Biological Computation At System Level

Information Theory

Amir Assadi

Information

Consider the random variable $X = \{x_{-k}, ..., x_0, \ldots, x_k\}$

Let the event $x = x_k$ occurs with probability

$$P(\ x = x_k\ ) = p_k$$

where $0 \le P_k \le 1$, $k = -k, \ldots, 0, \ldots, k$. We define the amount of *information gained after observing* the event $x = x_k$ as

$$I(\ x_k\ ) = \log(1/\ p_k) = -\log p_k$$

Entropy

The average information when X occurs is

$$H(X) = E(I(X))$$

$$= \sum_{k} I(x_k)\, p(x = x_k)$$

$$= -\sum_{-k}^{k} p_k \log p_k$$

The quantity H(x) is called the entropy of X.

Shannon (1948): Entropy maximization.

Historical Remarks

- Craik (1943): How does the brain build a model of the outside world? Answer: By incorporating the regularity of structure stimuli subject to the perceived constraints.
- Attneave (1954) introduced the function of Information: A major function of the perceptual machinery is to strip away some of the redundancy of stimuli, to describe or encode information in a form more economical than the form that impinges on the receptors.
  => **Encoding data from a scene with *redundancy reduction* is related to *identification of specific features of the scene.***
- Linsker (1988) formulated the Maximum Mutual Information Principle: synaptic connections of a multilayered neural network develop in such a way as to maximize the amount of information that is transformed at each processing stage.

Maximum Entropy principle (Jaynes 1957, 1982)

When inference is made based on *incomplete information*, it should be done from the choice of probability distribution that maximizes entropy subject to the constraints on the distribution.

Example: Maximum Differential Entropy

With the differential entropy of a random variable x defined by:

$$h(x) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx$$

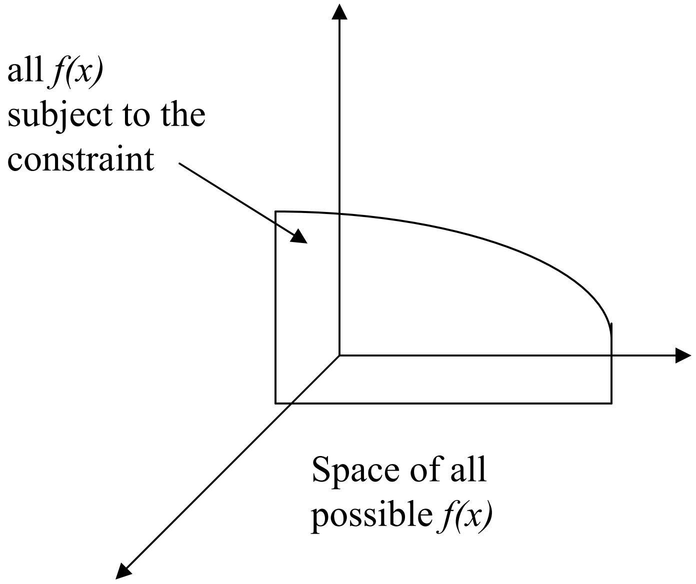
choose the probability function *f(x)* so that *h(x)* is optimized subject to the constraints on f(x) provided by experiments or observations.

For example, consider $\int_{-\infty}^{\infty} g_0(x) F(x) f(x) dx = \ldots.$

$g_0 = 1,$ *mean of F(x)*

$g_1(x) = (x-\mu)^2,$ $\int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \text{variance}$

Constraint Optimization Problem



The problem is a constraint optimization problem: find a point *f* for which *h* attain a max. We use the method of *Lagrange multiplier* to solve it.

Lagrange multiplier

$$J(f)=\int_{-\infty}^{\infty}[-f(x)\log f(x)+\lambda_0 f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)f(x)]dx$$

where λ is Lagrange's multiplier. Let *y = f(x)*,

$$J = \int_{-\infty}^{\infty}(-y\log y+\lambda_0 y+\sum_{i=1}^{m}\lambda_i g_i(x)y)dx$$

$$\frac{\partial J}{\partial y}=0 \Rightarrow -1-\log f(x)+\lambda_0+\sum \lambda_i g_i(x)=0$$

So, $$f(x)=\exp(-1+\lambda_0+\sum \lambda_i g_i(x))$$

Solve the problem

Suppose the constraints are *mean f(x)* = μ, *var(x)* = σ^2. Then

$$\int \exp(-1+\lambda_0+\lambda_1(x-\mu)^2)dx = 1$$

$$\int (x-\mu)^2 \exp(-1+\lambda_0+\lambda_1(x-\mu)^2)dx = 1$$

Solving for λ_0, λ_1, we get

$$\lambda_0 = 1 - \log(2\pi\sigma^2) \qquad \lambda_1 = \frac{1}{2\sigma^2}$$

Then $$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$