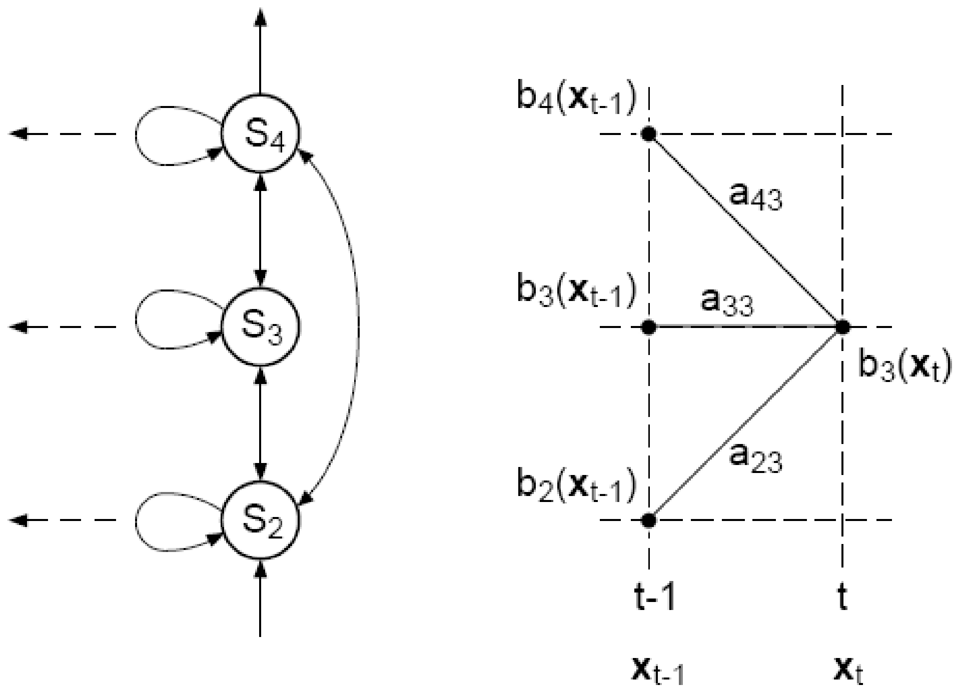
wir können auch berechnen, wie wahrscheinlich es ist, dass das HMM die Zustandsfolge s durchläuft und dabei die Beobachtungsfolge o erzeugt:

$$P(s,o) = P(o \mid s) \cdot P(s) \approx 0.00003$$

Trellis – Diagramm



Trellis – Diagramm



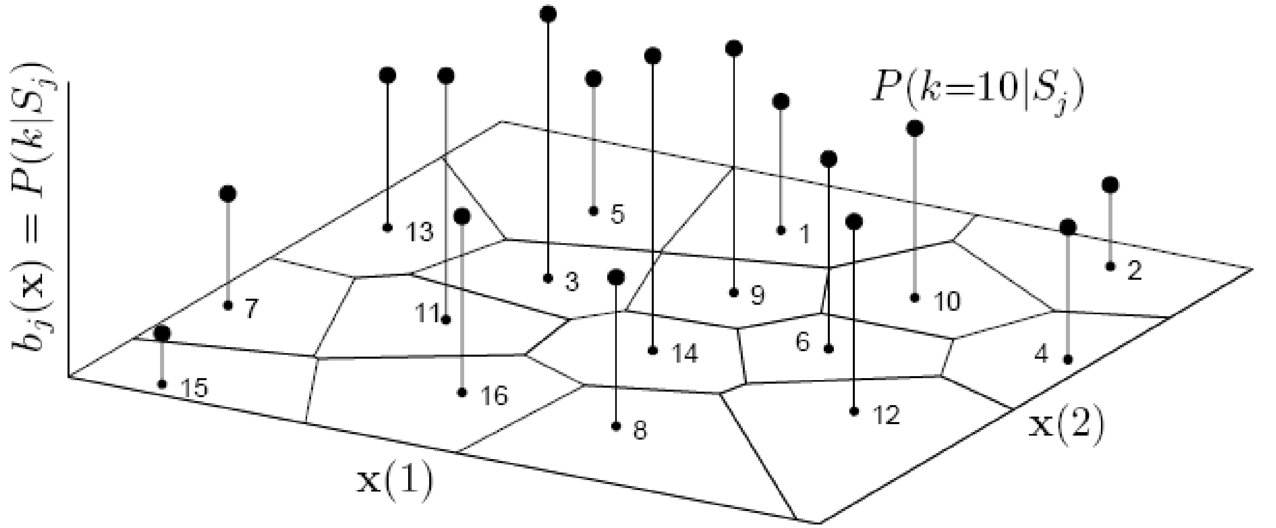
6.4.2 Emissionsmodellierung

Kontinuierliche Beobachtungen

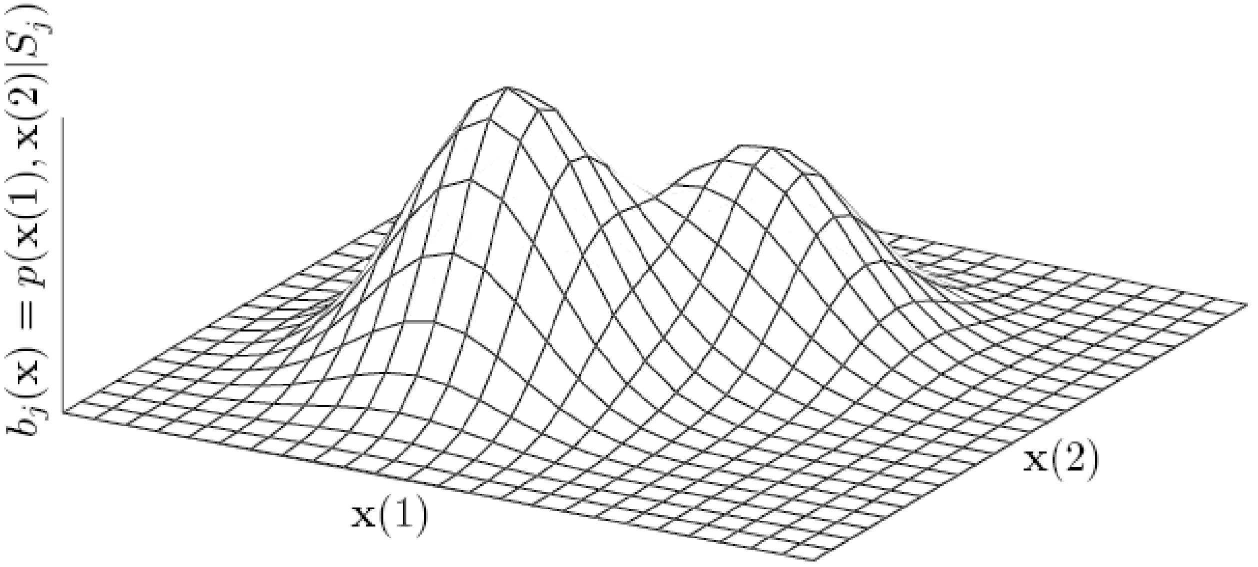
Kontinuierlichen Beobachtungen

- sind die Beobachtungen nicht diskret, dann spricht man von kontinuierlichen Beobachtungen
- die Wahrscheinlichkeitsverteilung solcher Beobachtungen wird als Wahrscheinlichkeitsdichte beschrieben
- HMM mit kontinuierlichen Beobachtungswahrscheinlichkeitsdichten werden als CDHMM (engl. *continuous density HMM*) bezeichnet

Diskrete Beobachtungen (DDHMM)



Kontinuierliche Beobachtungen (CDHMM)



Mathematische Beschreibung

Multivariate Gaussmischverteilung
(Gaussian mixture model – GMM)

$$b_j(\mathbf{x}) = \sum_{k=1}^{M_j} c_{jk} \cdot g_{jk}(\mathbf{x}) = \sum_{k=1}^{M_j} c_{jk} \cdot N(\mathbf{x}, \mu_{jk}, \Sigma_{jk})$$

$$\sum_{k=1}^{M_j} c_{jk} = 1, \qquad c_{jk} \geq 0$$

$$b_j(\mathbf{x}) = \sum_{k=1}^{M} c_{jk} \cdot g_k(\mathbf{x}) = \sum_{k=1}^{M} c_{jk} \cdot N(\mathbf{x}, \mu_k, \Sigma_k)$$

semikontinuierliche HMM

6.4.3 Grundlegende HMM – Probleme

Produktions (Erzeugungs)wahrscheinlichkeit

Gegeben: HMM λ

Beobachtungsfolge: $o = o_1, o_2, \ldots, o_T$

$$1 \leq o_i \leq L$$

Gesucht: $\boxed{P(o \mid \lambda)}$

Wahrscheinlichkeit, dass o von λ generiert wird

Beispiel: Wie groß ist die Wahrscheinlichkeit, dass die Temperaturfolge
10° 20° 20° 30° 30°
gemessen wird.

Berechnung der optimalen Zustandsfolge – Dekodierung

Gegeben: HMM λ

Beobachtungsfolge: $o = o_1, o_2, \ldots, o_T$

$$1 \leq o_i \leq L$$

Gesucht: $s^* = \underset{s}{\operatorname{argmax}}\, P(s \mid o, \lambda)$

$$s = s_1, s_2, \ldots, s_T \qquad 1 \leq s_i \leq N$$

Beispiel: Gegeben sei die Temperaturfolge 10° 20° 20° 30° 30°.
Welches Wetter hat am wahrscheinlichsten geherrscht?

Parameterschätzung – Trainingsproblem

Gegeben: Beobachtungsfolge: $O = O_1, O_2, \ldots, O_T$

$1 \leq o_i \leq L$

Gesucht: HMM λ

das die Beobachtungsfolge am wahrscheinlichsten generiert

6.4.4 Berechnung Erzeugungswahrscheinlichkeit

Problem

Gegeben: λ $\quad o = o_1, o_2, \ldots, o_T \quad 1 \leq o_i \leq L$

$$\Rightarrow \quad P(o \mid \lambda)$$

Einfache, ineffiziente Methode

$$P(o \mid \lambda) = \sum_{s \in S^T} P(o, s \mid \lambda) = \sum_{s \in S^T} P(o \mid s, \lambda) \cdot P(s \mid \lambda)$$

auf Grund der Formel zur Berechnung der bedingten Wahrscheinlichkeit

$$S^T = \underbrace{S \times S \times \cdots \times S}_{T\text{ - mal}} \qquad S = \{1, \ldots, N\}$$

$$s = s_1, s_2, \ldots, s_T \qquad 1 \leq s_i \leq N$$

entspricht einer Zustandsfolge

Einfache, ineffiziente Methode

$$P(o|\lambda) = \sum_{s \in S^T} P(o, s|\lambda) = \sum_{s \in S^T} P(o|s, \lambda) \cdot P(s|\lambda)$$

$$P(s|\lambda) = e_{s_1} \prod_{n=2}^{T} a_{s_{n-1}s_n} \qquad P(o|s, \lambda) = \prod_{n=1}^{T} b_{s_n o_n}$$

$$P(o|\lambda) = \sum_{s \in S^T} e_{s_1} b_{s_1 o_1} \prod_{n=2}^{T} a_{s_{n-1}s_n} b_{s_n o_n}$$

Der Rechenaufwand mit Hilfe dieser Formel ist schon für kleine Werte von N (und T) nicht mehr vertretbar (Komplexität $O(N^T T)$).

Vorwärts – Variable

$$\alpha_n(i) = P(o_1, o_2, \ldots, o_n, S_n = i \mid \lambda) \qquad 1 \leq n \leq T$$

Wahrscheinlichkeit, dass man sich zum Zeitpunkt n im Zustand i befindet und vom HMM die Beobachtungsfolge

$$O_1, O_2, \ldots, O_n$$

generiert wurde

Rückwärts – Variable

$$\beta_n(i) = P(o_{n+1},\ldots,o_T, S_n = i \mid \lambda) \qquad 1 \le n \le T-1$$

Wahrscheinlichkeit, dass man sich zum Zeitpunkt n im Zustand i befindet und vom HMM die Beobachtungsfolge

$$O_{n+1}, O_{n+2}, \ldots, O_T$$

generiert wurde

Wir setzen: $\beta_T(i) = 1$

Vorwärts – Rückwärts Algorithmus

$$\alpha_n(i) = P(o_1, o_2, \ldots, o_n, S_n = i \mid \lambda)$$

$$1 \leq n \leq T$$

$$\beta_n(i) = P(o_{n+1}, \ldots, o_T, S_n = i \mid \lambda)$$

$$\Rightarrow \quad \alpha_n(i) \cdot \beta_n(i) = P(o, S_n = i \mid \lambda) \qquad 1 \leq n \leq T$$

$$\boxed{P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_n(i) \cdot \beta_n(i)} \qquad 1 \leq n \leq T$$

Vorwärts – Rückwärts Algorithmus

$$\boxed{P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_n(i) \cdot \beta_n(i)}$$

$$1 \le n \le T$$

$n = T \Rightarrow P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$

$n = 1 \Rightarrow P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_1(i) \cdot \beta_1(i)$

Vorwärts – Algorithmus

$$\alpha_n(i) = P(o_1, o_2, \ldots, o_n, S_n = i \mid \lambda)$$

$$\alpha_1(i) = e_i \cdot b_{io_1} \qquad 1 \le i \le N$$

$$\alpha_{n+1}(i) = b_{io_{n+1}} \cdot \sum_{j=1}^{N} \alpha_n(j) \cdot a_{ji} \qquad \begin{array}{l} 1 \le i \le N \\ 1 \le n \le T-1 \end{array}$$

$$P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_n(i) \cdot \beta_n(i)$$

$$\xrightarrow{n = T}$$

$$\boxed{P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)}$$

Rückwärts – Algorithmus

$$\beta_n(i) = P(o_{n+1},\ldots,o_T, S_n = i \mid \lambda)$$

$$\beta_T(i) = 1 \qquad 1 \le i \le N$$

$$\beta_n(i) = \sum_{j=1}^{N} a_{ij} \cdot b_{jo_{n+1}} \cdot \beta_{n+1}(j) \qquad 1 \le i \le N, \quad 1 \le n \le T-1$$

$$P(o \mid \lambda) = \sum_{i=1}^{N} \alpha_n(i) \cdot \beta_n(i) \xrightarrow[n=1]{} \boxed{P(o \mid \lambda) = \sum_{i=1}^{N} e_i \cdot b_{io_1} \cdot \beta_1(i)}$$

Bemerkung

Komplexität: $O(N^2T)$

Zur Bestimmung von $P(o \mid \lambda)$

werden entweder die Werte $\alpha_n(i)$

oder die Werte $\beta_n(i)$ benötigt.

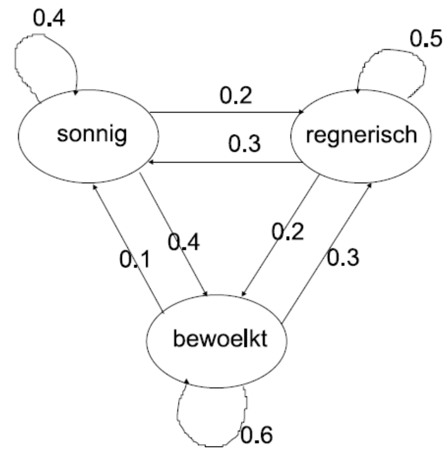
Die beiden Algorithmen unterscheiden sich hauptsächlich
durch die zeitliche Richtung der Rekursion.

Beispiel

Wie groß ist die Wahrscheinlichkeit, dass die Temperaturfolge

10° 20° 20° 30° 30°

gemessen wird.



$$\alpha_1(i) = e_i \cdot b_{io_1} \qquad 1 \le i \le N$$

$$M = \{0°, 10°, 20°, 30°\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\alpha_1(1) = e_1 \cdot b_{12} = 0.3 \cdot 0.1 = 0.03$$

$$\alpha_1(2) = e_2 \cdot b_{22} = 0.4 \cdot 0.2 = 0.08$$

$$\alpha_1(3) = e_3 \cdot b_{32} = 0.3 \cdot 0.4 = 0.12$$

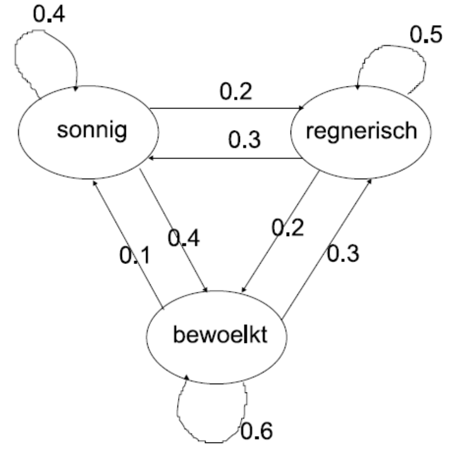
$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

Beispiel

Wie groß ist die Wahrscheinlichkeit, dass die Temperaturfolge

10° 20° 20° 30° 30°

gemessen wird.



$$M = \{0°, 10°, 20°, 30°\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

$$\alpha_1(1) = 0.03 \quad \alpha_1(2) = 0.08 \quad \alpha_1(3) = 0.12$$

$$\alpha_{n+1}(i) = b_{io_{n+1}} \cdot \sum_{j=1}^{N} \alpha_n(j) \cdot a_{ji}$$

$$\alpha_2(1) = b_{1o_2} \cdot \sum_{j=1}^{3} \alpha_1(j) \cdot a_{j1} = b_{13} \cdot \sum_{j=1}^{3} \alpha_1(j) \cdot a_{j1}$$

$$\alpha_2(1) = 0.3 \cdot (0.03 \cdot 0.4 + 0.08 \cdot 0.1 + 0.12 \cdot 0.3)$$

$$\alpha_2(1) = 0.0168$$

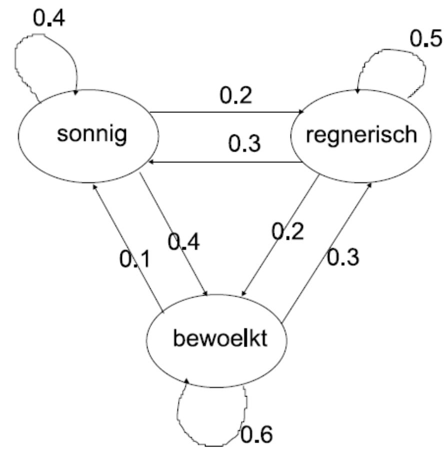
$$\alpha_2(2) = 0.0336 \qquad \alpha_2(3) = 0.0360$$

Beispiel

Wie groß ist die Wahrscheinlichkeit, dass die Temperaturfolge

10° 20° 20° 30° 30°

gemessen wird.



$$\alpha_1(1) = 0.03 \quad \alpha_1(2) = 0.08 \quad \alpha_1(3) = 0.12$$

$$\alpha_2(1) = 0.0168 \quad \alpha_2(2) = 0.0336 \quad \alpha_2(3) = 0.0360$$

$$M = \{0°, 10°, 20°, 30°\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\alpha_{n+1}(i) = b_{io_{n+1}} \cdot \sum_{j=1}^{N} \alpha_n(j) \cdot a_{ji}$$

$$\alpha_3(2) = b_{23} \cdot \sum_{j=1}^{3} \alpha_2(j) \cdot a_{j2}$$

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

$$\alpha_3(2) = 0.4 \cdot (0.0168 \cdot 0.4 + 0.0336 \cdot 0.6 + 0.036 \cdot 0.2)$$

$$\alpha_3(2) = 0.0136$$

Beispiel

Wie groß ist die Wahrscheinlichkeit, dass
die Temperaturfolge

10° 20° 20° 30° 30°

gemessen wird.

$\alpha_n(i)$	$10°(n=1)$	$20°(n=2)$	$20°(n=3)$	$30°(n=4)$	$30°(n=5)$
$sonnig(Z_1)$	0.03	0.0168	0.0063	0.0038	0.0011
$bewoelkt(Z_2)$	0.08	0.0336	0.0136	0.0026	0.0007
$regnerisch(Z_3)$	0.12	0.0360	0.0126	0.0012	0.0002

$$P(o\,|\,\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

$$P(o\,|\,\lambda) = \sum_{i=1}^{3} \alpha_5(i) = 0.0011 + 0.0007 + 0.0002 = 0.002$$

6.4.5 Berechnung der optimalen Zustandsfolge

Dekodierung

Zustandsfolge

- auch wenn die Folge der von einem HMM eingenommenen Zustände unserer Beobachtung nicht zugänglich ist, lässt die Kenntnis der Ausgabe und der HMMparameter dennoch einige statistische Rückschlüsse zu
- Viterbi – Algorithmus

Problem

Gegeben: Beobachtungsfolge: $o = o_1, o_2, \ldots, o_T \quad 1 \leq o_i \leq L$

HMM λ

wir maximieren die Wahrscheinlichkeit: $P(s \mid o, \lambda) = \dfrac{P(s, o \mid \lambda)}{P(o \mid \lambda)}$

über alle $s = s_1, s_2, \ldots, s_T \quad 1 \leq s_i \leq N$

$$P^*(o \mid \lambda) = \max_{s \in S^T} P(s \mid o, \lambda) = \max_{s \in S^T} \frac{P(s, o \mid \lambda)}{P(o \mid \lambda)} \Rightarrow \max_{s \in S^T} P(s, o \mid \lambda)$$

$$\boxed{P^*(o \mid \lambda) = \max_{s \in S^T} P(s \mid o, \lambda) = P(s^*, o \mid \lambda)}$$

Viterbi – Algorithmus

wir betrachten jetzt die maximal erzielbaren Wahrscheinlichkeiten:

$$\alpha'_n(i) = \delta_n(i) = \max_{s \in S^T, s_n = i} P(o_1, o_2, \ldots, o_n, s_1, \ldots, s_n \mid \lambda)$$

Die Berechnung geschieht wieder rekursiv. Die Zustände der optimalen Folge werden über die Funktion

$$\psi_n(i) \qquad\qquad 1 \le i \le N$$

berechnet

Initialisierung und Rekursion

$$\delta_1(i) = \alpha_1(i) = e_i \cdot b_{io_1} \qquad 1 \leq i \leq N$$

$$\psi_1(i) = 0$$

$$\delta_{n+1}(i) = b_{io_{n+1}} \cdot \max_{j=1,\ldots,N} \left\{ \delta_n(j) \cdot a_{ji} \right\} \qquad 1 \leq i \leq N$$

$$\psi_{n+1}(i) = \operatorname*{argmax}_{j=1,\ldots,N} \left\{ \delta_n(j) \cdot a_{ji} \right\} \qquad 1 \leq n \leq T-1$$

Terminierung und Rückverfolgung

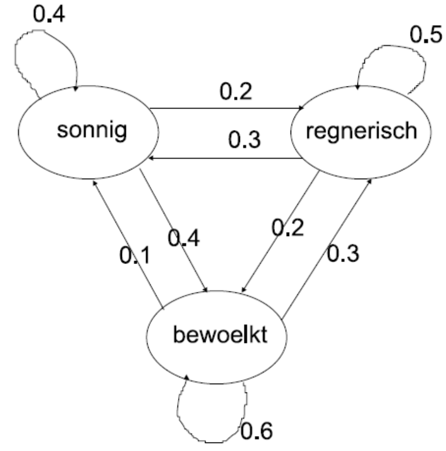
$$P^*(o \mid \lambda) = \max_{j=1,\ldots,N} \delta_T(j)$$

$$s_T^* = \operatorname*{argmax}_{j=1,\ldots,N} \delta_T(j)$$

$$s_n^* = \psi_{n+1}(s_{n+1}^*) \qquad 1 \leq n \leq T-1$$

Beispiel

Gegeben sei die Temperaturfolge
10° 20° 20° 30° 30°.
Welches Wetter hat am wahrscheinlichsten geherrscht?



$$\delta_1(1) = 0.03 \quad \delta_1(2) = 0.08 \quad \delta_1(3) = 0.12$$

$$\delta_{n+1}(i) = b_{io_{n+1}} \cdot \max_{j=1,\ldots,N} \left\{ \delta_n(j) \cdot a_{ji} \right\}$$

$$M = \{0°, 10°, 20°, 30°\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\delta_2(1) = b_{1o_2} \cdot \max_{j=1,\ldots,N} \left\{ \delta_1(j) \cdot a_{j1} \right\}$$

$$\delta_2(1) = 0.3 \cdot \max\{0.03 \cdot 0.4, 0.08 \cdot 0.1, 0.12 \cdot 0.3\}$$

$$\delta_2(1) = 0.3 \cdot 0.12 \cdot 0.3 = 0.0108$$

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

Beispiel

Gegeben sei die Temperaturfolge
10° 20° 20° 30° 30°.
Welches Wetter hat am wahrscheinlichsten geherrscht?

$\alpha'_n(i)$	$10°(n=1)$	$20°(n=2)$	$20°(n=3)$	$30°(n=4)$	$30°(n=5)$
$sonnig(Z_1)$	0.03	0.0108	0.00216	$\boxed{0.00072}$	$\boxed{0.00014}$
$bewoelkt(Z_2)$	0.08	0.0192	0.00461	0.00055	0.00007
$regnerisch(Z_3)$	$\boxed{0.12}$	$\boxed{0.0240}$	$\boxed{0.0048}$	0.00024	0.00002

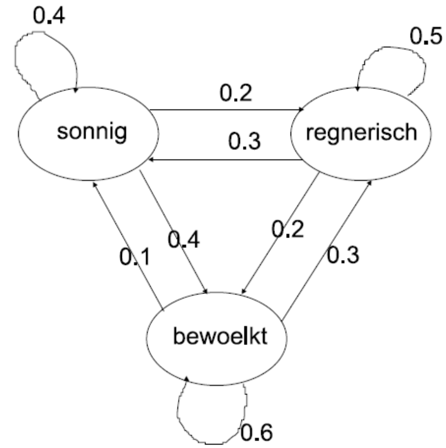
$$P^*(o \mid \lambda) = \max_{j=1,\ldots,N} \delta_5(j) = 0.00014$$

Beispiel

Gegeben sei die Temperaturfolge

10° 20° 20° 30° 30°.

Welches Wetter hat am wahrscheinlichsten geherrscht?



$$M = \{0°, 10°, 20°, 30°\}$$

$$\mathbf{e} = (0.3, 0.4, 0.3)$$

$$\delta_1(1) = 0.03 \qquad \delta_1(2) = 0.08 \qquad \delta_1(3) = 0.12$$

$$\psi_{n+1}(i) = \underset{j=1,\ldots,N}{\operatorname{argmax}} \left\{ \delta_n(j) \cdot a_{ji} \right\}$$

$$\psi_2(1) = \underset{j=1,\ldots,N}{\operatorname{argmax}} \left\{ \delta_1(j) \cdot a_{j1} \right\}$$

$$\psi_2(1) = \operatorname{argmax}\{0.03 \cdot 0.4, 0.08 \cdot 0.1, 0.12 \cdot 0.3\}$$

$$\psi_2(1) = 3$$

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0.1 & 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \end{pmatrix}$$

Beispiel

Gegeben sei die Temperaturfolge
10° 20° 20° 30° 30°.
Welches Wetter hat am wahrscheinlichsten geherrscht?

$\psi_n(i)$	$10°(n=1)$	$20°(n=2)$	$20°(n=3)$	$30°(n=4)$	$30°(n=5)$
$sonnig(Z_1)$	0	3	3	3	1
$bewoelkt(Z_2)$	0	2	2	2	2
$regnerisch(Z_3)$	0	3	3	3	2

$$s_T^* = \underset{j=1,\ldots,N}{\operatorname{argmax}} \delta_T(j) \qquad s_5^* = \underset{j=1,\ldots,N}{\operatorname{argmax}} \delta_5(j) = 1$$

Beispiel

Gegeben sei die Temperaturfolge

10° 20° 20° 30° 30°.

Welches Wetter hat am wahrscheinlichsten geherrscht?

$\psi_n(i)$	$10°(n=1)$	$20°(n=2)$	$20°(n=3)$	$30°(n=4)$	$30°(n=5)$
$sonnig(Z_1)$	0	3	3	3	1
$bewoelkt(Z_2)$	0	2	2	2	2
$regnerisch(Z_3)$	0	3	3	3	2

$$s_n^* = \psi_{n+1}(s_{n+1}^*)$$

$$s_4^* = \psi_5(s_5^*) = \psi_5(1) = 1$$

$$s_3^* = \psi_4(s_4^*) = \psi_4(1) = 3$$

$$s_2^* = \psi_3(s_3^*) = \psi_3(3) = 3$$

$$s_1^* = \psi_2(s_2^*) = \psi_2(3) = 3$$