Opt. Report
Friday, April 7, 2023 9:34 AM

GAUSSIAN PROCESS SEMIPARAMETRIC REGRESSION

Suppose we observe $y_1, \ldots, y_n$ such that

$$y_i = \eta(x_i, \underline{\theta}) + \delta(x_i) + \varepsilon_i \ , \ i = 1, \ldots, n,$$

where $x_1, \ldots, x_n$ are the independent variables of the experiment taking values in $\mathcal{X}$, $\varepsilon_1, \ldots, \varepsilon_n$ are unobserved i.i.d. normal r.v.'s w/ mean zero and std. dev. $\sigma_\varepsilon > 0$, $\underline{\theta}$ is an unknown parameter taking values in $\Theta \subseteq \mathbb{R}^p$, and δ is an unknown real-valued function on $\mathcal{X}$.

Let $\underline{y} = [y_1, \ldots, y_n]$, $\underline{X} = [x_1, \ldots, x_n]^T$, $\underline{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^T$, and $\underline{\delta} = [\delta(x_1), \ldots, \delta(x_n)]^T$. Let $\sigma_\varepsilon^2 I$ denote the covariance matrix of $\underline{\varepsilon}$.

In GPSR, we construct the integrated log-likelihood by taking $\{\delta(x) : x \in \mathcal{X}\}$ to be a mean-zero Gaussian process with covariance function $k(\cdot, \cdot; \underline{\varphi})$ that is uncorrelated w/ the errors, where $\underline{\varphi} = [\sigma_\delta^2, \psi]$ is a vector of the GP noise and len scale. The result is a Gaussian likelihood w/ a mean depending on θ and a covariance matrix that reflects the covariance function of δ. Moreover, instead of including δ in the specification of the mean of $\underline{y}$, we modify the covariance structure of $\underline{y}$ to reflect the presence of δ in the model.

the covariance structure of Σ to reflect the

presence of δ in the model.

For the squared exponential kernel $k(\cdot,\cdot;\underline{\varphi})$, $K_{\underline{\varphi}}$ denotes the $n \times n$ covariance matrix w/ (i,j)th element generated by

$$k(x_i, x_j; \underline{\varphi}) = \sigma_\delta^2 \exp\left(-\left(\frac{x_i - x_j}{\psi}\right)^2\right),$$

where $\underline{\varphi} = [\sigma_\delta^2, \psi]$. Thus, we define all nuissance parameters to be $\underline{\omega} = [\underline{\varphi}, \sigma_\varepsilon^2]$. The integrated log-likelihood for $(\underline{\theta}, \underline{\omega})$ is:

$$\ell(\underline{\theta}, \underline{\omega}) = -\frac{1}{2} \log|\sigma_\varepsilon^2 I + K_{\underline{\varphi}}| - \frac{1}{2}(\underline{y} - X\underline{\theta})'(\sigma_\varepsilon^2 I + K_{\underline{\varphi}})^{-1}(\underline{y} - X\underline{\theta})^T.$$

where X is an $n \times p$ matrix that repeats the vector of observations $\underline{x}$ p times.

ESTIMATION

For a given value of of the hyperparameters $\underline{\omega}$, $\underline{\theta}$ can be estimated by using generalized least squares, leading to the estimator

$$\hat{\underline{\theta}}(\underline{\omega}) = (X^T(K_{\underline{\varphi}} + \sigma_\varepsilon^2 I)^{-1} X)^{-1} X^T (K_{\underline{\varphi}} + \sigma_\varepsilon^2 I)^{-1} \underline{y}.$$

However, $\underline{\omega}$ is generally unknown and must be estimated.

The MLE of $\underline{\omega}$ can be obtained by maximizing the profile log-likelihood for $\underline{\varphi}$, based on $\ell(\underline{\theta}, \underline{\omega})$ given by

$$\ell(\underline{\varphi}, \underline{\omega}) = -\frac{1}{2} \log|\sigma_\varepsilon^2 I + K_{\underline{\varphi}}| - \frac{1}{2}(\underline{y} - X\hat{\underline{\theta}}(\underline{\omega}))^T(\sigma_\varepsilon^2 I + K_{\underline{\varphi}})^{-1}(\underline{y} - X\hat{\underline{\theta}}(\underline{\omega}))^T.$$

Thus, we make predictions for new inputs $x^* \in \mathcal{X}$ w/ the true model $\zeta(\cdot)$ using the estimators $\hat{\underline{\theta}}$ and $\hat{\delta}(\cdot) = \delta(\cdot,\cdot;\hat{\underline{\varphi}})$, i.e.,

$$\zeta(x^*) = \eta(x^*; \hat{\underline{\theta}}) + \hat{\delta}(x^*)$$

OPTIMIZATION

The optimization problem is formulated as follows:

$$\underset{\underline{\omega} \in \Omega}{\arg\max} \quad -\frac{1}{2}\log|\sigma_\varepsilon^2 I + K_{\underline{\varphi}}| - \frac{1}{2}(\underline{y} - X\hat{\underline{\theta}}(\underline{\omega}))^T(\sigma_\varepsilon^2 I + K_{\underline{\varphi}})^{-1}(\underline{y} - X\hat{\underline{\theta}}(\underline{\omega}))^T$$

$$\text{s.t.} \quad \hat{\theta}(\underline{\varphi}) = (X^T(\sigma_\varepsilon^2 I + K_{\underline{\varphi}})^{-1}X)^{-1}X)^{-1}X^T(\sigma_\varepsilon^2 + K_{\underline{\varphi}})^{-1}\underline{y}$$

$$[K_{\underline{\varphi}}]_{i,j} = k(x_i, x_j; \underline{\varphi})$$

$$\underline{\omega} = [\sigma_\varepsilon^2, \sigma_\delta^2, \psi] \in \Omega \subseteq \mathbb{R}_+^3$$

$$\underline{x} \in \mathbb{R}^n, y \in \mathbb{R}^n, \underline{\theta} \in \mathbb{R}^p.$$