Multilinear Gaussian Process

Mingliang Wang(1), Riccardo Sven Risuleo(1), Håkan Hjalmarsson(1)

Abstract

We study the identification of multilinear models with Gaussian distributed noises. We model each component in the multilinear structure as zero-mean independent Gaussian process (GP) models. We optimize hyperparameters of the GP model by maximizing the marginal likelihood of the data with the expectation maximization (EM) algorithm. In EM, we draw samples from the non-available posterior by Gibbs sampling. After EM converged, we use the optimized parameters to get the sample mean for each component to further identifications. The method is successfully applied to the model selection of bio-kinetics.

Motivation

Bilinear model

$$y[t] = f[t] \cdot g[t] + e[t]$$

- Independent GP models $f \sim \mathcal{N}(0, K_\alpha), g \sim \mathcal{N}(0, K_\beta)$
- White Gaussian noise $e[t] \sim \mathcal{N}(0, \sigma^2)$

Multilinear case

$$y[t] = \prod_{j=1}^{m} h_j[t] + e[t]$$

Objective *Estimate each component $h_j[t]$ from the data.*

Bilinear GP

Marginal likelihood

$$L(\theta) = \log \mathrm{p}(y|X, \alpha, \beta)$$

Expectation Maximization

E step: $Q^{(k)}(\alpha, \beta) = \iint \log \mathrm{p}(y, g, f|\alpha, \beta)\mathrm{p}(g, f|y, \alpha_k, \beta_k) df dg$

M step: $\alpha_{k+1}, \beta_{k+1} = \arg\max_{\alpha,\beta} Q^{(k)}(\alpha, \beta)$

- Posterior $\mathrm{p}(g, f|y, \alpha_k, \beta_k)$ are not available

$\Rightarrow$ We replace the integration in E step with a sampling step

$$\begin{aligned} Q^{(k)}(\alpha, \beta) =& \frac{1}{M}\sum_{\ell=1}^{M} \log \mathrm{p}(y|f_\ell^{(k)}, g_\ell^{(k)}, \sigma^2) \\ &+ \frac{1}{M}\sum_{j=1}^{M} \left\{ \log \mathrm{p}(f_\ell^{(k)}|\alpha) + \log \mathrm{p}(g_\ell^{(k)}|\beta) \right\} \end{aligned}$$

- Conditional distribution are Gaussian.

$\Rightarrow$ We we use Gibbs sampling to sample $f_j^{(k)}, g_j^{(k)}$ from the condition distributions,

$$\mathrm{p}(g|f, y, \alpha) = \mathcal{N}(m_g, P_g), \quad \mathrm{p}(f|g, y, \alpha) = \mathcal{N}(m_f, P_f),$$

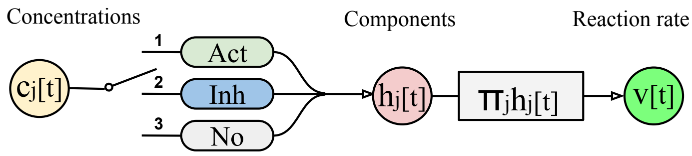
where $P_g^{-1} = \frac{1}{\sigma^2}D_f^2 + K_\beta^{-1}$, $m_g = \frac{1}{\sigma^2}P_g D_f y$, D_f is the diagnal matrix of f. Similar expressions hold for m_f and P_f.

Estimates of f, g

$$\hat{f} = \frac{1}{M}\sum_{\ell=1}^{M} f_\ell^*, \quad \hat{g} = \frac{1}{M}\sum_{\ell=1}^{M} g_\ell^*, \quad \hat{y} = \hat{f} \cdot \hat{g}$$

where f_ℓ^* and g_ℓ^* are samples from the Gibbs sampling of the posterior with optimized α_*, β_*.

Application: bio-kinetics selection



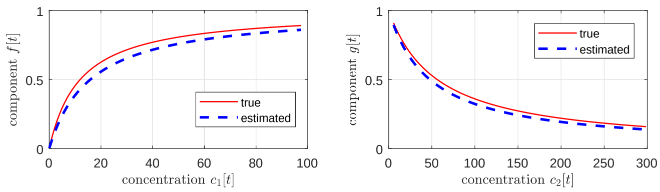
- **Act**: activation effect, $c_j[t]/(c_j[t] + \gamma_j)$
- **Inh**: inhibition effect, $\gamma_j/(c_j[t] + \gamma_j)$
- **No**: no effect, 1

Experiment 1

Bio-kinetics: $v[t] = v_m \frac{c_1[t]}{c_1[t]+12} \frac{\gamma_2}{c_2[t]+56} + e[t], \;\; v_m = 1.$

Experimental settings $e[t] \sim \mathcal{N}(0, 0.1\mathrm{Var}[v]), \; c_1[t] \in [0, 100], \; c_2[t] \in [0, 300].$

Prediction results

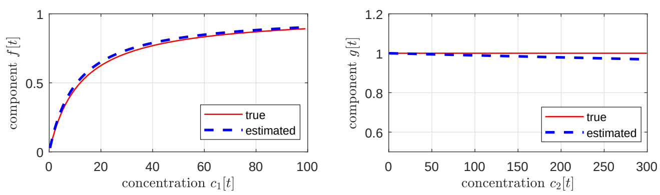


Experiment 2

Bio-kinetics: $v[t] = v_m \frac{c_1[t]}{c_1[t]+12} + e[t]$.

Same experimental settings as experiment 1.

Prediction results



Conclusions

We are able to estimate each component in the multi-linear model from the data.

Important results

- The noise variance σ^2 can be estimated as well.
- The method can be easily extended to multilinear cases.

(1) Department of Automatic Control, School of Electrical Engineering, KTH, Stockholm, Sweden