8.581 Entropy, information and the brain

Problems for Lecture 6.

---

1. **Discrimination and asymmetry.** Imagine that you will see a signal x drawn from one of two probability distributions, $P_+(x)$ or $P_-(x)$. A priori, either possibility is equally likely. The first problem is to evaluate some function $f(x)$ that allows you to assign your observation to one distribution or the other. In principle this assignment could be probabilistic, so that $f(x)$ represents the probability of assigning x to the $+$ distribution. Then the average probability of an error is

$$E = \frac{1}{2}\int dx P_+(x)[1 - f(x)] + \frac{1}{2}\int dx P_-(x)f(x), \tag{1}$$

where the two terms reflect the two kinds of errors, misassignment of $+$ points to the $-$ distribution, and the vice versa.

   - Find the assignment function $f(x)$ that minimizes the error probability. Show that this function is in fact deterministic ($f(x) = 1$ or $f(x) = 1$), and that the optimal assignment is based on maximum likelihood: we assign x to $+$ if $P + (x) > P_-(x)$.
   - Consider the case where $P_\pm(x)$ are Gaussian distributions with the same variance but different means. Evaluate the minimum error probability (formally) and give asymptotic results for large and small differences in mean. How large do we need to make this 'signal' to be guaranteed only 1% errors?
   - Generalize these results to multidimensional Gaussian distributions, and give a geometrical picture of the assignment rule. This problem is easiest if the different Gaussian variables are independent and have equal variances. What happens in the more general case of arbitrary covariance matrices?
   - Now consider the case where $P_\pm(x)$ have the same mean but different variances. Show that the two different types of errors occur with different probabilities, as sketched in the lecture.

2. **Clusters of relevant information.** In the lecture we considered the following problem: A signal x provides information about another signal y, and we know the joint probability distribution $P(x, y)$; compress the signal $x \to C$ into fewer bits while preserving as much information about y as possible. Formally, this means that we should minimize

$$\mathcal{I} = I[C \to x] - \lambda I[C \to y], \tag{2}$$

where the variables over which we minimize are the assignment rules $P(C|x)$. Derive the self-consistent equation for these rules,

$$P(C|x) = \frac{P(C)}{Z(x)} \exp\left[-\lambda \sum_y P(y|x) \ln\left(\frac{P(y|x)}{P(y|C)}\right)\right]; \tag{3}$$

Remember that in minimizing you must preserve normalization,

$$\sum_C P(C|x) = 1. \tag{4}$$