Musterlösungen zur schriftlichen Wiederholungsklausur
für die Übungen zu Statistische Methoden im WS 2020/21

Aufgabe 1)

a)

I) Modell: Naive Bayes-Modell

II) Anwendung: Textklassifikation

Beschreibung der Variablen: c=Klasse, d=Wortfolge des Textes, L=Textlänge k=Wortposition

III) Beispiel:

log p(Wirtschaft|Aktien fallen) = log p(Wirtschaft) + log p(Aktien|Wirtschaft) + log p(fallen|Wirtschaft)

b)

I) Modell: Markowmodell C-ter Ordnung

II) Anwendung: Sprachmodellierung

Beschreibung der Variablen: d=Wortfolge, L=Länge der Wortfolge, k=Wortposition

III) Beispiel für C=1:

log p(Aktien, fallen) = log p(Aktien|$\langle s \rangle$) + log p(fallen|Aktien) + log p($\langle s \rangle$|fallen)

c)

I) Modell: HMM C-ter Ordnung

II) Anwendung: Wortart-Tagging

Beschreibung der Variablen: c=Tagfolge, d=Wortfolge, L=Satzlänge, k=Wortposition

III) Beispiel:

log p(Aktien, NN, fallen, VVFIN) = log p(NN|$\langle s \rangle$) + log p(Aktien|NN) + log p(VVFIN|NN) + log p(fallen|VVFIN) + log p($\langle s \rangle$|VVFIN) + log p(ϵ|$\langle s \rangle$)

d)

I) Modell: loglineares Modell

II) Anwendung: Textklassifikation

Beschreibung der Variablen: c=Klasse, d=Wortfolge, z(d)=Normalisierungskonstante,
w_k=Gewicht, m_k=Merkmalsfunktion, k=Merkmalsindex

III) Beispiel: (Als Merkmalsfunktionen dienen die Wort-Häufigkeiten)

log p(Wirtschaft|Aktien, fallen) =
- log z(Aktien, fallen) + w(Aktien,Wirtschaft) · 1 + w(fallen,Wirtschaft) · 1

Aufgabe 2)

$$
\begin{aligned}
\delta_{\langle s \rangle}(0) &= 1 \\
\delta_{PRO}(1) &= \delta_{\langle s \rangle}(0)\, p(PRO|\langle s \rangle)\, p(we|PRO) = 1 \cdot 0.2 \cdot 0.2 = 0.04 \\
\psi_{PRO}(1) &= \langle s \rangle \\
\delta_{MD}(2) &= \delta_{PRO}(1)\, p(MD|PRO)\, p(can|MD) = 0.04 \cdot 0.3 \cdot 0.3 = 0.0036 \\
\psi_{MD}(2) &= PRO \\
\delta_{N}(2) &= \delta_{PRO}(1)\, p(N|PRO)\, p(can|N) = 0.04 \cdot 0.1 \cdot 0.1 = 0.0004
\end{aligned}
$$

$$
\begin{aligned}
\psi_N(2) &= PRO \\
\delta_{\langle s\rangle}(3) &= max(\delta_{MD}(2)\, p(\langle s\rangle|MD)\, p(\epsilon|\langle s\rangle), \delta_N(2)\, p(\langle s\rangle|N)\, p(\epsilon|\langle s\rangle)) \\
&= max(0.0036 \cdot 0.1 \cdot 1, 0.0004 \cdot 0.2 \cdot 1) \\
&= max(0.00036, 0.00008) = 0.00036 \\
\psi_{\langle s\rangle}(3) &= MD \\
t_2 &= \psi_{\langle s\rangle}(3) = MD \\
t_1 &= \psi_{MD}(2) = PRO
\end{aligned}
$$

Ergebnis-Tagfolge: PRO MD

Aufgabe 3)

Wir müssen zählen, wieviele Wörter nur TaggerA korrekt annotiert hat, und wieviele Wörter nur TaggerB richtig annotiert hat.

nur TaggerA korrekt: 5
nur TaggerB korrekt: 4

Wir haben also 9 Beispiele, die genau ein Tagger korrekt annotiert hat. Wir nehmen (nicht ganz korrekt) an, dass diese Beispiele eine Stichprobe von statistisch unabhängigen Ergebnissen bildet.

Nullhypothese: TaggerA ist nicht besser als TaggerB

Bei jedem Element der Stichprobe ist die Wahrscheinlichkeit, dass TaggerA richtig lag, maximal 0.5.

Wir summieren die Werte der Binomialfunktion für r-Werte ab 5: $p = \sum_{r=5}^{9} b(r, 0.5, 9)$

TaggerA ist signifikant besser als TaggerB, falls $p \leq 0.05$ gilt.

Aufgabe 4)

$$
\begin{aligned}
p(s|H,a,u) &= r(s|H,a,u) + \alpha(H,a,u)(\\
&\quad r(s|a,u) + \alpha(a,u)(\\
&\quad r(s|u) + \alpha(u)(\\
&\quad r(s))))
\end{aligned}
$$

Aufgabe 5)

bedingte Wahrscheinlichkeiten:

p(a|a) = 1/4
p(b|a) = 3/4
p(c|a) = 0

p(a|b) = 1/7
p(b|b) = 1/7
p(c|b) = 5/7

p(a|c) = 2/3
p(b|c) = 0
p(c|c) = 1/3

normal:

f(a) = 4
f(b) = 4
f(c) = 6

Kneser-Ney:

f(a) = 3
f(b) = 2
f(c) = 2