Bayes' Rule, Hyperbolic Tangents

Questions 1–4 adapted from "Bayesian Statistics" by P. M. Lee, Edward Arnold, 1989

1. Monozygotic twins (M) always have the same sex, whereas dizygotic twins (D) can be of opposite sex. If the sexes of a pair of twins are denoted GG, BB and $GB \equiv BG$, and given

$$\mathrm{P}(GG|M) = \mathrm{P}(BB|M) = \frac{1}{2}, \qquad \mathrm{P}(GB|M) = 0$$

$$\mathrm{P}(GG|D) = \mathrm{P}(BB|D) = \frac{1}{4}, \qquad \mathrm{P}(GB|D) = \frac{1}{2}$$

and that the probability of finding GG twins in the twin population as a whole is $\mathrm{P}(GG) = 0.4$, estimate the proportion of monozygotic twins.

2. There was a referendum held in 1975 to decide whether the U.K. should remain within the European Economic Community. At the time, 52% of the electorate supported the Labour Party (L) and the remainder the Conservative Party (C). 55% of Labour supporters and 85% of Conservative supporters voted "Yes" (Y) to remaining within the EEC.

   Suppose you met a man saying he voted "Yes", how could you infer which political party he supported, *i.e.* estimate $\mathrm{P}(L|Y)$ or $\mathrm{P}(C|Y)$.

3. 48 of the students taking a test were left–handed (L) and 259 were right–handed (R). Of the left–handed students 26 passed (P) whereas 206 of the right–handed students passed. Is there evidence for discrimination against the left–handed students? In answering this question, is it relevant to ask whether

$$\mathrm{P}(L|P) < \mathrm{P}(R|P) \qquad \text{or} \qquad \mathrm{P}(P|L) < \mathrm{P}(P|R) \quad ?$$

   Show that $\mathrm{P}(L|P)$ and $\mathrm{P}(P|L)$ are related by Bayes' theorem.

4. You are using two similar tensile test machines A and B to measure the strength of steel. You later discover that both machines have an intermittent fault, machine A giving an invalid result 10% of the time and machine B 5% of the time. 40% of your accumulated test data come from machine A. Find the probability that a particular invalid–result from these data comes from machine A.

5. Plot the function $\tanh(x)$ for the range $x = -360^\circ$ to $x = +360^\circ$. Notice that this gives you a sigmoidal curve. Now plot, using the same range, the function $\tanh(0.1 \times x)$. You should find that this is almost a straight line. This illustrates the flexibility of the hyperbolic tangent function; over a specified range, the shape of the function depends on the *weight*, which in the second case is equal to 0.1. This is why hyperbolic tangents are good model functions for neural networks. However, a single hyperbolic tangent is not sufficiently flexible to represent complex data.

   As another excercise, plot the function $\tanh(x)$-$\tanh(x - 180)$. This combination of two hyperbolic tangents gives a function which goes through a maximum. By combining hyperbolic tangents it is possible to create extremely flexible, non–linear functions. These functions are the basis of the neural networks that we shall use in the next examples class.

6. One difficulty with non–linear function is the problem of overfitting experimental data. After all, the function can be made sufficiently complex to pass through every data point, neglecting noise in the data. Such a function can be expected to generalise badly. Think of one way of avoiding overfitting experimental data.