

9. Übungsblatt - Abgabe: 7.1.2014

Aufgabe 9.1 - Statistische Modellierung

Ein statistisches Modell zur Klassifikation von Wörtern w in Adjektive (*ADJA*) und Nichtadjektive (*NADJA*) wurde auf einigen Hundert Sätzen trainiert. Man erhält dabei die folgende Frequenztafel:

Wort w Artikel?	Wort $w+1$ großgeschrieben?	w endet auf -er/-es/-e/-en/-em?	ADJA	NADJA
Falsch	Falsch	Falsch	0	1758
Falsch	Falsch	Wahr	44	932
Falsch	Wahr	Falsch	3	473
Falsch	Wahr	Wahr	255	127
Wahr	Falsch	Falsch	0	0
Wahr	Falsch	Wahr	0	128
Wahr	Wahr	Falsch	0	32
Wahr	Wahr	Wahr	0	233

- Wie viele Klassen gibt es, wie heißen sie?
- Wie viele Features gibt es? Wie viele Werte haben die Features jeweils?
- Welches sind die möglichen Ereignisse, die sich aus unterschiedlichen Featurekombinationen ergeben? Wie groß ist also der Ereignisraum? Geben Sie zu jeder Kombination aus einem Ereignis aus dem Ereignisraum und einer Klasse falls möglich einen Beispielsatz/eine Beispielphrase an, der/die zu dieser Kombination passt. Markieren Sie darin jeweils das Wort, auf das sich die Kombination bezieht. Wenn Sie Probleme haben, ein geeignetes Beispiel zu finden, beschreiben Sie, woran das liegt.
- Für welchen Teil des Ereignisraumes (also für wieviele von wievielen potenziell möglichen Ereignissen) hat das Modell Trainingsinstanzen gesehen? Was bedeutet das für die Abdeckung des Modells auf neuen Daten?
- Formulieren Sie Regeln für einen simplen Klassifikator, die jedem Ereignis genau eine Klasse zuordnen.

Aufgabe 9.2 - Evaluation

Die Folien zur letzten Vorlesung zeigen eine Beispielevaluation für eine binäre Klassifikationsaufgabe in Adjektive (*ADJA*) und Nichtadjektive (*NADJA*) und gibt Precision, Recall und F-Score für die Klasse *ADJA* an.

- Berechnen Sie Precision und Recall für die Klasse *NADJA*.

- (b) Vergleichen Sie die Ergebnisse für *NADJA* mit den Ergebnissen für *ADJA*. Was fällt Ihnen auf? Beschreiben Sie umgangssprachlich, warum dieses Ergebnis zustande kommt.
 - (c) Welche der beiden Evaluationen ist für die tatsächliche Brauchbarkeit des Modells in der Praxis aussagekräftiger? Begründen Sie kurz.
-

Feedback

Bitte beantworten Sie kurz folgende Fragen:

- (a) Wie schwer fanden Sie den Stoff der aktuellen Vorlesung? (sehr einfach, einfach, durchschnittlich, schwer, sehr schwer)
 - (b) Wie viel Zeit haben Sie mit der Bearbeitung des aktuellen Übungsblattes verbracht?
 - (c) Haben Sie sonstige Fragen oder Anmerkungen zur aktuellen Vorlesung? Wenn ja, welche?
-

Abgabe in Gruppen von bis zu drei Studierenden bis **7.1.2014** 10 Uhr entweder als Email- Anhang im pdf-Format an **i2cl@coli.uni-sb.de** mit dem Betreff **i2cl homework 9** und Ihrer Matrikelnummer oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.