Analytical exercises for the module "Models of Higher Brain Function"

Ingo Fründ

Natural Image Statistics I
April 29, 2010

The exercise is due at Monday, the 17^{th} of May at 6 p.m. Please upload your solutions at
`http://www.bccn-berlin.de/Graduate+Programs/Courses+and+Modules/Models+of+Higher+Brain+Functions/`

In the lecture the entropy of a random variable X was defined as

$$H(X) := -\mathbb{E}(\log_2(\mathbb{P}(X))).$$

Here, $\mathbb{E}(\,\cdot\,)$ denotes expectation, that is for a discrete random variable X with possible values $\{x_1, x_2, x_3, \ldots, x_N\}$, we have

$$\mathbb{E}(X) = \sum_{k=1}^{N} x_k \mathbb{P}(X = x_k).$$

If X is a continuous random variable with a density function f, we have

$$\mathbb{E}(X) = \int x f(x)\, \mathrm{d}x = \int X \, \mathrm{d}\mathbb{P}(X).$$

The indefinite integral $\int$ is here always assumed to go over the complete real axis.

In this exercise, we will study some properties of entropy.

Exercise 1 Discrete Entropy

We start with discrete entropy. What is the distribution with maximum entropy? Any transformation T that reduces redundancy will transform X to TX such that T is invertible and the entropy of TX is lower than the entropy of X. Thus, in natural image statistics, we are often concerned with transformations that avoid the distribution with maximum entropy.

Assume that X can take a finite number of states, i.e. $X \in \{x_i\}_{i=1}^{n}$. Show that the distribution with maximum entropy is the discrete uniform distribution, i.e.

$$\mathbb{P}(X = x_i) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Exercise 2 Properties of differential entropy

Typically, the luminances of natural images can be assumed to be continuous values. With continuous values the expectation in the definition of entropy becomes an integral. Here, we study the effect of some transformations on a continuous random variable.

Consider a continuous random variable X with density f and differential entropy $H(X)$. Derive expressions for $H(X+a)$ and $H(aX)$, for $a \in \mathbb{R}$. What happens if the standard deviation of X is changed?

Exercise 3 Entropy, Redundancy reducing transformations

In the lecture, we heart, that natural image statistics typically search for redundancy reducing transformations of input stimuli. We will explore this in more depth in this exercise. As a starting point, we look at the joint entropy of multiple random variables. This is defined in a straight forward way for random variables X and Y as

$$H(X,Y) = -\mathbb{E}(\log_2(\mathbb{P}(X,Y))).$$

First show, that if X and Y are independent (that is $\mathbb{P}(X,Y) = \mathbb{P}(X)\mathbb{P}(Y)$) then

$$H(X,Y) = H(X) + H(Y).$$

Now look at two binary random variables X, Y that can take the values 1 or 0. To code each of these random variables in isolation, we would need two bits (one for X and one for Y). How could the pair (X,Y) be transformed to a single variable such that the redundancy of this code is reduced. The joint probability distribution of X and Y is given by

$$\begin{aligned}
\mathbb{P}(X = 0 \wedge Y = 0) &= 0.5, \\
\mathbb{P}(X = 0 \wedge Y = 1) &= 0.25, \\
\mathbb{P}(X = 1 \wedge Y = 0) &= 0.125, \\
\mathbb{P}(X = 1 \wedge Y = 1) &= 0.125.
\end{aligned}$$

A short reminder about Lagrange multipliers

Lagrange multipliers are used to determine extrema under constraints. Suppose, we want to find an extremum of a function $f : \mathbb{R} \to \mathbb{R}$, then we know that we only need to look at the critical points that solve the condition

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = 0.$$

Consider now the situation, that we want to maximize the function $f : \mathbb{R}^n \to \mathbb{R}$. We can again find critical points that solve the conditions

$$\frac{\partial}{\partial x_i} f(x) = 0, \quad i = 1, \ldots, n.$$

Lagrange multipliers extend this idea by allowing for special conditions on the x_i. These conditions are typically given in the form of another function $h : \mathbb{R}^n \to \mathbb{R}$, in the form

$$h(x) = 0.$$

If we want to find the critical points of f subject to the condition $h(x) = 0$, we have to find the critical points of the function

$$L(x) = f(x) + \lambda h(x), \quad \lambda \in \mathbb{R}.$$

Obviously, the critical points depend on the special choice of λ. Thus, the critical points can be written as $x(\lambda)$. To find the critical points of the constrained optimization problem, we can simply solve the constraint equation

$$h(x(\lambda)) = 0,$$

for λ.

We will elaborate this in an example:

Find the maximum (with $x > 0$, $y > 0$, $z > 0$) of

$$V : \mathbb{R}^3_+ \to \mathbb{R}, \quad V(x, y, z) = 8xyz.$$

under the constraint that

$$x^2 + y^2 + z^2 = 1.$$

We set $h(x, y, z) := x^2 + y^2 + z^2 - 1$ and determine the derivatives

$$\begin{aligned}
\frac{\partial}{\partial x} L(x, y, z) &= 8yz + 2\lambda x, \\
\frac{\partial}{\partial y} L(x, y, z) &= 8xz + 2\lambda y, \\
\frac{\partial}{\partial z} L(x, y, z) &= 8xy + 2\lambda z.
\end{aligned}$$

Setting the derivatives to zero, we obtain

$$\begin{aligned}
0 &= 8yz + 2\lambda x \qquad \Big| \cdot \frac{x}{2} \\
0 &= 4xzy + \lambda x^2, \\
\implies 4xyz &= -\lambda x^2.
\end{aligned}$$

With analogous calculations for the other parameters, we obtain

$$4xyz = -\lambda x^2 = -\lambda y^2 = -\lambda z^2.$$

With $\lambda = 0$, this implies $xyz = 0$ and thus $V(x, y, z) = 0$. This is not the maximum.

With $\lambda \neq 0$, this implies $x^2 = y^2 = z^2$ and because, $x > 0, y > 0, z > 0$ also $x = y = z$. From the constraint, we see that in that case, we have

$$3x^2 = 1 \implies x = \frac{1}{\sqrt{3}}.$$

Thus, the constrained maximum of V is

$$V_{\max} = 8\frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} = \frac{8}{3\sqrt{3}} = \frac{8}{9}\sqrt{3}.$$