Exercise Sheet #9

Fabian Schubert <fschubert@itp.uni-frankfurt.de>
Oren Neumann <neumann@itp.uni-frankfurt.de>

Problem 1 (*Bayesian Inference*) 10 Pts

We are observing a drunk person who has difficulties walking only in one direction. Looking at him, in the beginning we believed that the probability of a step forward (p) equals that of a step backwards ($1 - p$). However, looking at him again, we see that after an initial step back, he has walked straight forward for the last 4 steps (that is, you have the evidence $D = \{\Delta x_0 = 1, \Delta x_1 = 1, \Delta x_2 = 1, \Delta x_3 = 1, \Delta x_4 = 1\}$, where Δx i is the direction of each step). We can still assume that each step has two possible outcomes, with a fixed probability p of forward and $1 - p$ of backwards, but now with an unknown p value.

As a model of the world, we consider the $P(p)$ prior probability distribution of the parameter p. According to our expectations, this should be symmetric with respect to $\frac{1}{2}$.

(a) Compare two prior models, namely $P_1(p) \propto p(1 - p)$ and $P_2(p) = 1$:

(1) Calculate the $P_{1,2}(p|D)$ distributions for p, given the prior models P_1 and P_2 and the evidence.

(2) Using the chain rule and the previous result, calculate the probability that the next step will be forward, i.e. $P(\Delta x_5 = 1|D)$.

(b) How can you enhance your prediction of Δx_{n+1} after you watched the person taking one more step?

(c) A "naive" approach to estimating p from the evidence would be to divide the number of forward steps by the number of total steps, which, in this case would be $4/5 = 0.8$. Compare this to the estimated p of your inferred distributions, which is the value of maximal likelihood. Discuss the reason for possible differences.

Problem 2 (*Communication of Information from Correlated Sources*) 10 Pts

Imagine that we want to communicate data from two data sources X_A and X_B to a central location C via two noise-free communication channels. The signals x_A and x_B are strongly dependent, so their joint information content is only a little greater than the marginal information content of either of them. For example, C is a central weather forecasting agency that wishes to receive a string of binary reports saying whether it is raining in Aachen (x_A) and Bonn (x_B). The joint probability of x_A and x_B might be:

$P(x_A, x_B)$:

		x_A = 0	x_A = 1
x_B	0	0.49	0.01
	1	0.01	0.49

C would like to know N successive values of x_A and x_B exactly, but, since there is a cost to sending bits of information, it would like to avoid buying N bits from source X_A *and* N bits from source X_B. If we define the information rate $R_{A,B} = N/N_{A,B}$, where $N_{A,B}$ is the number of bits that are being sent over the respective channels:

Can X_A and X_B encoded in such a way that C can reconstruct all the variables, with the sum $R_A + R_B$ of information transmission rates being less than two bits?

(a) Using Shannon's source coding theorem, find expressions that define lower bounds for R_A and R_B, i.e. equations of the form $R_A \geq ...$, $R_B \geq ...$, $R_A + R_B \geq ...$. *Hint*: You need to consider two quantities: A lower bound for the total transmission rate, as well as for each individual channel/binary variable.

(b) Sketch the region of admissible transmission rates in a 2D-coordinate system representing R_A, R_B.

(c) Use the values given in the table and calculate the lower bound for the total transmission rate.