

**Schriftliche Prüfung zur
Vorlesung Statistische Sprachverarbeitung
WS 2015/16
Dozent: Helmut Schmid**

Aufgabe 1) Geben Sie an, wie bei einem Markowmodell vierter Ordnung (= 5Gramm-Modell) die Wahrscheinlichkeit einer Buchstabenfolge a_1, \dots, a_n definiert ist. (3 Punkte)

Aufgabe 2) Beschreiben Sie, wie ein Sprachidentifizierer mit Hilfe von solchen Markowmodellen implementiert werden kann. (3 Punkte)

Aufgabe 3) Wie lautet die Formel für die Berechnung der Wahrscheinlichkeit $p(a_5|a_1, \dots, a_4)$, wenn Sie eine Maximum-Likelihood-Schätzung (= relative Häufigkeiten) verwenden? Welches Problem ergibt sich bei dieser Art der Parameterschätzung? (3 Punkte)

Aufgabe 4) Was ist die Grundidee der Backoff-Glättung? Geben Sie die zugehörige Formel zur Berechnung der geglätteten Wahrscheinlichkeiten mit einem interpolierten Backoff-Modell an. (3 Punkte)

Aufgabe 5) Erklären Sie, wie der Perzeptron-Algorithmus funktioniert und geben Sie die Formel für die Gewichts Anpassung an. (3 Punkte)

Aufgabe 6) Viele Lernverfahren weisen Metaparameter auf, welche das Lernverfahren selbst nicht lernen kann. Beispielsweise ist bei Hidden-Markow-Modellen die Kontextgröße (Bigramm, Trigramm etc.) ein solcher Metaparameter. Wie können gute Werte für solche Metaparameter bestimmt werden? (3 Punkte)

Aufgabe 7) Erklären Sie ausführlich die Grundidee des Berkeley-Parsers von Petrov und Klein und wie er trainiert wird (ohne Formeln). (4 Punkte)

Aufgabe 8) Mit Conditional Random Fields (CRFs) kann man unter anderem Wortart-Tagger implementieren. Wie ist bei einem CRF die Wahrscheinlichkeit einer Tagfolge t_1, \dots, t_n für eine gegebene Wortfolge w_1, \dots, w_n genau definiert (mit Formel)?

Warum wird eine bedingte Wahrscheinlichkeit definiert? Welches Problem tritt auf, wenn stattdessen eine gemeinsame Wahrscheinlichkeit von Tagfolge und Wortfolge wie bei den HMMs definiert wird? (3 Punkte)

Aufgabe 9) Erklären Sie die Grundidee des Gradientenanstiegsverfahrens. Welche Zielfunktion optimiert man, wenn man einen CRF-basierten Tagger trainiert? Wie ist der Gradient hier definiert? Wie wird er beim Training eines Wortart-Taggers berechnet (welcher Algorithmus wird verwendet)? Wie vermeidet man bei CRFs Overfitting? In dieser Aufgabe sind keine Formeln verlangt. (5 Punkte)

(30 Punkte insgesamt)