Schriftliche Wiederholungsprüfung zur Vorlesung
Statistische Methoden in der maschinellen Sprachverarbeitung
WS 2020/21
Dozent: Helmut Schmid

Sie haben **90 Minuten** Zeit plus 5 Minuten zum Absenden Ihrer Lösungen per Email. Sie können eine Textdatei oder einen Scan schicken.

Wenn Sie einen Fehler in einer der Aufgaben entdecken sollten, dann melden Sie sich bitte per Zoom (aber nicht Zoom-Chat). Notfalls können Sie mich auch unter der Nummer 07121 44240 anrufen.

Aufgabe 1) Die folgenden Formeln definieren jeweils die logarithmierte Wahrscheinlichkeit für ein bestimmtes statistisches Modell, das wir in der Vorlesung behandelt haben.

a)
$$\log p(c, d_1, ..., d_L) = \log p(c) + \sum_{k=1}^{L} \log p(d_k|c)$$

b)
$$\log p(d_1, d_2, ..., d_L) = \sum_{k=1}^{L+1} \log p(d_k|d_{k-C}, ..., d_{k-1})$$

c)
$$\log p(c_1, d_1, ..., c_L, d_L) = \sum_{k=1}^{L+1} \log p(c_k|c_{k-C}, ..., c_{k-1}) + \sum_{k=1}^{L+1} \log p(d_k|c_k)$$

d)
$$\log p(c|\mathbf{d}) = -\log z(\mathbf{d}) + \sum_{k} w_k m_k(\mathbf{d}, c)$$

wobei w_k und $m_k(...)$ Skalare sind und außerdem gilt: $\mathbf{d} = d_1, ..., d_L$.

Bearbeiten Sie für jede der Formeln a) bis d) die folgenden Teilaufgaben:

I) Wie heißt das entsprechende **statistische Modell**?

II) Nennen Sie eine konkrete computerlinguistische **Anwendung** für dieses Modell. Welche Bedeutung hat jede einzelne **Variable** (inklusive der Variable k) bei dieser Anwendung? (Eine Anwendung genügt hier.)

III) Wählen Sie für die in II) gewählte Anwendung ein **Beispiel** für die Argumente "..." der log. Wahrscheinlichkeitsverteilung log p(...) auf der linken Seite der Formel und wenden Sie die Formel auf das Beispiel an: Schreiben Sie also hin, wie gemäß der Formel die logarithmierte Wahrscheinlichkeit für Ihr Beispiel zu **berechnen** ist.

Denken Sie sich ein **neues** Beispiel aus dem Themen-Bereich **Wirtschaft** aus, bei dem $L = 2$ ist. Nehmen Sie kein Beispiel, das im Kurs behandelt wurde.

Beispiel: (allerdings kein statistisches Modell)
Formel: $p(a|b) = p(b|a)p(a)/p(b)$
I) Name: Bayes'sches Theorem
II) Anwendung: Transformation der lexikalischen Wahrscheinlichkeiten beim HMM-Tagger zur besseren Behandlung unbekannter Wörter; a ist bei dieser Anwendung ein Wort und b ein Tag.
III) $p(Aktie|N) = p(N|Aktie)p(Aktie)/p(N)$

(10 Punkte)

Aufgabe 2) Ein HMM ist gegeben durch die (unvollständige) Tabelle:

	PRO	MD	N	$\langle s \rangle$	we	can	ϵ
PRO	0.1	0.3	0.1	0.1	0.2	0	0
MD	0.1	0.0	0.1	0.1	0	0.3	0
N	0.1	0.2	0.2	0.2	0	0.1	0
$\langle s \rangle$	0.2	0.1	0.1	0	0	0	1

mit $p(\langle s \rangle|PRO) = 0.1$ und $p(I|PRO) = 0.2$.

Berechnen Sie für die Tokenfolge *"we can"* und das obige HMM die **Viterbi**-Wahrscheinlichkeiten $\delta_t(i)$ und die besten Vorgänger-Tags $\psi_t(i)$ nach den Formeln:

$$\begin{aligned}
\delta_t(0) &= \begin{cases} 1 & \text{falls } t = \langle s \rangle \\ 0 & \text{sonst} \end{cases} \\
\delta_t(k) &= \max_{t'} \; \delta_{t'}(k-1)\; p(t|t')\; p(w_k|t) \quad \text{für } 0 < k \leq n+1 \\
\psi_t(k) &= \arg\max_{t'} \; \delta_{t'}(k-1)\; p(t|t')\; p(w_k|t) \quad \text{für } 0 < k \leq n+1
\end{aligned}$$

Schreiben Sie nicht nur das Ergebnis hin, sondern zeigen Sie auch den Rechenweg.
Extrahieren Sie dann die beste **Tagfolge** nach den Formeln

$$\begin{aligned}
t_n &= \psi_{\langle s \rangle}(n+1) \\
t_k &= \psi_{t_{k+1}}(k+1) \quad \text{für } n > k > 0
\end{aligned}$$

(5 Punkte)

Aufgabe 3) Sie sollen untersuchen, ob TaggerA **signifikant** besser ist als der TaggerB.
Auf Testdaten haben Sie folgende Ergebnisse erhalten:

Satz:	Ich	sah	den	Mann	auf	dem	Hügel	mit	dem	Stock	unter	dem	Arm
Tags:	PPER	VVFIN	ART	NN	APPR	ART	NN	APPR	ART	NN	APPR	ART	NN
TaggerA:	PPER	VVINF	ART	NNS	APPR	ART	NE	APPR	PDS	NN	APPO	PDS	NN
TaggerB:	NN	VVFIN	ART	NE	APPO	PDS	NN	APPR	PDS	NE	APPR	ART	NE

Führen Sie einen **Vorzeichen-Test** durch. Geben Sie dabei alle Zwischenschritte an.
Wie lautet die Nullhypothese?
(Werte der Binomialfunktion müssen Sie hier nicht ausrechnen.) (5 Punkte)

Aufgabe 4) Wie wird die Wahrscheinlichkeit $p(s|H, a, u)$ bei **interpolierter Backoff-Glättung** berechnet, wenn die Backoff-Faktoren $\alpha(a_1, ..., a_i)$ und die relativen Häufigkeiten mit Discount $r(a_{i+1}|a_1, ..., a_i) = (f(a_1, ..., a_{i+1}) - \delta_i)/f(a_1, ..., a_i)$ für $i \leq 3$ gegeben sind? Die Backoff-Glättung soll bei der Unigramm-Wahrscheinlichkeit $p(a_{i+1})$ abbrechen.
(Das bedeutet, dass dort der Discount 0 ist.) (5 Punkte)

Aufgabe 5) Gegeben sind die Buchstabenpaar-Häufigkeiten

x,y	a,a	a,b	a,c	b,a	b,b	b,c	c,a	c,b	c,c
f(x,y)	1	3	0	1	1	5	2	0	1

Berechnen Sie die **ungeglätteten Wahrscheinlichkeiten** $p(y|x)$.

Berechnen Sie dann Unigramm-Häufigkeiten $p(y)$ nach der normalen Methode und nach der **Kneser-Ney**-Methode.
(Diese können für die Schätzung von Backoff-Wahrscheinlichkeiten benutzt werden.)

(5 Punkte)

(30 Punkte insgesamt)

Viel Erfolg!