Derivations NIP course

Matthias Hennig

Maximum entropy, mean constraint

Here we derive the maximum entropy distribution for binary neurons, where the mean firing rates are known and used as constraints. We have N neurons with activity σ_i, which yields $M = 2^N$ possible states or patterns, each with unknown probability p_i.

The mean firing rates are estimated from the data like this:

$$< \sigma_i >_t = \frac{1}{T} \sum_{t=1}^{T} \sigma_i(t) = \nu_i$$

We can formulate this problem as a Lagrangian:

$$L = -\sum_{i=1}^{M} p_i \log p_i + \lambda \left(\sum_{i=1}^{M} p_i - 1 \right) + \sum_{j=1}^{N} h_j \left(\sum_{i=1}^{M} p_i \sigma_j - \nu_i \right)$$

The first term is the entropy, and our objective is to maximise this given the constraints in the next two terms. The second term ensures that all probabilities sum to one, this always has to be included. The third term implements the firing rate constraint, and yields N new parameters h_i.

To obtain expressions for these parameters, we have to find the global maximum of the Lagrangian. We can do this for all parameters in one go as they all enter the Lagrangian in the same way. So we pick neuron i and calculate:

$$\frac{\partial L}{\partial p_i} = -\log p_i - 1 + \lambda + \sum_{j=1}^{N} h_j \sigma_j$$

This we have to set to zero, which yields:

$$p_i = \exp \left(\lambda - 1 + \sum_{j=1}^{N} h_j \sigma_j \right)$$

Now we use the fact that $\sum_{i=1}^{M} p_i = 1$ by summing over all patterns:

$$
\begin{aligned}
1 &= \sum_{i=1}^{M} \exp\left(\lambda - 1 + \sum_{j=1}^{N} h_j \sigma_j\right) \\
\lambda &= 1 - \log \sum_{j=1}^{M} e^{\sum_i h_i \sigma_i}
\end{aligned}
$$

Putting this together gives:

$$
p_i = \frac{1}{\sum_{j=1}^{M} \exp(\sum_{i=1}^{N} h_i \sigma_i)} \exp(\sum_{i=1}^{N} h_i \sigma_i)
$$

This is the Maxwell Boltzmann distribution. A constraint on correlations can be added in the same way:

$$
< \sigma_i \sigma_j >_t = \frac{1}{T} \sum_{t=1}^{T} \sigma_i(t) \sigma_j(t)
$$

$$
\begin{aligned}
L &= -\sum_{i=1}^{N} p_i \log p_i + \lambda \left(\sum_{i=1}^{N} p_i - 1\right) + \sum_{j=1}^{N} h_j \left(\sum_{i=1}^{M} p_i \sigma_j - \nu_i\right) \\
&\quad + \sum_{j=1}^{N} \sum_{k=1}^{N} J_{jk} \left(\sum_{i=1}^{M} p_i \sigma_j \sigma_k - c_{jk}\right) \\
\implies & \\
p_i &= \frac{\exp\left(\sum_{j=1}^{N} h_j \sigma_j + \frac{1}{2} \sum_{j \neq k}^{N} J_{jk} \sigma_j \sigma_k\right)}{\sum_{i=1}^{M} \exp\left(\sum_{j=1}^{N} h_j \sigma_j + \frac{1}{2} \sum_{j \neq k} J_{jk} \sigma_j \sigma_k\right)}
\end{aligned}
$$

Note the factor $\frac{1}{2}$ reflects the fact that each pair is counted twice as the sums run over $i \neq j$, and that implicitly $J_{ii} = 0$.