PRACTICAL WORK 3.

2.2.9 Exercises

Exercise 2.9 [★]

Take a (short) piece of text and compute the relative frequencies of the letters in the text. Assume these are the true probabilities. What is the entropy of this distribution?

Exercise 2.10 [★]

Take another piece of text and compute a second probability distribution over letters by the same method. What is the KL divergence between the two distributions? (You will need to 'smooth' the second distribution and replace any zero with a small quantity E.)

Exercise 2.11 [★]

Cast the problem of word sense disambiguation as a noisy channel model, in analogy to the examples in table 2.2. Word sense disambiguation is the problem of determining which sense of an ambiguous word is used (e.g., 'industrial plant' vs. 'living plant' for *plant)* and will be covered in chapter 7.

Exercise 2.12 [★]

Show that the KL divergence is not symmetric by finding an example of two distributions p and q for which $D(p||q) \neq D(q||p)$.

Exercise 2.13 [★]

Prove the equality shown in the first two lines of (2.46).

Exercise 2.14 [★]

We arrived at the simplified way of computing cross entropy in equation (2.49) under the premise that the process we are dealing with is ergodic and stationary. List some characteristics of natural languages that show that these two properties are only approximately true of English.

Exercise 2.15 [★★]

Reproduce Shannon's experiment. Write a program that shows you a text one letter at a time. Run it on a text you have not seen. Can you confirm Shannon's estimate of the entropy of English?

Exercise 2.16 [★★]

Repeat the last exercise for one text that is 'easy' (e.g., a newsgroup posting) and one text that is 'hard' (e.g., a scientific article from a field you don't know well). Do you get different estimates? If the estimates are different, what difficulties does the experiment raise for interpreting the different estimates of the entropy of English?