Maschinelles Übersetzen natürlicher Sprachen

Aufgabe 1

1. Die Wahrscheinlichkeit einer Wortfolge $w_1 \cdots w_n$ kann auf die folgenden beiden Weisen dekomponiert werden (beachten Sie, dass wir in der hier verwendeten Kurzschreibweise die benutzten Zufallsvariablen unterschlagen):

$$\begin{aligned}
&P(w_1 \cdots w_n) \\
&= P(w_1) \cdot P(w_2 \mid w_1) \cdot P(w_3 \mid w_1, w_2) \\
&\quad \cdots P(w_n \mid w_1, \ldots, w_{n-1}) \cdot P(n \mid w_1, \ldots, w_n) \\
&= P(w_n) \cdot P(w_{n-1} \mid w_n) \cdot P(w_{n-2} \mid w_{n-1}, w_n) \\
&\quad \cdots P(w_1 \mid w_2, \ldots, w_n) \cdot P(n \mid w_1, \ldots, w_n)\,.
\end{aligned}$$

   Heißt das also, dass $P(w_1 \cdots w_n) = P(w_n \cdots w_1)$ gilt?

2. Sei b eine Bigramm-Wahrscheinlichkeitsverteilung über einem Alphabet X mit dem Stoppsymbol $\#$ (natürlich nehmen wir an, dass $\# \notin X$) und sei $w_1 \cdots w_n \in X^*$.

   Ist die Gleichung

$$b(w_1 \mid \#) \cdot \Big(\prod_{i \in \{2,\ldots,n\}} b(w_i \mid w_{i-1})\Big) \cdot b(\# \mid w_n) = b(w_n \mid \#) \cdot \Big(\prod_{i \in \{n-1,\ldots,1\}} b(w_i \mid w_{i+1})\Big) \cdot b(\# \mid w_1)$$

   allgemeingültig?

Aufgabe 2

1. Erweitern Sie das Bigramm-Modell aus der Vorlesung zu einem Trigramm-Modell, also einem Sprachmodell mit der Annahme, dass die Wahrscheinlichkeit eines jeden Worts von den *zwei* Worten vor ihm abhängt.

2. Definieren Sie ein allgemeines n-gramm-Modell!

Aufgabe 3

Gegeben seien m, $n \in \mathbb{N}$ und eine Abbildung $f : [m] \times [n] \to \mathbb{R}$, dabei sei $[n] = \{1, \ldots, n\}$ für jedes $n \in \mathbb{N}$. Zeigen Sie die Gültigkeit der Gleichung

$$\sum_{a \in [n]^{[m]}} \prod_{i \in [m]} f(i, a(i)) = \prod_{i \in [m]} \sum_{j \in [n]} f(i, j)$$

und setzen Sie sie in Bezug mit dem IBM-Modell 1.

Zur Erinnerung: A^B bezeichnet die Menge aller totalen Funktionen von der Menge B in die Menge A.

Aufgabe 4

Gegeben sei das folgende Wörterbuch t.

			f		
	$t(f \mid e)$	kra	ban	las	gha
	du	0.2	0.4	0.4	0
e	su	0	0.1	0.8	0.1
	ur	0.3	0.4	0.3	0
	fur	0.4	0.3	0.1	0.2

1. Zusätzlich dazu betrachten wir das Längenmodell ε, wobei

$$\varepsilon(m \mid l) = \begin{cases} 0.5 & \text{falls } m = l, \\ 0.25 & \text{falls } |m - l| = 1, \\ 0 & \text{sonst} \end{cases}$$

   für alle $n, m \in \mathbb{N}^+$.

   Bestimmen Sie, für alle $i \in \{1, 2, 3\}$, mit dem IBM-Modell 1 die Übersetzungswahrscheinlichkeiten $P(f_i \mid e_i)$. Dabei seien

   - $f_1 =$ kra las gha, $e_1 =$ du su ur,
   - $f_2 =$ kra las gha, $e_2 =$ du su
   - $f_3 =$ gha gha, $e_3 =$ su du su.

2. Nun betrachten wir ein anderes Längenmodell ε'. Für dieses gelte

$$\varepsilon'(m \mid l) = 2^{-m},$$

   wieder für alle $n, m \in \mathbb{N}^+$.

   Bestimmen Sie $\arg\max_f P(f \mid e_i)$, $i \in \{1, 2, 3\}$, für

   - $e_1 =$ du,
   - $e_2 =$ du su ur,
   - $e_3 =$ fur du fur su fur.

   Wie verändern sich die besten Übersetzungen, wenn wir stattdessen das Längenmodell ε'' mit

$$\varepsilon''(m \mid l) = \begin{cases} 1 & \text{falls } l = m \\ 0 & \text{sonst} \end{cases}$$

   für alle $n, m \in \mathbb{N}^+$, verwenden?