Lecture 22: Gaussian process regression

Professor Ilias Bilionis

Gaussian process regression with measurement noise

PREDICTIVE
SCIENCE LABORATORY

The likelihood of the observations

$\varepsilon_i \sim N(0, \sigma^2)$

$x_{1:n} = (x_1, \dots, x_n)$; $y_{1:n} = (y_1, \dots, y_n)$; $y_i = f(x_i) + \varepsilon_i$.

$$p(y_i \mid f(x_i)) = N(y_i \mid f(x_i), \sigma^2)$$

$$f_{1:n} = (f(x_1), \dots, f(x_n))$$

$$p(y_{1:n} \mid f_{1:n}) = \prod_{i=1}^{n} p(y_i \mid f(x_i)) = N(y_{1:n} \mid f_{1:n}, \sigma^2 I)$$

Likelihood of observed data.

PREDICTIVE
SCIENCE LABORATORY

The joint probability density over observations and test points

$$x^*_{1:n^*} = (x^*_1, \dots, x^*_n)$$

$$f^*_{1:n^*} = (f(x^*_1), \dots, f(x^*_n))$$

$$f(\cdot) \sim GP(m(\cdot), c(\cdot,\cdot))$$

$$p(f_{1:n}, f^*_{1:n^*} \mid x_{1:n}, x^*_{1:n^*}) = N\left(\begin{pmatrix} f_{1:n} \\ f^*_{1:n^*} \end{pmatrix} \middle| \begin{pmatrix} m_{1:n} \\ m_{1:n^*} \end{pmatrix}, \begin{pmatrix} C_n & B \\ B^T & C_{n^*} \end{pmatrix}\right)$$

PREDICTIVE
SCIENCE LABORATORY

Conditioning on observations

Likelihood:

$$p(y_{1:n} \mid f_{1:n}) = \prod_{i=1}^{n} p(y_i \mid f(x_i)) = N(y_{1:n} \mid f_{1:n}, \sigma^2 I)$$

Joint:

$$p(f_{1:n}, f^*_{1:n^*} \mid x_{1:n}, x^*_{1:n^*}) = N\left(\begin{matrix} f_{1:n} \\ f^*_{1:n^*} \end{matrix} \,\middle|\, \begin{pmatrix} m_{1:n} \\ m_{1:n^*} \end{pmatrix}, \begin{pmatrix} C_n & B \\ B^T & C_{n^*} \end{pmatrix} \right)$$

We are after:

$$p(f^*_{1:n^*} \mid x_{1:n}, y_{1:n}, x^*_{1:n^*}) \overset{\text{Sum Rule}}{=} \int \underline{p(f_{1:n}, f^*_{1:n^*} \mid x_{1:n}, y_{1:n}, x_{1:n^*})}_{G} \, df_{1:n}$$

$$\overset{\text{Bayes' Rule}}{\propto} \int \underbrace{p(y_{1:n} \mid x_{1:n}, f_{1:n})}_{G} \times \underbrace{p(f_{1:n}, f^*_{1:n^*} \mid x_{1:n}, x^*_{1:n^*})}_{G} \, df_{1:n}$$

$$\overset{\text{complete the square}}{=} N(f_{1:n^*} \mid m^*_{1:n^*}, C^*_{n^*})$$

$$m^*_{1:n} = m_{1:n^*} - B^T \left[C_n + \sigma^2 I_n\right]^{-1} (y_{1:n} - m_{1:n})$$

$$C^*_{n^*} = C_{n^*} - B^T \left[C_n + \sigma^2 I_n\right]^{-1} B$$

PREDICTIVE SCIENCE LABORATORY

The posterior Gaussian process

!!

$$f(\cdot) \mid x_{1:n}, y_{1:n} \sim GP\left(\underbrace{m_n^*(\cdot)}_{\text{posterior mean function}}, \underbrace{c_n^*(\cdot,\cdot)}_{\text{posterior covariance function}}\right)$$

$$m_n^*(x) = m(x) - c(x, x_{1:n})\left[C_n + \sigma^2 I_n\right]^{-1}(y_{1:n} - m_{1:n})$$

$$c_n^*(x, x') = c(x, x') - \underbrace{c(x, x_{1:n})}_{= \; 1\times n \; = \; (c(x,x_1), \dots, c(x,x_n))}\left[C_n + \sigma^2 I_n\right]^{-1} \underbrace{c(x_{1:n}, x')}_{= \; n \times 1 \; = \; \begin{pmatrix} c(x_1, x') \\ \vdots \\ c(x_n, x') \end{pmatrix}}$$

The point predictive distribution

$$p(f(x) \mid \underline{x_{1:n}, y_{1:n}}) \overset{\text{post.}}{\underset{GP}{=}} N\left(f(x) \mid m_n^*(x), \underset{\underset{c_n^*(x,x)}{\parallel}}{\sigma_n^{*2}(x)}\right)$$

→ Best I can say about the function values.
Uncertainty here is epistemic.

$$p(y \mid x, x_{1:n}, y_{1:n}) \overset{\text{sum}}{\underset{\text{Rule}}{=}} \int \underbrace{p(y \mid f(x))}_{N(y \mid f(x), \sigma^2)} \overset{G \times G}{\overline{p(f(x) \mid x_{1:n}, y_{1:n})}}\, df(x)$$

$$\overset{\text{complete}}{\underset{\text{the square}}{=}} N\left(y \mid m_n^*(x), \underbrace{\sigma_n^{*2}(x)}_{\text{epistemic}} + \underbrace{\sigma^2}_{\text{aleatory}}\right)$$

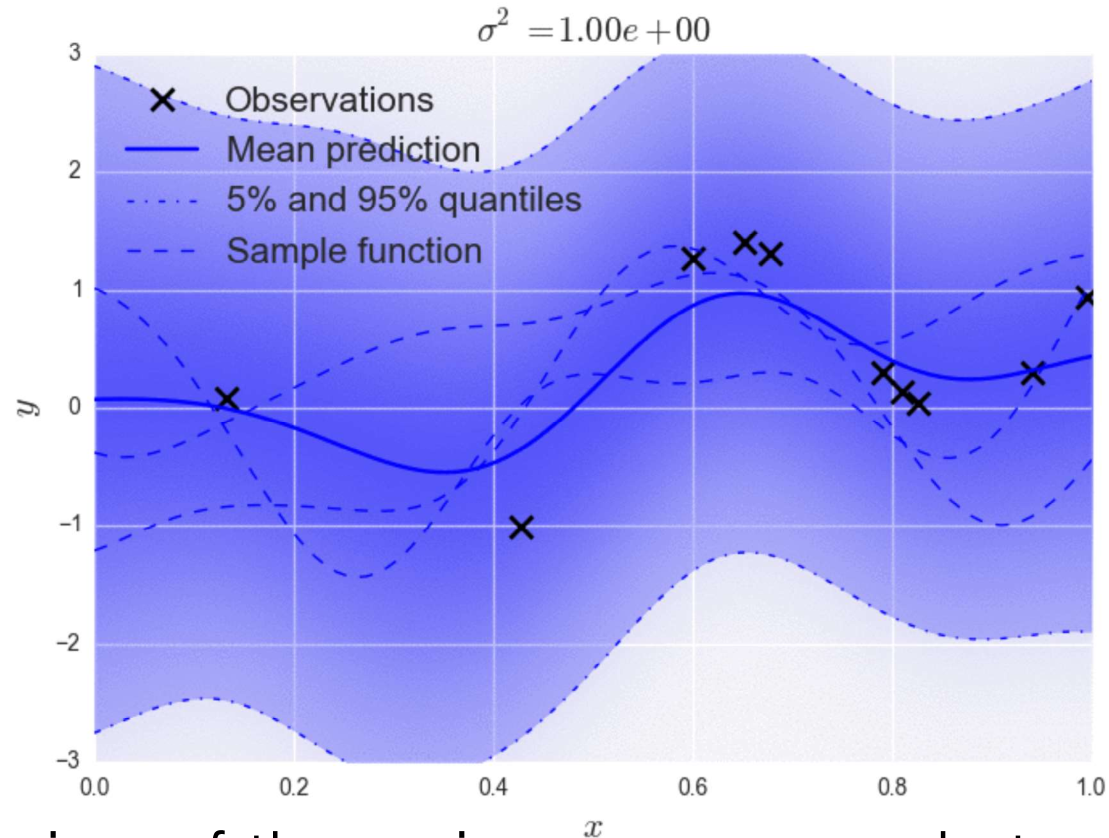
$\Downarrow$ 95% credible int.

$$f(x) \in \left[m_n^*(x) - 2\sigma_n^*(x),\ m_n^*(x) + 2\sigma_n^*(x)\right]$$

$$y \in \left[m_n^*(x) - 2\sqrt{\sigma_n^{*2}(x) + \sigma^2},\ m_n^*(x) + 2\sqrt{\sigma_n^{*2}(x) + \sigma^2}\right]$$

PREDICTIVE SCIENCE LABORATORY

Gaussian process regression - Noisy observations



Each choice of the noise corresponds to a different interpretation of the data.

PREDICTIVE SCIENCE LABORATORY

Even when there is not any noise, including it improves numerical stability

- It is common to use small noise even if there is not any in the data.
- Cholesky fails when covariance is close to being semi-positive definite.
- Adding a small noise improves numerical stability.
- It is known as the “jitter” or as the “nugget” in this case.