Ergänzungen zum maschinellen Übersetzen natürlicher Sprachen

1. Übungsblatt

2016-04-12

Exercise 1
Let $a, c \in \mathbb{R}$. The *logarithm of c to base a*, denoted by $\log_a c$, is the unique $b \in \mathbb{R}$ such that $a^b = c$.

1. Recall some logarithmic identities.
2. We assume that $0^0 = 1$ and $\log 0 = -\infty$. Show that $0 \cdot (-\infty) = 0$.

Exercise 2
Let X be an arbitrary set, and $f : X \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ 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. $\operatorname{argmax} f = \operatorname{argmax} g \circ f$.

Exercise 3
The k-means algorithm partitions n data points $x_1, \dots, x_n \in \mathbb{R}^q$ into k clusters. The objective is to minimize $\sum_{j=1}^{n} d(x_j, \mu_{z_j})$ where z_j is the cluster assigned to x_j, μ_i is the mean of the i-th cluster, and d is a distance function. For each cluster i in $\{1, \dots, k\}$ there is an initial mean $\mu_i^0 \in \mathbb{R}^q$. The following two steps are iterated until convergence:

1. A cluster z_j is assigned to each data point x_j such that
$$z_j \in \operatorname{argmin}_{z \in \{1,\dots,k\}} d(\mu_z^t, z) \ .$$
2. New means are calculated:
$$\mu_i^{t+1} = \operatorname{average}(\{x_j \mid z_j = i\}) \ .$$

Apply the 2-means algorithm to the data points
$$(-2,-1), (0,-1), (0,-3), (2,2), (2,4), (4,2), (4,4)$$
with initial means $\mu_1^0 = (2,2)$ and $\mu_2^0 = (5,4)$, using the Euclidean distance.

Exercise 4
Mrs. Brown flips two fair coins.

1. Assume that the first coin comes up head. What is the probability that the other coin comes up head also?
2. Assume that at least one coin comes up head. What is the probability that the other coin comes up head also?

Exercise 5
[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?

Literatur

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