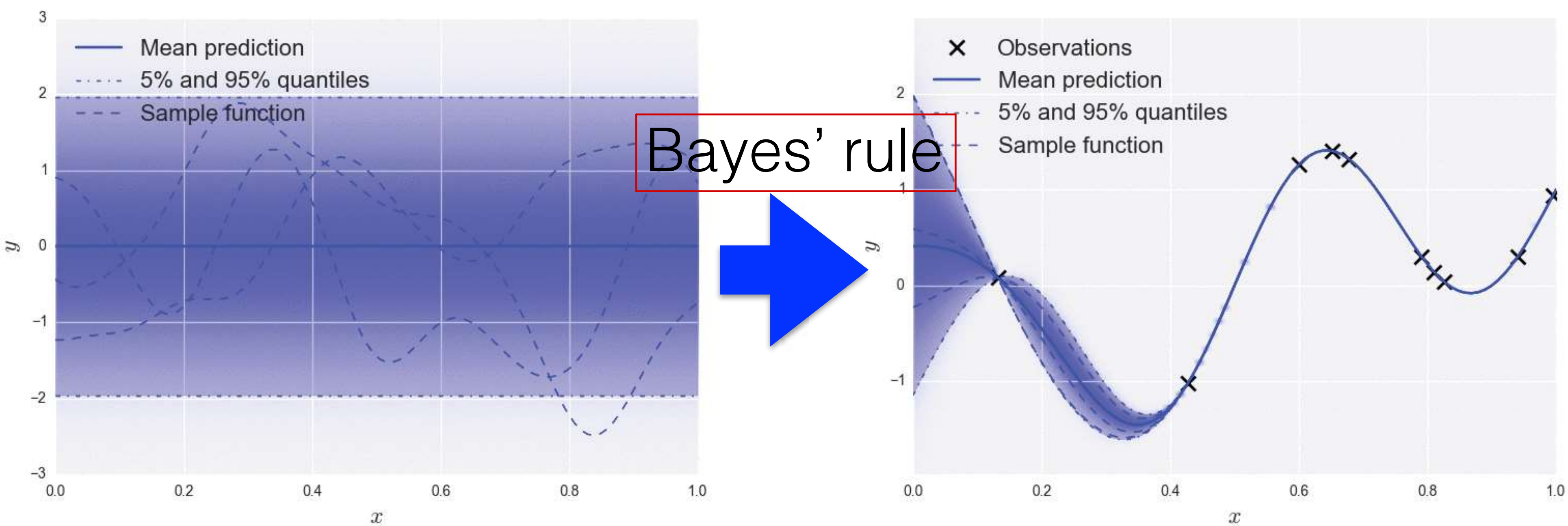
# Lecture 22: Gaussian process regression

Professor Ilias Bilionis

## Gaussian process regression without measurement noise

PREDICTIVE
SCIENCE LABORATORY

# How does Gaussian process regression work?




Prior GP

Posterior GP

PREDICTIVE
SCIENCE LABORATORY

# The joint probability density of observations

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

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

$$\Downarrow$$

joint observations
condioned on the inputs

$$p(f_{1:n} | x_{1:n}) = N\Big(f_{1:n} \,\Big|\, \underbrace{m_{1:n}}_{= \begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}}, \underbrace{C_n}_{= \left(c(x_i, x_j)\right)_{i,j=1}^{n}}\Big)$$

# The joint probability density over observations and test points

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

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

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

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

joint of observed & test function values conditioned on all inputs

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

(n+n*) x (n+n*) size

$$m_{1:n^*} = \begin{pmatrix} m(x^*_1) \\ \vdots \\ m(x^*_{n^*}) \end{pmatrix}$$

$$C_{n^*} = \left(c(x^*_i, x^*_j)\right)_{i,j=1}^{n^*}$$

$$B = \left(c(x_i, x^*_j)\right)_{i=1, j=1}^{n \quad n^*}$$

observed, test

cross-covariance $n \times n^*$

# Conditioning on observations

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

$$\Big\Downarrow \text{ completing the square}$$

$$p(f^*_{1:n^*} \mid x_{1:n}, f_{1:n}, x^*_{1:n^*}) = N\left(f^*_{1:n^*} \mid m^*_{1:n^*}, C^*_{n^*}\right)$$

$$m^*_{1:n^*} = m_{1:n^*} - B^T C_n^{-1}(f_{1:n} - m_{1:n})$$

$$C^*_{n^*} = C_{n^*} - \underbrace{B^T C_n^{-1} B}_{\text{correction terms}}$$

PREDICTIVE
SCIENCE LABORATORY

# The posterior Gaussian process

summarizes our state of knowledge after the observations are made

$$p(f^*_{1:n^*} \mid x_{1:n}, f_{1:n}, x^*_{1:n^*}) = \mathcal{N}\left(f^*_{1:n^*} \mid m^*_{1:n^*}, C^*_{n^*}\right)$$

$\Downarrow$ test inputs are arbitrary

posterior Gaussian Process

$$f(\cdot) \mid x_{1:n}, f_{1:n} \sim GP\left(\underbrace{m^*_n(\cdot)}_{\text{posterior mean function}}, \underbrace{c^*_n(\cdot,\cdot)}_{\text{posterior covariance function}}\right)$$

$$m^*_n(x) = m(x) - \underbrace{c(x, x_{1:n})}_{\substack{\text{cross covariance} \\ 1\times n \\ = \\ (c(x,x_1), \dots, c(x,x_n))}} C_n^{-1}(f_{1:n} - m_{1:n})$$

$$c^*_n(x,x') = c(x,x') - c(x, x_{1:n})\, C_n^{-1}\, c(x_{1:n}, x')$$

$$c(x_{1:n}, x') = \begin{pmatrix} c(x_1, x') \\ \vdots \\ c(x_n, x') \end{pmatrix} \quad n\times 1$$

# The point predictive distribution

$$f(\cdot) \mid x_{1:n}, f_{1:n} \sim GP\left(m_n^*(\cdot), c_n^*(\cdot, \cdot)\right)$$

$$\Downarrow$$

single test point

$$p(f(x) \mid x_{1:n}, f_{1:n}) = \mathcal{N}\left(f(x) \mid m_n^*(x), \sigma_n^{*2}(x)\right)$$

posterior covariance function

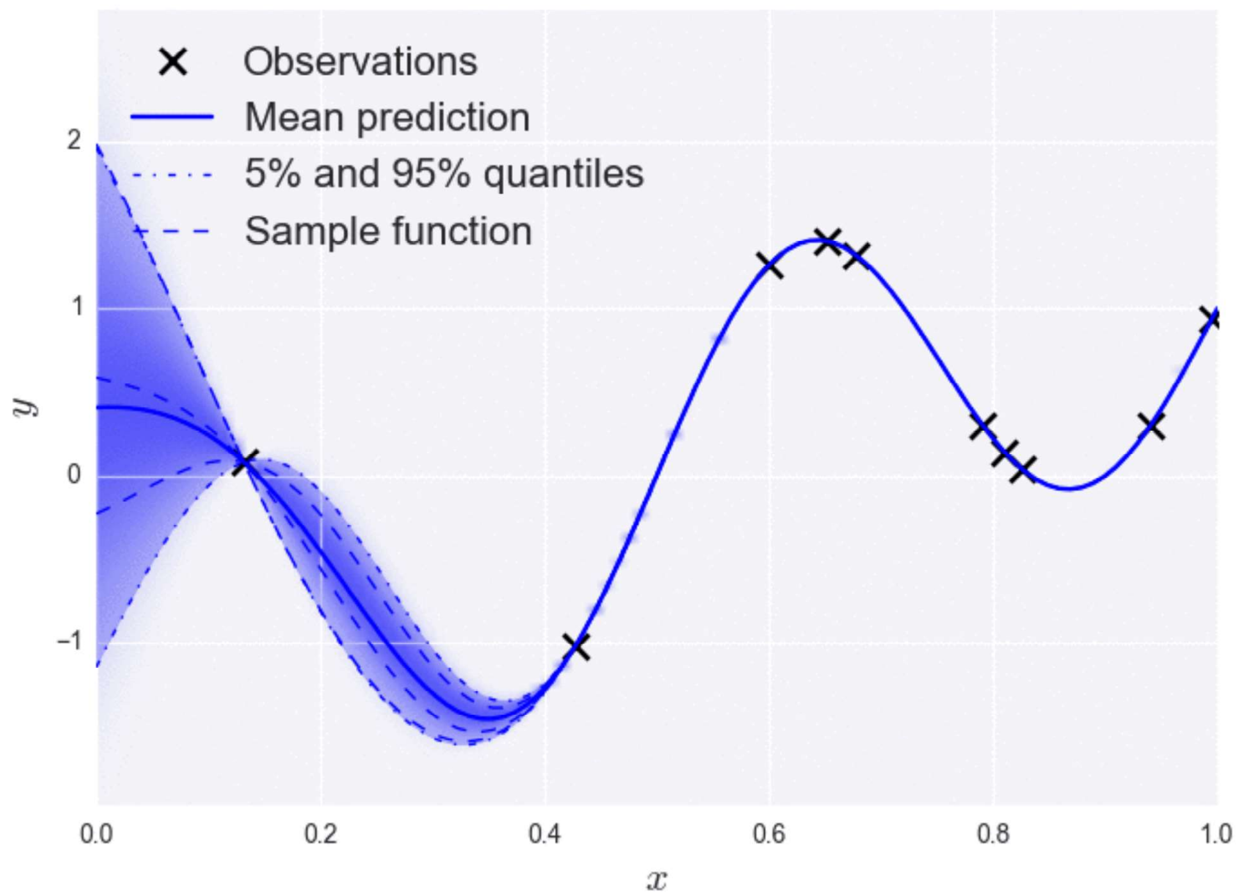
$$\sigma_n^{*2}(x) = c_n^*(x, x).$$

$\Downarrow$ credible interval $\sim 95\%$

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

PREDICTIVE
SCIENCE LABORATORY

# Example




PREDICTIVE
SCIENCE LABORATORY