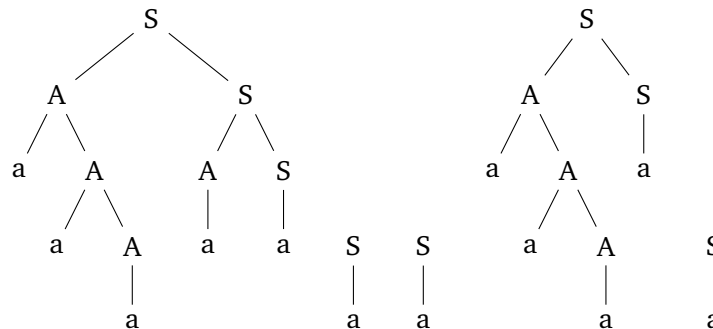# Maschinelles Übersetzen natürlicher Sprachen
10. Übungsblatt
2017-01-19

Aufgabe 1

Consider the context-free grammar $G = (Z, \Sigma, S, P)$ with nonterminal symbols $Z = \{S, A\}$, terminals $\Sigma = \{a, b\}$, start symbol S, and a set of productions $P = \{z \rightarrow x \mid z \in Z, x \in X\}$ where $X = \{AS, a, aA, b\}$. We observe the following sequence of parse trees:

a) Specify an $X \times Z$-corpus c that reflects this observation.

Compute $c_S(AS) = \qquad$, $c_S(a) = \qquad$, $c_A(aA) = \qquad$, $c_A(a) = \qquad$.

b) Let $p \in \mathcal{M}(X|Z)$ be such that

$$p(AS \mid S) = 0.3 \qquad p(aA \mid A) = 0.5$$
$$p(a \mid S) = 0.7 \qquad p(a \mid A) = 0.5.$$

Compute $p(c)$:

$p(c) =$

c) Let $p \in \mathcal{M}(X \mid Z)$. Then

$\bar{p}(c)(S) = \qquad\qquad\qquad\qquad |c_S| =$

$\widetilde{c}_S(AS) = \qquad\qquad\qquad\qquad \widetilde{c}_S(a) =$

$\bar{p}(c)(A) = \qquad\qquad\qquad\qquad |c_A| =$

$\widetilde{c}_A(aA) = \qquad\qquad\qquad\qquad \widetilde{c}_A(a) =$.

d) Now assume, that we observe the following sequence of parse trees.

$$
\begin{array}{cc}
\text{S} & \text{S} \\
| & | \\
\text{a} & \text{a}
\end{array}
$$
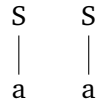
Specify an $X \times Z$-corpus c' that reflects this observation and repeat tasks b) and c) with c'.

e) Let $\Omega = \{(u, v) \mid u, v \in [0, 1]\}$ and $p : \Omega \to \mathcal{M}(X \mid Z)$ such that

$$
\begin{aligned}
\big(p(u, v)\big)(AS \mid S) &= u & \big(p(u, v)\big)(aA \mid A) &= v \\
\big(p(u, v)\big)(\ a \mid S) &= 1 - u & \big(p(u, v)\big)(\ a \mid A) &= 1 - v.
\end{aligned}
$$

Instantiate the optimization problem $\mathrm{cmle}_p(c)$ for p and c.

$$\mathrm{cmle}_p(c) =$$

Aufgabe 2

Imagine the following game: Two independent coins are thrown and you win, if both coins land on the same side. You are only told, if you won. Assume that one coin is rather thick and may land on the edge. We represent the possible events with the set $Y = \{M_1, M_2, M_3\} \times \{N_1, N_2\}$. You win with the events (M_1, N_1) and (M_2, N_2).

After playing the game several times, you won 6 times and lost 18 times. Instantiate the corpus-based EM algorithm with this scenario and calculate one EM step. Start with the probability of $2/5$ for M_1 and M_2, and $1/3$ for N_1.