Notes on GP CaKe: Effective Brain Connectivity with Causal Kernels

hannahpinson[1]

[1]Affiliation not available

January 12, 2018

Model Equations

The dynamics of each neuronal population $x_j(t)$ is modeled with an integro-differential equation:

$$\mathcal{D}_j x_j(t) = \sum_{i=1}^{N} (c_{i \to j} \star x_i)(t) + n_j(t) \tag{1}$$

where $c_{i \to j}$ is the kernel between source i and target j and n_j is Gaussian white noise with mean 0 and variance σ_n^2, and where $\mathcal{D}_j$ is a differential operator, given by:

$$\mathcal{D}_j = \alpha_0 + \sum_{p=1}^{P} \alpha_p \frac{d^p}{dt^p} \tag{2}$$

The construct the model, we thus need to estimate both the parameters of the differential operator as the shape of the kernels $c_{i \to j}$.

Estimating the Differential Operator

Given the largest power p, the parameters of the differential operator can be estimated by maximizing the univariate marginal likelihood of each individual source. The quality of this estimate is of course dependent on the strength of the coupling effects it neglects.

Example for $D = \frac{d}{dt} + \alpha$ (which is already implemented by the authors):

Equation for each source, but without connections, then becomes:

$$(\frac{d}{dt} + \alpha) x_j(t) = n_j(t) \tag{3}$$

with solution (a Gaussian Process)

$$y_j(t) = x_j(t) = c \exp(-\alpha t) + n_j(t) \tag{4}$$

where c is called the amplitude and α is called the relaxation constant. If we define f

$$f(t) = c\exp(-\alpha t) \tag{5}$$

then the (stationary) covariance function of f, σ_f, is given by:

$$\sigma_f^2(t_1, t_2) = c^2 \exp(-\alpha \mid t_1 - t_2 \mid) \tag{6}$$

Now we ought to estimate the univariate marginal likelihood as:

$$\log p(y \mid X) = -\frac{1}{2} y^t (\sigma_f^2 + \sigma_n^2 I) y - \frac{1}{2} \log \mid \sigma_f^2 + \sigma_n^2 I \mid - \frac{N}{2} \log 2\pi \tag{7}$$

with N the number of timesteps, y being the timeseries data, and X being the 'features' vector, here a vector of timesteps. (Note that σ_f^2 is implemented as a matrix and σ_n^2 as a constant.) A full discussion and derivations can be found in (Rasmussen, 2004).

In real applications we don't immediately know σ_n^2, and at this point in the code, the authors seem to just neglect it in their estimate. The maximal likelihood is determined from varying the parameters c and α over a grid (we could definitely optimize this approach). The value of c is for driving noise with $\sigma_n^2 = 1$ and no other input, equal to the timestep Δt of the discretization scheme used to generate the data.

Estimating the Causal Kernels

Fourier transforming Eq. into the frequency domain, we obtain:

$$\mathcal{P}_j(\omega) x_j(\omega) = \sum_{i=1}^{N} x_i(\omega) c_{i \to j}(\omega) + n_j(\omega), \tag{8}$$

or

$$x_j(\omega) = \sum_{i=1}^{N} \frac{x_i(\omega)}{\mathcal{P}_j(\omega)} c_{i \to j}(\omega) + \frac{n_j(\omega)}{\mathcal{P}_j(\omega)}. \tag{9}$$

This equation has again the form of a Gaussian Process, this time in the frequency domain. We can thus estimate the mean function, or the 'posterior expected value', of each kernel $c_{i \to j}(\omega)$:

$$\bar{c}_{i \to j} = K_{i \to j} \frac{x_j}{\mathcal{P}_j} \Big(\sum_l \frac{x_l}{\mathcal{P}_j} K_{l \to j} \frac{x_l}{\mathcal{P}_j} + \frac{\sigma_n^2}{\mid \mathcal{P}_j \mid^2} \Big)^{-1} x_j \tag{10}$$

with $K_{i \to j}$ the covariance matrix of the kernel from i to j $c_{i \to j}$ (aka "the kernel of the kernel"). These covariance functions $\mathcal{K}$ are restricted in shape, such that they correspond to causal, smooth and temporal localized kernels ($c_{i \to j}$) in the time domain. In order to achieve this, they have the proposed shape of:

$$\mathcal{K}(\omega 1, \omega 2) = \exp^{(-\nu \frac{\omega_1^2 + \omega_2^2}{2})} \big(\mathcal{K}_{temporal} + i \mathcal{H}(\mathcal{K}_{temporal}) \big) \tag{11}$$

with $\mathcal{H}$ the Hilbert transform, and

$$\mathcal{K}_{temporal}(\omega 1, \omega 2) = \exp^{\left(-\theta \frac{(\omega_1 - \omega_2)^2}{2} + it_s(\omega_1 - \omega_2)\right)}. \tag{12}$$

See the original paper for elaboration. To estimate the kernel $\bar{c}_{i\to j}$, we thus need to estimate 4 parameters: the smoothing factor ν (or the 'timescale'), the delay t_s, the temporal localization factor θ and finally the noise level σ_n. Once these are given, we can use them in the closed form expression 10.

In the code found on Github, the author implemented a procedure to estimate the timescale ν and the noise level σ_n :

1. use equation 9 to estimate the kernels $c_{i\to j}$ with a least squares regression
2. compute the covariance functions of these kernels
3. fit gaussians with timescale parameters to these covariance functions, choose the timescale of the best fit as the best parameter
4. use the residual variance as an estimate of the noise level
5. use k-mean clustering to find the centroid value of these parameters over all samples for each connection

An estimation of θ and t_s doesn't seem to be implemented.

References

Carl Edward Rasmussen. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*. Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-540-28650-9_4. URL https://doi.org/10.1007%2F978-3-540-28650-9_4.