Information Theory — Tutorial 5

Iain Murray

October 15, 2014

1. **More inference and prediction:** Ex. 3.14, p58.

2. **Entropy of more than one variable:** Review the definitions introduced in Chapter 8 of the textbook p138–140. Do Exercise 8.6, p140. And review the answer, p147.

   Definitions 8.1 to 8.8, and Figure 8.1, will all be examinable. If we haven't got to them in lectures yet, reviewing these for yourself to do this question should be entirely possible, and will help you follow the lectures. We won't use Definitions 8.9 or 8.10, the conditional mutual information expressions, in this course.

3. **Life in high-dimensions:** MacKay's book Exercise 2.20, p37.

4. **The data processing inequality:** Read MacKay Ex. 8.9, p141, and try to come up with a plain English explanation of what the inequality means. MacKay's answer, which is the standard approach also found in Cover and Thomas, uses technology we haven't covered in the lectures. It is possible to prove the inequality, albeit less neatly, using only methods we have covered. One strategy is given at the bottom of the page.

---

5. **More life in high-dimensions** *(optional)*:
   (a) Where is most of the volume of a (hyper-)cube in N dimensions?
   (b) Read: http://bit-player.org/2011/the-n-ball-game
       The high-N generalization of the 5 discs in a square is really mind-bending.
   (c) MacKay's book Exercise 6.14, p124.

Hint: one relatively brute force way of showing the data processing inequality is to average $D_{\mathrm{KL}}[P(w \mid d) \,\|\, P(w \mid r)]$ *under* $P(d, r)$.