

**Schriftliche Prüfung zur
Vorlesung Statistische Sprachverarbeitung
WS 2016/17
Dozent: Helmut Schmid**

Aufgabe 1) Was sagt das **Zipf'sche Gesetz** aus? (2 Punkte)

Aufgabe 2) Was ist eine **Zufallsvariable**? Was haben Zufallsvariablen mit der Berechnung eines Notendurchschnittes zu tun? (Denken Sie an die Noten sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend.) (2 Punkte)

Aufgabe 3) Wie wird der **Erwartungswert** für die Zahl der Augen eines fairen Würfels berechnet? (2 Punkte)

Aufgabe 4) Wie lautet das **Theorem von Bayes**? Nennen Sie ein Beispiel, wo wir dieses Theorem benutzt haben. (2 Punkte)

Aufgabe 5) Geben Sie die Formeln an, mit denen Sie die Kontext-Wahrscheinlichkeiten $p(t|t', t'')$ und die lexikalischen Wahrscheinlichkeiten $p(w|t)$ für einen HMM-Wortart-Tagger bei Verwendung einer **Maximum-Likelihood**-Schätzung berechnen. (2 Punkte)

Aufgabe 6) Wozu dient **Parameter-Glättung**? Wie funktioniert eine Backoff-Glättung (mit Formel)? Wie unterscheidet sich die Kneser-Ney-Glättung von einer normalen Backoff-Glättung? (4 Punkte)

Aufgabe 7) Erklären Sie den Unterschied zwischen **überwachtem** und **unüberwachtem** Training anhand eines Beispiels. (1 Punkt)

Aufgabe 8) Welche Vorteile hat ein **Conditional Random Field**-Tagger gegenüber einem HMM-Tagger? (1 Punkt)

Aufgabe 9) Bei einem **CRF-Tagger** ist die Wahrscheinlichkeit einer Tagfolge wie folgt definiert:

$$p(t_1^n | w_1^n) = \frac{1}{Z(w_1^n)} \prod_{i=1}^{n+1} e^{\theta \cdot f(t_{i-1}, t_i, w_1^n, i)}$$

Wofür stehen die Ausdrücke θ , $f(t_{i-1}, t_i, w_1^n, i)$ und $Z(w_1^n)$?

Beschreiben Sie, was hier berechnet wird.

Wie ist $Z(w_1^n)$ definiert?

Wie wird ein solcher CRF-Tagger trainiert? Welcher Algorithmus wird dabei verwendet? Geben Sie die Formel an, mit der die wahrscheinlichste Tagfolge berechnet wird.

(5 Punkte)

>>>>>>>>>>>>>>> weiter auf der nächsten Seite >>>>>>>>>>>>>>>

Aufgabe 10) Wie funktioniert das Gradientenanstiegsverfahren (gradient ascent)? Wozu dient es? Was ist der Unterschied zwischen Gradientenabstieg und Gradientenanstieg? Welche Unterschiede gibt es zwischen batch gradient ascent, stochastic gradient ascent und minibatch gradient ascent? (3 Punkte)

Aufgabe 11) Wozu dient Regularisierung beim CRF-Training? Welche Varianten der Regularisierung kennen Sie? (1 Punkt)

Aufgabe 12) Wie **evaluieren** Sie einen Wortart-Tagger, einen Parser bzw. ein Sprachmodell? Welche Daten verwenden Sie jeweils? Welches Maß berechnen Sie, um verschiedene Systeme zu vergleichen? Wozu dient eine Baseline? (3 Punkte)

Aufgabe 13) In der Vorlesung haben wir den Einsatz von **Naive-Bayes**-Modellen für die Wortbedeutungsdesambiguierung behandelt. So wie dort die Lesart eines Wortes auf Basis der Kontextwörter desambiguiert wurde, so kann man auch die Klasse (bspw. das Themengebiet) eines Dokumentes auf Basis der darin enthaltenen Wörter bestimmen.

Geben Sie an, welchen mathematischen Ausdruck Sie maximieren müssen, wenn Sie mit einem Naive-Bayes-Modell die wahrscheinlichste Klasse c für ein Dokument mit den Wörtern w_1, \dots, w_n berechnen wollen. (Sie sollten dabei nicht davon ausgehen, dass die Klassen alle gleich wahrscheinlich sind.) (2 Punkte)

(30 Punkte insgesamt)