Problem Set 3

Due November 17, 2022

Linear-nonlinear model

The optimal kernel derived from a linear model has two problems. First, there is nothing to prevent the predicted response to become negative. Neuron is a nonlinear device. The response of a neuron is typically characterized by the firing rate. This number *cannot* be negative. Moreover, in a linear model, the predicted response does not saturate. As the magnitude of the response increases, the response would also increase without bound. If we use L to represent the linear term we have been discussing thus far:

$$L(t) = \int_0^\infty d\tau D(\tau) s(t-\tau) \tag{1}$$

The modification is to replace the linear prediction $R_{est}(t) = R_0 + L(t)$ with the generalization

$$r_{est}(t) = r_0 + F[L(t)] \tag{2}$$

For example, one can choose $F(x) = [x]^+ = \mathbf{max}(x, 0)$, one can also set an upper bound so that the function saturates for large x, i.e., $F(x) = \mathbf{min}(x, F_0)$.

However, when nonlinearity is added, there is no guarantee that the derived kernel is optimal. A self-consistent solution for the optimal kernel should satisfy

$$\begin{aligned} D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma^2} &= \frac{1}{\sigma^2 T}\int_0^T r(t)s(t-\tau)dt \\ &\approx \frac{1}{\sigma^2 T}\int_0^T r_{est}(t)s(t-\tau)dt \\ &= \frac{1}{\sigma^2 T}\int_0^T F[L(t)]s(t-\tau)dt \end{aligned} \tag{3}$$

In general, the above equation does not hold. There is one exception. if the stimulus is Gaussian white noise, please show that the expected value of the integral satisfies

$$\frac{1}{\sigma^2 T}\int_0^T F[L(t)]s(t-\tau)dt = \frac{D(\tau)}{T}\int_0^T dt \frac{dF(L(t))}{dL} \tag{4}$$

The integral on the right hand side is a normalization condition. By properly scaling F, we can make $\frac{1}{T}\int_0^T dt \frac{dF(L(t))}{dL} = 1$.

Hint:

(1) For a Gaussian random variable x with zero mean and standard deviation σ, prove using integration by part that

$$\langle xF(\alpha x)\rangle = \alpha\sigma^2 \langle F'(\alpha x)\rangle, \tag{5}$$

where F is any function, α is a constant, and $\langle ... \rangle$ denotes the gaussian weighted average (or expected value),

$$\langle g(x)\rangle = \int dx g(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-x^2}{2\sigma^2}). \tag{6}$$

Extend your argument to multivariate functions and then to functional. For those who do not understand functional, please read my lecture notes carefully.

(2) Gaussian white noise has a gaussian probability density, with zero mean and variance $\sigma^2/\Delta t$, where Δt is the size of the time window.

Maximization of Entropy under Constraints

The Entropy of a variable X drawn from a distribution $p(X)$ is given by the following formula

$$H(X) = \int dX p(X) \ln p(X) \tag{7}$$

Use the Lagrange Multiplier method to evaluate the maximum entropy probability distribution $p(X)$ in the following cases:

(a) X is one dimensional continuous random variable, which takes only positive values and its mean is fixed. Hint: In addition to the mean, you should also take into account the constraint imposed by the normalization of p.

(b) There is no constraint on the range of X but its variance is given.

(c) X is an N-dimensional continuous random variable with constraint on the total variance,

$$\sum_i^N \langle x_i^2 \rangle = N\sigma^2 \tag{8}$$

(d) Show that the entropy of the multivariate Gaussian $N(\mathbf{X}|\mu, \mathbf{\Sigma})$ is given by

$$H(\mathbf{X}) = \frac{1}{2}\ln|\mathbf{\Sigma}| + \frac{D}{2}(1 + \ln(2\pi)) \tag{9}$$

where D is the dimensionality of $\mathbf{X}$, $|\mathbf{\Sigma}|$ is the determinant of the covariance matrix $\mathbf{\Sigma}$.

Fisher Information and Mutual Information

Consider a group of N cells responding to a stimulus x. The response of each neuron i is given by $r_i = wx + z_i$ where z_i varies from trial to trial as a gaussian random variable with zero mean and variance σ_i^2. The z_i are uncorrelated. The stimulus x is drawn from a gaussian distribution with zero mean and variance σ_0.

(a) Compute the Fisher Information of the system.

(b) Recall what I discussed in the class about the Maximum Likelihood (ML) estimate

$$\operatorname*{argmax}_x p(r|x).$$

Give an expression for the ML estimator of $\hat{x}$. Evaluate the mean square error of the estimate $\langle (x - \hat{x})^2 \rangle$, and compare with the result of (a).

(c*) Instead of using ML estimate, now let us consider the Byes rule:

$$p(x|r) = \frac{p(r|x)p(x)}{p(r)}. \tag{10}$$

We would like to find a Bayesian estimator $\hat{x}$ that would maximize the posterior distribution $p(x|r)$. Evaluate its mean square error and explain its dependence on σ_0. Compare your results with (b) and Under what conditions you expect that the two estimators will differ significantly? Explain the relationship between the results and the Cramer-Rau Bound.

(d) Compute the Mutual Information between the neuronal responses and the stimulus. What is the relationship between your result and the Fisher Information? Assume now that all $\sigma_i = \sigma$. How do both quantities depend on the population size.

Guide: In the above, the mean square error average $\langle ... \rangle$ is defined as an average over both neuronal noise and stimulus values. (*) indicates that this part has not been covered by our lectures. But solving it would give you extra credit.