Statistical Models and Data Analysis

Summer term 2017

Problem Set 8 **5.7.2017**

The solutions to this exercise should be ready by 11 am on 13.7.2017. If you have any questions, please send me an email at stemmler@bio.lmu.de. You may submit your exercises by email in one file, but please be legible!

---

1. (Probabilities) Adapted from C. Bishop's "Pattern Recognition and Machine Learning", Springer 2006.

Imagine we have a red box and a blue box. In the red box, there are 2 apples and 6 oranges, and in the blue box there are 3 apples and 1 orange. Now suppose we randomly pick one of the boxes B, with probabilities $p(B = \mathsf{red}) = 40\%$ and $p(B = \mathsf{blue}) = 60\%$. From the chosen box we randomly select an item of fruit.

(a) Give the probabilities of getting an apple or an orange from the red and blue boxes, respectively. Formally this is written as the conditional probability $p(F \mid B)$ with F denoting the fruit.

(b) Determine the probability of choosing an apple (independently of which of which box was chosen). *Hint:* Use the sum and product rules of probability.

(c) Suppose that an orange has been chosen. What is the probability that it came from the red box?

(d) Given that you chose an orange, what is the probability that the next fruit you choose is an apple? You have not replaced the orange in the box that it came from. Optionally, please write a short poem comparing apples to oranges. Haikus, limericks, sonnets, you're free to choose.

2. (Bayes' Rule) Suppose that a disease occurs in a population with probability $p(d) = 0.001$. A diagnostic test claims to be 95% accurate, i.e., the test will give the correct answer in 95% of all cases, whether the person being tested actually has the disease or not. Let + indicate a positive test result, − a negative one.

(a) What are the probabilities of the test giving a false positive or a miss? In notational terms, we write

$$P(+|\neg\, d) \qquad \text{and} \qquad P(-|d),$$

where $\neg\, d$ stands for the person being tested not actually having the disease.

(b) Suppose the test result is positive. How likely is the person to have the disease? I.e., compute

$$P(d|+)$$

using Bayes' Rule.