Lecture 11, Friday w6, 2014-10-24

Finished off source coding loose end:

- Predictions can't just count how often things happened in a context.
  - Small contexts: model's predictions will be too vague.
  - Large contexts: won't have enough data to get good counts. No matter how much data we have, it makes sense to consider contexts of multiple sizes.
- Prediction by Partial Match (PPM). One simple way (frequently combined with arithmetic coding) to back off between contexts.

Started noisy channel coding:

- Communication: compress (remove redundancy), encode so robust to noise (add redundancy again!), send over channel, decode, decompress.
- Discrete memoryless channels:
  - Discrete: finite number of possible input and output symbols
  - Memoryless: nothing that happens to the channel affects its properties (unlike a neuron that can't fire any more after sustained activity, or a disk that flips more bits if too many 1's are stored in a row).
  - Can have different input and output alphabets (including different sizes).
  - Synchronized: we know the sequence of every received symbol. None are *deleted*, removed without knowing we lost a symbol, no extra symbols are *inserted*. The order is maintained.
- Example channels: BSC, BEC, Z.
- Channel defines Q matrix of transition probabilities. We don't usually get to choose that. We get to set vector of input probabilities: $\mathbf{p}_x$.
- Q and $\mathbf{p}_x$ specify whole joint distribution of x and y. Marginal distribution of output is $\mathbf{p}_y = Q\mathbf{p}_x$.
- Joint entropy of channel $H(X,Y) < H(X) + H(Y)$. Equality would mean channel is of zero use!
- The mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$ tells us how much information is shared between the input and output.

Check your progress

- Why do we compress (remove redundancy) if we're just going to add redundancy again?
- What's the difference between a *deletion* channel (which is hard to deal with), and an *erasure* channel (which is fairly easy)?
- Why is $\mathbf{p}_y = Q\mathbf{p}_x$?

We're up to around the 13th slide in the 'week 6' slides. Mark anything that's unclear or that needs expanding on NB.

This lecture covered around pp146–148 of the textbook.

Extra reading

If you'd like to know how compressors like gzip work (non-examinable) read MacKay p119, but also do Exercise 6.13 and then read its answer.

If you want to read ahead, look at Chapter 8 and the rest of Chapter 9.