Lecture 22: Gaussian process regression

Professor Ilias Bilionis

Tuning the hyper-parameters

↳ may include parameters of the mean function, the covariance function, and the likelihood

PREDICTIVE
SCIENCE LABORATORY

The posterior over parameters and latent function values

$$f(\cdot)|\theta \sim GP(0, c(\cdot,\cdot;\theta))$$

$$c(x,x';\theta) = s^2 \exp\left\{-\frac{(x-x')^2}{2\ell^2}\right\}, \quad \theta = (s,\ell)$$

Observations: $x_{1:n},\ y_{1:n}$

Likelihood: $p(y_i|f(x_i),\sigma^2) = N(y_i|f(x_i),\sigma^2)$

$$\Downarrow$$

posterior G.P. $\rightarrow$ $f(\cdot)|x_{1:n},y_{1:n},\theta,\sigma \sim GP(m_n^*(\cdot), c_n^*(\cdot,\cdot))$

$\theta \sim p(\theta)$

$\sigma \sim p(\sigma)$

posterior of parameters conditioned on the data

$$p(\theta,\sigma|x_{1:n},y_{1:n}) \overset{\boxed{\text{Sum Rule}}}{=} \int p(f_{1:n},\theta,\sigma|x_{1:n},y_{1:n})\,df_{1:n}$$

$$\overset{\boxed{\text{Baye's Rule}}}{\propto} \int \overset{\text{likelihood}}{p(y_{1:n}|f_{1:n},\sigma)}\, p(f_{1:n}|x_{1:n},\theta)\, \overset{\text{priors}}{p(\theta)p(\sigma)}\, df_{1:n}$$

$$= \int \underbrace{p(y_{1:n}|f_{1:n},\sigma)}_{G} \times \underbrace{p(f_{1:n}|x_{1:n},\theta)\,df_{1:n}}_{G}\ p(\theta)p(\sigma)$$

$$= \underbrace{N(y_{1:n}|m_{1:n}, C_n+\sigma^2 I_n)\, p(\theta)p(\sigma)}_{\text{not Gaussian}}$$

Estimating the parameters by maximizing the marginal likelihood

MAP - maximum a posteriori estimate

$$\max_{\theta, \sigma} \log p(\theta, \sigma \mid x_{1:n}, y_{1:n})$$

$$= \log N(y_{1:n} \mid m_{1:n}, C_n + \sigma^2 I_n) + \log p(\theta) + \log p(\sigma)$$

$$= -\frac{1}{2}\log\left|C_n + \sigma^2 I_n\right| - \frac{1}{2}(y_{1:n} - m_{1:n})^T \left[C_n + \sigma^2 I_n\right]^{-1}(y_{1:n} - m_{1:n})$$

$$+ \text{constants} \ldots$$