

Schriftliche Wiederholungsprüfung zur Vorlesung
Statistische Sprachverarbeitung
SS 2016
Helmut Schmid

Aufgabe 1) Geben Sie an, wie bei einem Markowmodell für Buchstabenfolgen die Wahrscheinlichkeit einer Buchstabenfolge a_1^n definiert ist. (2 Punkte)

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

Aufgabe 3) Erklären Sie, zu welchem Problem die Parameterschätzung mit relativen Häufigkeiten in der Sprachverarbeitung führen kann und wie das Problem prinzipiell gelöst wird. (2 Punkte)

Aufgabe 4) Erklären Sie das Prinzip des EM-Algorithmus am Beispiel des unüberwachten Trainings eines HMM-Taggers. (3 Punkte)

Aufgabe 5) Der Forward-Backward-Algorithmus kann zum Training von Hidden-Markow-Modellen auf unannotierten Daten benutzt werden. Welche Daten brauchen Sie, um einen Wortart-Tagger auf diese Weise zu trainieren? Wie lauten die Formeln zur Berechnung der Forward- und Backward-Wahrscheinlichkeiten? Wozu werden die berechneten Forward- und Backward-Wahrscheinlichkeiten anschließend verwendet? (4 Punkte)

Aufgabe 6) Erklären Sie, wie der Perzeptron-Algorithmus funktioniert und geben Sie die Formel für die Gewichts Anpassung an. (4 Punkte)

Aufgabe 7) Wie funktioniert Parse-Reranking? Welchen Vorteil hat das Reranking? (3 Punkte)

Aufgabe 8) Viele Lernverfahren weisen Metaparameter auf, welche das Lernverfahren selbst nicht lernen kann. Beispielsweise ist bei einem N-Gramm-Modell die Kontextgröße (Bigramm, Trigramm etc.) ein solcher Metaparameter. Wie werden gute Werte für Metaparameter bestimmt? (3 Punkte)

Aufgabe 9) Erklären Sie die Grundidee des Berkeley-Parsers von Petrov und Klein. Wie wird der Parser trainiert? (3 Punkte)

Aufgabe 10) Wie wird üblicherweise die Genauigkeit von (Konstituenten-)Parsern gemessen? Erklären Sie die Methode und wie das Maß berechnet wird. (3 Punkte)

(30 Punkte insgesamt)

Viel Erfolg!