

Language Technology

Exercise 1

Prof. Tobias Scheffer
Uwe Dick

Summer 2016

Distribution at: 19-04-16
Discussion at: 25-04-16

Task 1

Stochastic Matrix

Three persons are playing the „telephone game“, where the first person rolls a dice with only three possible outcomes $\{1, 2, 3\}$ and whispers the outcome to the second person which in turn whispers the perceived number to the third person. The probability of the dice showing '1' is $P(1) = 0.3$ and the probability of '2' is $P(2) = 0.5$. The probabilities that the listening person interprets the whispered number wrongly can be described using a stochastic matrix

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

The probability that number $h \in \{1, 2, 3\}$ is heard while number $s \in \{1, 2, 3\}$ is whispered—that is, $P(h|s)$ —corresponds to the h -th element of row s of matrix S .

- What are the probabilities of each digit as perceived by the second person?
- What are the probabilities of each digit as perceived by the third person?
- What is the expected value of the dice's outcome?
- What is the expected value of the digit as perceived by the third person?

It might be helpful to use matrix multiplication and matrix vector multiplication to answer the questions.

Task 2

Bayes' Theorem

Email users often receive spam emails. In fact, let us assume that about 85% of received emails are spam. Additionally, let us assume that 45% of all spam mails contain the string "that will save you taxes". On the other hand, the same string is only contained in 2% of all other emails. What is the probability that an email which contains this string is in fact spam?

Task 3

Bayes' Theorem

A survey shows that 28% of all male and 24% of all female participants are smoking. Additionally, 15% of male participants are interested in cycling, in contrast to only 5% of surveyed women. Assuming that 51% of the population are female and both interests are independent from each other, how likely is it that somebody who smokes and who is not interested in cycling is a man? Which attribute is more expressive?