

Schriftliche Wiederholungsprüfung
Vorlesung Statistische Methoden in der Sprachverarbeitung
WS 2016/17
Dozent: Helmut Schmid

Aufgabe 1) Wie ist bei einem Markowmodell zweiter Ordnung (= Trigramm-Modell) über Buchstaben die Wahrscheinlichkeit einer Buchstabenfolge a_1, a_2, \dots, a_n allgemein definiert?

Wie ist die Wahrscheinlichkeit der Buchstabenfolge 'abc' konkret definiert? Was ist am Anfang und Ende der Buchstabenfolge zu beachten? (4 Punkte)

Aufgabe 2) Wie wird die bedingte Wahrscheinlichkeit $p(w|w', w'')$ (ohne Glättung) geschätzt? Geben Sie die Formel an und erklären Sie kurz, was die verwendeten Ausdrücke/Variablen bedeuten. (1 Punkt)

Aufgabe 3) Das Hidden-Markow-Modell (HMM) ist durch die folgende Formel charakterisiert:

$$\hat{t}_1^n = \arg \max_{t_1^n} \prod_{i=1}^{n+1} p(t_i | t_{i-k}, \dots, t_{i-1}) p(w_i | t_i)$$

Erklären Sie

- wofür w_i und t_i genau stehen
- welche Bedeutung k hat
- wieso das Produkt bis $i = n + 1$ berechnet wird
- welche Abhängigkeiten die beiden bedingten Wahrscheinlichkeiten jeweils beschreiben
- ob w_i von w_{i-1} statistisch unabhängig ist
- ob w_i von w_{i-1} statistisch unabhängig ist, wenn t_i bekannt ist

Geben Sie außerdem an, wie bei einem Bigramm-Tagger die Wahrscheinlichkeit der getaggten Wortfolge *es/PPER regnet/VVFIN* konkret definiert ist. (4 Punkte)

Aufgabe 4) Wie kann ein HMM-Tagger am besten mit unbekannten Wörtern umgehen? Erklären Sie, welche Änderungen an der obigen Formel des Modelles dazu notwendig sind. (3 Punkte)

Aufgabe 5) Geben Sie an, wie bei einer probabilistischen kontextfreien Grammatik folgende Werte definiert sind:

- die Wahrscheinlichkeit eines Parsebaumes

- die Wahrscheinlichkeit eines Satzes (= Menge von Parsebäumen)
- die Wahrscheinlichkeit eines Korpus (= Folge von Sätzen) (2 Punkte)

Aufgabe 6) Welche Maße werden üblicherweise verwendet bei der Evaluierung von

- Sprachmodellen
- Wortart-Taggern
- Parsern (2 Punkte)

Aufgabe 7) Der Berkeley-Parser von Petrov und Klein funktioniert für viele Sprachen recht gut.

- Was ist die Grundidee dieses Parsers?
- Geben Sie Pseudocode für das Parser-Training an (circa 5-10 Zeilen; wenn Sie nicht wissen, wie Sie Pseudocode schreiben sollen, dann beschreiben Sie, wie das Training funktioniert.) (4 Punkte)

Aufgabe 8) Lineare Modelle berechnen für ein Objekt x und eine Klasse y einen Merkmalsvektor $f(x, y)$ und multiplizieren ihn mit einem Gewichtsvektor θ . Diejenige Klasse, bei der das Produkt am größten ist, wird ausgegeben. Solche Modelle können mit dem Perzeptron-Algorithmus trainiert werden. Geben Sie Pseudocode für das Perzeptron-Training an. (3 Punkte)

Aufgabe 9) Wie kann auf Basis des Produktes $f(x, y) \cdot \theta$ von Merkmalsvektor und Gewichtsvektor eine Wahrscheinlichkeitsverteilung $p(y|x)$ definiert werden? (2 Punkte)

Aufgabe 10) Wie definiert ein loglinearen Modell die bedingte Wahrscheinlichkeit der Klasse y für ein gegebenes Objekt x auf Basis der Merkmale $f(x, y)$? (2 Punkte)

Aufgabe 11) CRF-Tagger können mit dem Gradientenanstiegs-Verfahren trainiert werden.

- Welche Zielfunktion wird beim Training eines CRF-Taggers maximiert?
- Die Ableitung dieser Zielfunktion ist die Differenz zwischen der beobachteten Häufigkeiten der Merkmale im Korpus und der erwarteten Häufigkeit der Merkmale. Mit welchem Algorithmus berechnet man die erwarteten Häufigkeiten?
- Worin unterscheiden sich Batch Gradient Ascent, Stochastic Gradient Ascent und Minibatch Gradient Ascent?

- Welche Funktion hat die Lernrate?
- Wozu dient eine Regularisierung?

(5 Punkte)

(32 Punkte insgesamt)

Viel Erfolg!