Ergänzungen zum maschinellen Übersetzen natürlicher Sprachen

1. Übungsblatt

2015-04-28

Aufgabe 1

Let $\mathbb{N}_{\geq 1}$ be the set of positive integers. The corpora $c_1, c_2, c_3 \colon \mathbb{N}_{\geq 1} \to \mathbb{R}_{\geq 0}$ are defined as follows:

$$c_1(1) = 5, \quad c_1(2) = 10, \quad c_1(3) = 5, \quad \text{and 0 otherwise;}$$
$$c_2(n) = 2^{-n};$$
$$c_3(n) = \frac{1}{n}.$$

Determine $\overline{p}(c_i)$ for every $i \in \{1, 2, 3\}$ and arbitrary $p \in \mathcal{M}(\mathbb{N}_{\geq 1})$.

Aufgabe 2

Let X be an arbitrary set, and $f \colon X \to \mathbb{R}$ and $g \colon \mathbb{R} \to \mathbb{R}$ be mappings such that g is strictly increasing, i.e., $\forall x, y \in \mathbb{R} \colon x < y \implies g(x) < g(y)$. Show that

1. $\forall x, y \in \mathbb{R} \colon g(x) < g(y) \implies x < y$ and
2. $\operatorname{argmax} f = \operatorname{argmax} g \circ f$.

Aufgabe 3

Mrs. Brown has two children.

- Assume that the firstborn child is a girl. What is the probability that the other child is also a girl?
- Assume that at least one child is a girl. What is the probability that the other child is also a girl?

Aufgabe 4

[Ben08] Suppose that 1 in 10000 people is a carrier of a certain virus. We have a test for this virus which gives a positive result if a person is a carrier with probability 0.99. The test also shows false positive results, i.e., a non-carrier tests positive, say with probability 0.0001. This sounds like a reliable and valuable test.

Suppose a person chosen at random from the population takes the test and the result is positive, what is the probability that the person is actually a carrier?

Aufgabe 5

You are a guest on a game show. During the last round of the show you are presented with three doors. A prize is hidden behind one of the doors (this door is chosen randomly). You are allowed to pick one door (you decide for the first door). After that the show host randomly chooses one of the doors that you did not pick and that does not contain the prize and then he opens it (he opens the second door).

If you are allowed to rethink your choice in the resulting situation and decide for the other door that has not been opened yet (i.e., the third door), would you do it?

Aufgabe 6
In a box there are two coins which look very much the same. However, one of them is fair and the other one shows twice as often head as ship. How can you decide which coin is the fair one?

Aufgabe 7
Again, you have two coins. Both coins look the same. This time the probabilities of head of coin 1 and coin 2 are 0.4 and 0.7, respectively. Assume that you choose one coin arbitrarily, toss it 10 times, and obtain the outcome "head" 5 times. Now you toss this coin another time and win a prize if you predict correctly. What do you predict? (Hint: "tail" is a bad choice! Why?) What if tossing the coin 100 times yields 50 times "head"?

Discuss implications of this result for the field of natural language processing, in particular the training of translation models.

Literatur

[Ben08] A. Ben-Naim. *A farewell to Entropy: Statistical Thermodynamics Based on Information*. World Scientific Pub Co Inc, 2008. ISBN: 9812707077.