13. Übung AuD

Dominic Deckert

26. Januar 2017

Previously on ...

- Semi-Ring
- Aho-Algorithmus

mA_G

$$\begin{pmatrix} \{\varepsilon\} & \{b\} & \{b\} \\ \emptyset & \{\varepsilon\} & \{a\} \\ \{c\} & \emptyset & \{\varepsilon\} \end{pmatrix}$$

b)

$$\begin{pmatrix} \{\varepsilon\} & \{b\} & \{b\} \\ \emptyset & \{\varepsilon\} & \{a\} \\ \{c\} & \{cb\} & \{\varepsilon, cb\} \end{pmatrix} = D_G^1$$

$$\begin{pmatrix} \{\varepsilon\} & \{b\} & \{b, ba\} \\ \emptyset & \{\varepsilon\} & \{a\} \\ \{c\} & \{cb\} & \{\varepsilon, cb, cba\} \end{pmatrix} = D_G^2$$

c), d)

$$\begin{pmatrix} \{bc, bac\}^* & \{bc, bac\}^* \circ \{b\} & \{bc, bac\}^* \circ \{b, ba\} \\ \{a\} \circ \{cb, cba\}^* \circ \{c\} & \{a\} \circ \{cb, cba\}^* \circ \{cb\} & \{a\} \circ \{cb, cba\}^* \circ \{cba\} \\ \{cb, cba\}^* \circ \{c\} & \{cb, cba\}^* \circ \{cb\} & \{cb, cba\}^* \end{pmatrix} = D_G$$

$$D_{G'}(3,3) = \{cb, ba, cba\}^*$$

Zufallsexperimente

Zufallsexperiment:

- Ausgang kann nicht vorhergesagt werden
- wiederholbar(ohne dass sich verschiedene Experimente beeinflussen)

X: mögliche Ausgänge eines Zufallsexperiment
Korpus $h : X \to \mathbb{R}_{\geq 0}^{\infty}$ “zählt” Ausgänge bei mehrfacher Wiederholung
Größe eines Korpus: $|h| = \sum_{x \in X} h(x)$

Likelihood

Zufallsexperimente können durch Wahrscheinlichkeitsverteilung p beschrieben werden
$p : X \rightarrow [0, 1]$ mit $\sum_{x \in X} p(x) = 1$
Menge aller Wahrscheinlichkeitsverteilungen über X: $\mathcal{M}(X)$
"Wahrscheinlichkeit", dass ein Korpus von einem bestimmten p erzeugt wird:
Likelihood $L(h, p) = \prod_{x \in X} p(x)^{h(x)}$
Aufgabe: Gegeben einen Korpus, welche Wahrscheinlichkeitsverteilung hat diesen Korpus höchstwahrscheinlich erzeugt?

Wahrscheinlichkeitsmodell

Einschränkung der möglichen Wahrscheinlichkeiten: $M \subseteq \mathcal{M}(X)$
Häufig anhand von Wissen/ Annahmen über das Zufallsexperiment
Wahrscheinlichstes p: Maximum-Likelihood-Schätzung
$mle(M, h) = argmax_{p \in M}\{L(h, p)\}$
relative Häufigkeitsverteilung $rfe(h)(x) = \frac{h(x)}{|h|}$
Wenn $rfe(h) \in M$, dann ist $mle(M, h) = rfe(h)$

a)

$$X = \{1, 2, 3, 4, 5, 6\}$$
$$|h| = 18$$
$$M = \{p \in \mathcal{M}(X) | \forall i \in X : p(i) = p(6 - i)\}$$

x	1	2	3	4	5	6
h(x)	3	5	1	1	5	3
rfe(h)	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{5}{18}$	$\frac{1}{6}$

$rfe \in M$, also $rfe(h) = mle(M, h)$

b)

$X = \{K, Z\}^2$
$|h| = 18$
$M = \{p \in \mathcal{M}(X) | \exists p_1, p_2 \in \mathcal{M}(\{K, Z\}) : p(x_1, x_2) = p_1(x_1) \cdot p_2(x_2)\}$

x	(K, K)	(K, Z)	(Z, K)	(Z, Z)
h(x)	2	4	4	8
rfe(h)	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{4}{9}$

Mit $p_1(K) = p_2(K) = \frac{1}{3}$ folgt: $rfe \in M$, also $rfe(h) = mle(M, h)$

c)

$X = \{R, S, W\}$
$|h| = 10$
$M = \{p \in \mathcal{M}(X) | \forall i \in X : 5p(i) \in \mathbb{N}\}$

x	W	S	R
h(x)	4	2	4
rfe(h)	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

$rfe \in M$, also $rfe(h) = mle(M, h)$

unvollständiger Korpus

Manchmal sind von Zufalsexperiment nicht die Ausgänge (aus X) bekannt, sondern nur Interpretation der Ausgänge
Y: mögliche Interpretationen eines Ausgangs $x \in X$
$\mathit{yield} : X \to Y$ interpretiert Ausgänge
Analysator $A : Y \to Y$ analysiert Beobachtungen d.h. $x \in A(\mathit{yield}(x))$ Korpus unvollständiger Daten $h : Y \to \mathbb{R}^{\infty}_{\geq 0}$ gibt Anzahl der Beobachtungen an (tatsächliche Ausgänge nicht bekannt)

EM

Ziel: Gegeben einen unvollständigen Korpus, welches $q \in M$ hat dann die höchste Likelihood?
$\rightarrow$ nichtlineare Optimierung
Ansatz: schrittweises Erzeugen immer besserer q^i durch EM-Algorithmus
E-Schritt: Erzeuge aus letzter Wahrscheinlichkeitsverteilung (hypothetischen) vollständigen Korpus

$$h^i(x) = h(yield(x)) \cdot \frac{q^{i-1}(x)}{\sum_{x' \in A(yield(x))} q^{i-1}(x')}$$

$\rightarrow$ vollständiger Korpus wird aus unvollständigem anhand bedingter Wahrscheinlichkeit normiert
M-Schritt: Bestimme $q^i = mle(M, h^i)$

Berechnungen

Rechenregeln für diese Aufgabe:

- Münzen unabhängig, dh. $q_0(x_1, x_2) = q_0^1(x_1) \cdot q_0^2(x_2)$
- Teil-Korpora für erste / zweite Münze: addiere im vollständigen Korpus alle zugehörigen Elemente auf
  Bsp: $h_1^1(R) = h_1(R, Z) + h_1(R, K)$

Aufgabe 3

$$A(win) = \{(K,K),(Z,Z)\}$$

X	(K,K)	(Z, Z)	(K, Z)	(Z, K)	(R, K)	(R, Z)
yield(x)	win	win	lose	lose	lose	lose
q_0	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{4}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$
h_1	2	4	8	4	2	4
q_1	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$

Aufgabe 3

X^1	K	Z	R
h_1^1	10	8	6
q_1^1	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$

X^2	K	Z
h_1^2	8	16
q_1^2	$\frac{1}{3}$	$\frac{2}{3}$

Aufgabe 4

$A(\text{lose}) = \{(r, r), (g, g), (b, b)\}$

X	(r,r)	(g,g)	(b,b)	(r, g)	(r, b)	(g, r)	(g, b)	(b, r)	(b, g)
yield(x)	lose	lose	lose	win	win	win	win	win	win
q_0	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{12}$
h_1	9	12	3	6	3	2	2	1	2
q_1	$\frac{9}{40}$	$\frac{12}{40}$	$\frac{3}{40}$	$\frac{6}{40}$	$\frac{3}{40}$	$\frac{2}{40}$	$\frac{2}{40}$	$\frac{1}{40}$	$\frac{2}{40}$

Aufgabe 4

X^1	r	g	b
h_1^1	18	16	6
q_1^1	$\frac{9}{20}$	$\frac{2}{5}$	$\frac{3}{20}$

X^2	r	g	b
h_1^2	12	20	8
q_1^2	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{5}$