

Tutorium 9 - Wahrscheinlichkeiten, POS-Tagging & Hidden Markov Models

Übungen

Andreas Wassermayr

15.01.2021

Übung 1 - Wahrscheinlichkeitstheorie

- 1 Wodurch zeichnet sich eine diskrete Wahrscheinlichkeitsverteilung aus?
- 2 Wann sind zwei Ereignisse X und Y voneinander statistisch unabhängig?
- 3 Was besagt das Theorem von Bayes und wozu kann man es verwenden?

Übung 1 - Musterlösung

- 1 Wodurch zeichnet sich eine diskrete Wahrscheinlichkeitsverteilung aus?
Es gibt endlich (oder abzählbar unendlich) viele mögliche Versuchsergebnisse & die Wahrscheinlichkeiten addieren zu 1 auf
- 2 Wann sind zwei Ereignisse X und Y voneinander statistisch unabhängig?
*Wenn gilt $p(X \wedge Y) = p(X) * p(Y)$*
- 3 Was besagt das Theorem von Bayes und wozu kann man es verwenden?
$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)}$$

um bedingte Wahrscheinlichkeiten zu Berechnen mit Hilfe von anderen Werten, die man leichter schätzen kann

Übung 2 - Theorem von Bayes

Verwendet das Theorem von Bayes, um die fehlenden Werte zu berechnen:

$P(X)$	0.90
$P(Y)$	0.05
$P(X Y)$	0.50
$P(Y X)$	???

$P(X)$	0.25
$P(Y)$	0.75
$P(X Y)$	???
$P(Y X)$	0.15

$P(X)$	0.20
$P(Y)$	???
$P(X Y)$	0.55
$P(Y X)$	0.90

Übung 2 - Musterlösung

$P(X)$	0.90
$P(Y)$	0.05
$P(X Y)$	0.50
$P(Y X)$	0.03

$P(X)$	0.25
$P(Y)$	0.75
$P(X Y)$	0.05
$P(Y X)$	0.15

$P(X)$	0.20
$P(Y)$	0.33
$P(X Y)$	0.55
$P(Y X)$	0.90

Übung 3 - Tagfolgen

Gegeben dem Satz "I can can a can" und dem folgenden HMM, berechnet die Wahrscheinlichkeiten für folgende Tagfolgen: (Leere Zellen stehen für Wahrscheinlichkeit 0)

A-priori-Wahrscheinlichkeiten

Wort	AUX	V	PRO	DET	NN
the	-	-	-	0.50	-
a	-	-	-	0.50	-
I	-	-	0.60	-	-
you	-	-	0.40	-	-
read	-	0.50	-	-	-
can	0.33	0.50	-	-	1.00
must	0.33	-	-	-	-
should	0.33	-	-	-	-

Übergangswahrscheinlichkeiten

Vorgänger↓	PRO	PN	NN	DET	AUX	V	Ende
Beginn	0.25	0.25	0.05	0.45	-	-	-
PRO	0.10	-	-	0.15	0.25	0.50	-
PN	-	0.10	-	0.15	0.25	0.50	-
NN	0.10	0.05	0.10	-	0.25	0.45	0.05
DET	0.05	0.05	0.90	-	-	-	-
AUX	0.10	0.05	0.10	-	-	0.75	-
V	0.20	0.10	0.05	0.60	-	0.05	-

- 1 PRO - AUX - V - DET - NN
- 2 PRO - NN - NN - DET - NN
- 3 PRO - V - V - DET - NN

Übung 3 - Musterlösung

- 1 PRO - AUX - V - DET - NN
0.000062648
- 2 PRO - NN - NN - DET - NN
0
- 3 PRO - V - V - DET - NN
0.000012656

Übung 4 - Maximum Likelihood

Gegeben folgender Sätze als Trainingsdaten, wie würde das HMM aussehen?

- *Das_{DET} Kin_{NN} hat_V heute_{ADV} geöffnet_V*
- *Ich_{PRO} kann_{AUX} heute_{ADV} mit_{PP} dir_{PRO} ins_{PP} Kin_{NN} gehen_V*
- *Im_{PP} Saal_{NN} wird_V es_{PRO} plötzlich_{ADV} dunkel_{ADJ}*

Übung 4 - Musterlösung

A-priori-Wahrscheinlichkeiten								
Wort	AUX	V	PRO	DET	NN	ADJ	ADV	PP
kann	1.00	-	-	-	-	-	-	-
hat	-	0.25	-	-	-	-	-	-
gehen	-	0.25	-	-	-	-	-	-
geöffnet	-	0.25	-	-	-	-	-	-
wird	-	0.25	-	-	-	-	-	-
dir	-	-	0.33	-	-	-	-	-
es	-	-	0.33	-	-	-	-	-
Ich	-	-	0.33	-	-	-	-	-
Das	-	-	-	1.00	-	-	-	-
Kino	-	-	-	-	0.67	-	-	-
Saal	-	-	-	-	0.33	-	-	-
dunkel	-	-	-	-	-	1.00	-	-
heute	-	-	-	-	-	-	0.67	-
plötzlich	-	-	-	-	-	-	0.33	-
Im	-	-	-	-	-	-	-	0.33
ins	-	-	-	-	-	-	-	0.33
mit	-	-	-	-	-	-	-	0.33

Übung 4 - Musterlösung

Übergangswahrscheinlichkeiten

Vorgänger↓	AUX	V	PRO	DET	NN	ADJ	ADV	PP	Ende
Beginn	-	-	0.33	0.33	-	-	-	0.33	-
AUX	-	-	-	-	-	-	1.00	-	-
V	-	-	0.25	-	-	-	0.25	-	0.50
PRO	0.33	-	-	-	-	-	0.33	0.33	-
DET	-	-	-	-	1.00	-	-	-	-
NN	-	1.00	-	-	-	-	-	-	-
ADJ	-	-	-	-	-	-	-	-	1.00
ADV	-	0.33	-	-	-	0.33	-	0.33	-
PP	-	-	0.33	-	0.67	-	-	-	-

Übung 5 - Viterbi-Algorithmus

Verwendet den Viterbi-Algorithmus und das folgende Hidden Markov Model, um die wahrscheinlichste Tagfolge für den Satz **Violinist involved in crash blossoms** zu bestimmen!

A-priori-Wahrscheinlichkeiten

Wort	ADJ	ADV	NN	PP	V
blossoms	-	-	0.15	-	0.25
crash	0.40	-	0.50	-	0.40
in	0.50	1.00	0.05	1.00	0.10
involved	0.10	-	-	-	0.25
violinist	-	-	0.30	-	-

Übergangswahrscheinlichkeiten

Vorgänger↓	ADJ	ADV	NN	PP	V	Ende
Beginn	0.20	0.15	0.50	0.10	0.05	-
ADJ	0.33	0.05	0.60	-	-	0.02
ADV	0.33	0.33	0.33	-	-	-
NN	-	-	0.05	0.05	0.40	0.50
PP	0.33	-	0.67	-	-	-
V	0.20	-	0.20	0.10	0.05	0.25

Übung 5 - Musterlösung

< s >	Violinist	involved	in	crash	blossoms	< e >
Beginn	NN	ADJ V	PP ADJ ADV V NN	NN ADJ V	NN V	Ende

$$p(\text{NN}, \text{V}, \text{PP}, \text{NN}, \text{V} | \text{Violinist}, \text{involved}, \text{in}, \text{crash}, \text{blossoms}) = 0.000012563$$