

Sprachtechnologie

1. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2016

Ausgabe am: 19.04.16
Besprechung am: 25.04.16

Aufgabe 1

Wahrscheinlichkeitsrechnung

Drei Personen spielen „Stille Post“, indem die erste Person einen Würfel mit nur 3 unterschiedlichen Ziffern würfelt und die gewürfelte Ziffer der 2. Person ins Ohr flüstert, die die gehörte Zahl dann der dritten Person weitervermittelt. Die drei möglichen Ziffern sind $\{1,2,3\}$. Die Wahrscheinlichkeit, dass '1' gewürfelt wird ist $P(1) = 0,3$ und die Wahrscheinlichkeit für '2' ist $P(2) = 0,5$. Beim Flüstern der Zahlen kommt es vor, dass die zuhörende Person eine Zahl falsch versteht. Die diskrete Verteilung, die die Wahrscheinlichkeiten von gehörten Zahlen gegeben gesagte Zahlen beschreibt, kann mithilfe folgender stochastischer Matrix beschrieben werden.

$$S = \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$$

D.h. die Wahrscheinlichkeit, dass Ziffer $h \in \{1,2,3\}$ gehört wurde, gegeben dass Ziffer $s \in \{1,2,3\}$ gesagt wurde, also $P(h|s)$, entspricht dem h -ten Element der Zeile s der Matrix S .

- Mit welchen Wahrscheinlichkeiten hat die zweite Person die einzelnen Ziffern gehört?
- Mit welchen Wahrscheinlichkeiten hat die dritte Person die einzelnen Ziffern gehört?
- Was ist der Erwartungswert der gewürfelten Zahl?
- Was ist der Erwartungswert der Zahl, die von der zweiten Person gehört wird?

Überlegen Sie, ob sich die Aufgaben z.B. mit Matrixmultiplikation oder Matrix-Vektor-Multiplikationen lösen lässt.

Aufgabe 2

Satz von Bayes

E-Mail-Nutzer bekommen häufig neben den eigentlich wichtigen E-Mails ungewollt Werbemails (Spam) zugeschickt. 85% aller erhaltenen E-Mails können zu Spam gezählt werden. In 45% der Spam-Mails und in 2% der eigentlich wichtigen Mails kommt der Satz "Sparen auch Sie Steuern" vor. Wie hoch ist die Wahrscheinlichkeit, dass eine Mail, die diesen Satz enthält, auch tatsächlich Spam ist?

Aufgabe 3

Satz von Bayes

Über Umfragen wurde festgestellt, dass 28% der Männer und 24% der Frauen rauchen. Desweiteren geben 15% der Männer ein Interesse für Radsport an. Der Anteil der Frauen mit der Vorliebe für diesen Sport beschränkt sich auf 5%. Angenommen 51% der Bevölkerung sind Frauen und die Interessen sind unabhängig voneinander. Mit welcher Wahrscheinlichkeit handelt es sich bei jemandem, der raucht und sich nicht für Radsport interessiert um einen Mann? Welches Attribut sagt mehr über das Geschlecht aus?