10-424/624: Bayesian Methods in ML Lecture 9: Supplement & Figures

Henry Chai

2/11/25

Some old friends

Gaussian process =

Bayesian linear regression + Kernels

A new perspective

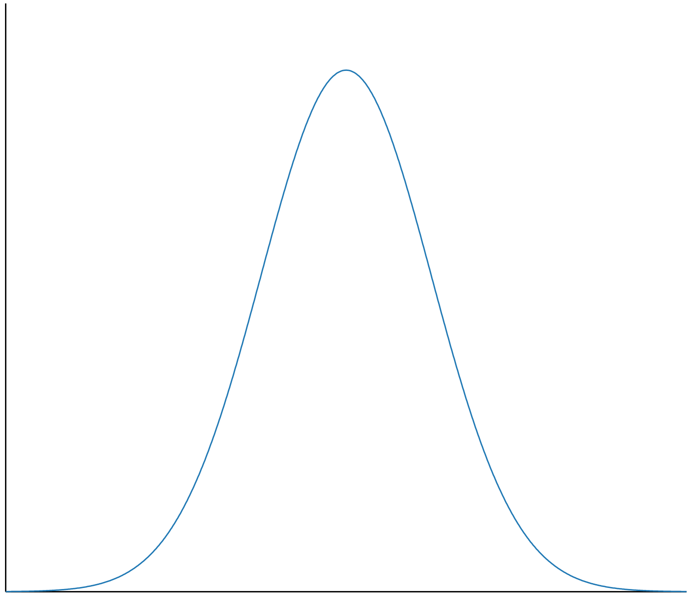
Gaussian process =

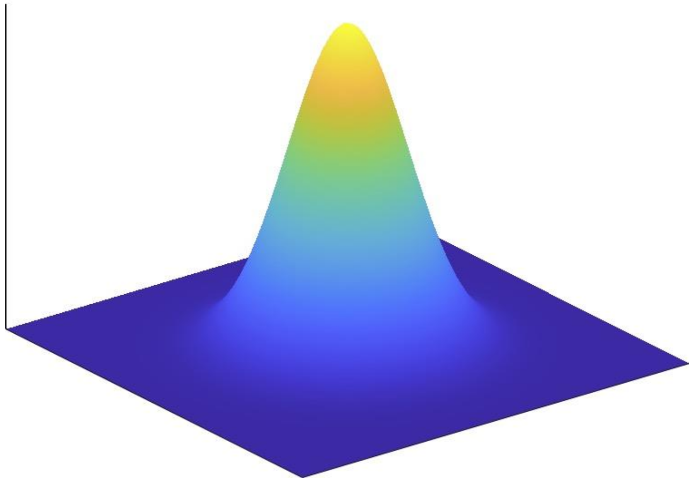
The extension of a Gaussian distribution to functions

Gaussians

- (Univariate) Gaussians:

  $x \sim \mathcal{N}(x; \mu = 0, \sigma^2 = 1)$





- Multivariate Gaussians:

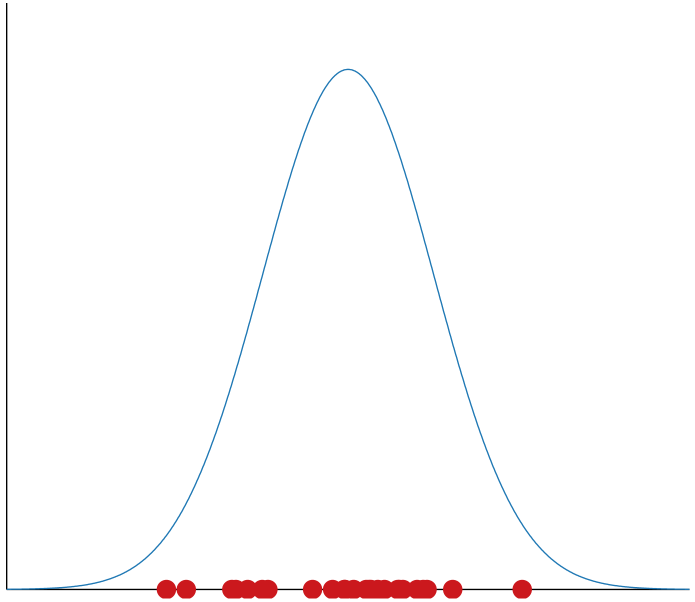
  $\boldsymbol{x} = [x_1, \ldots, x_D]^T$

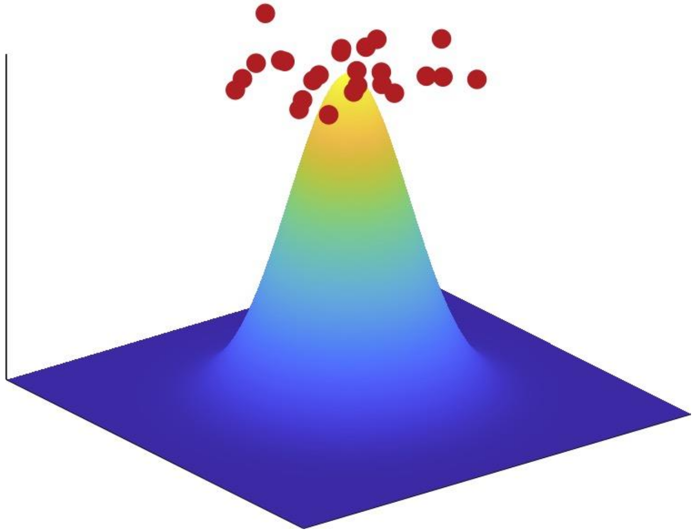
  $\sim \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu} = \boldsymbol{0}_D, \Sigma = I_D)$

Gaussians

- (Univariate) Gaussians:

  $x \sim \mathcal{N}(x; \mu = 0, \sigma^2 = 1)$





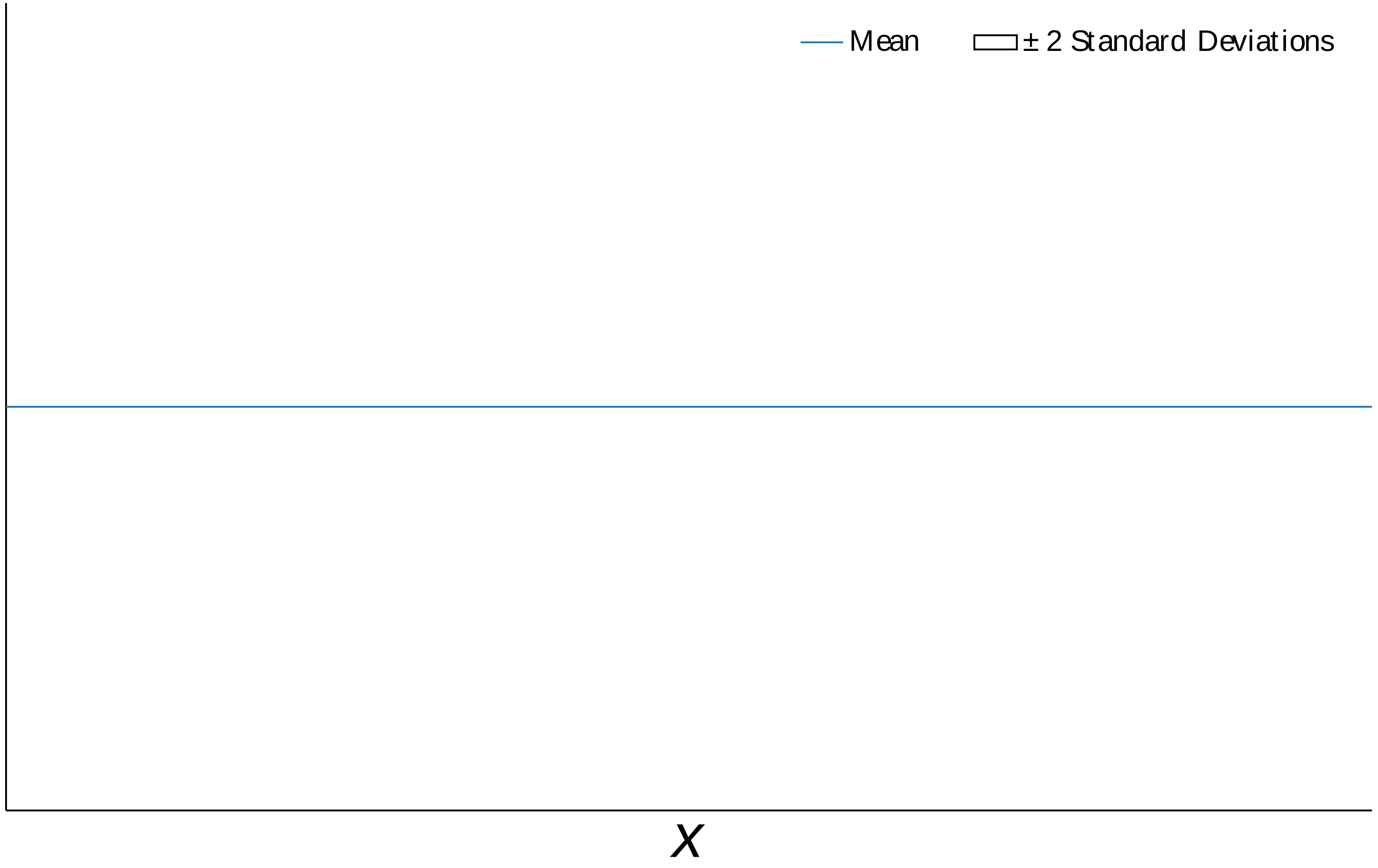
- Multivariate Gaussians:

  $\boldsymbol{x} = [x_1, \dots, x_D]^T$

  $\sim \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu} = \mathbf{0}_D, \Sigma = I_D)$

Gaussian Process (GP)

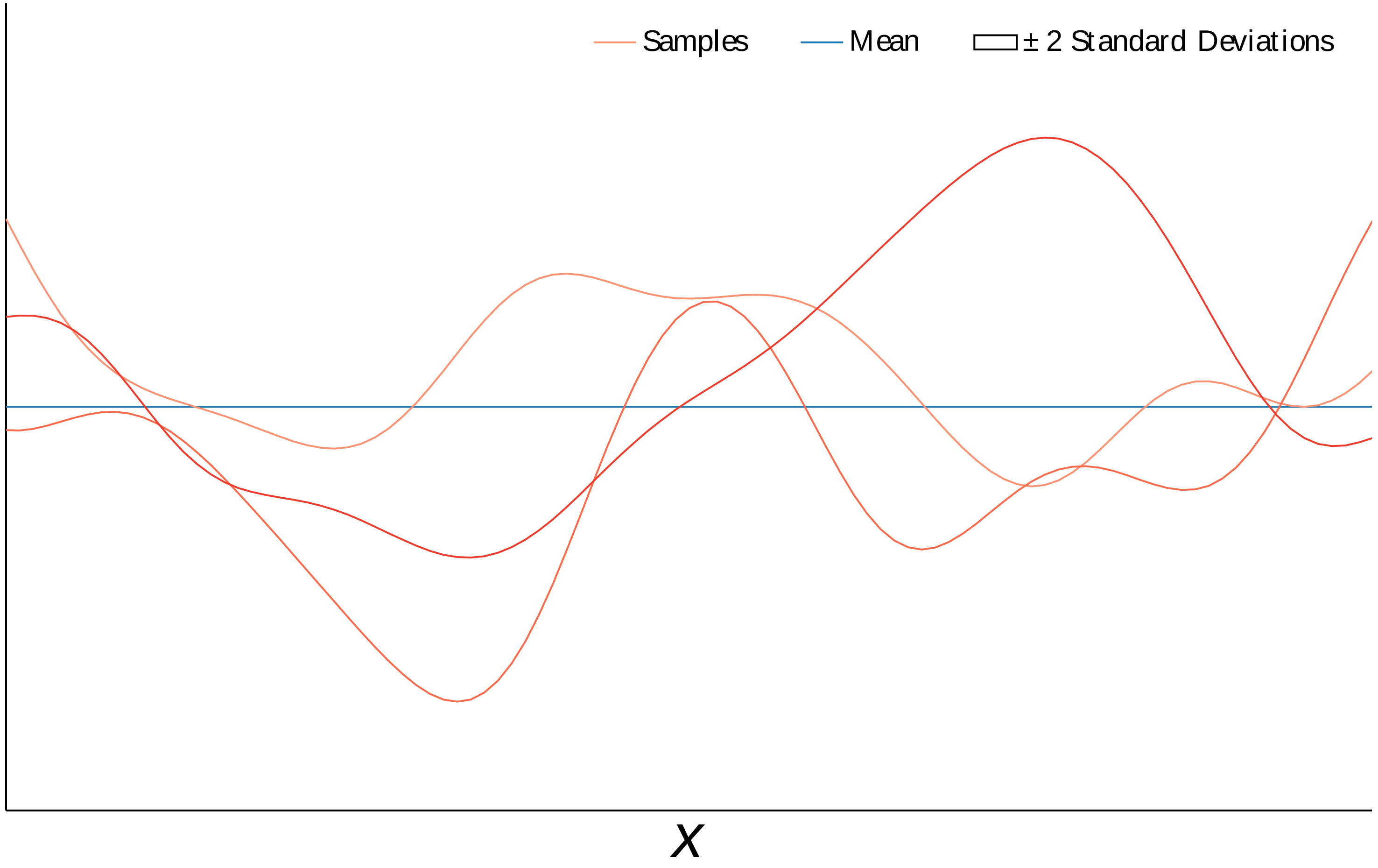
$f: \mathbb{R}^p \mapsto \mathbb{R} \sim \mathcal{GP}(f; \mu(x) \quad , K(x, x') \quad))$



$f \sim \mathcal{GP}(\mu, \Sigma) \rightarrow f(x) \sim \mathcal{N}\big(\mu(x), \Sigma(x, x)\big)$

Gaussian Process (GP)

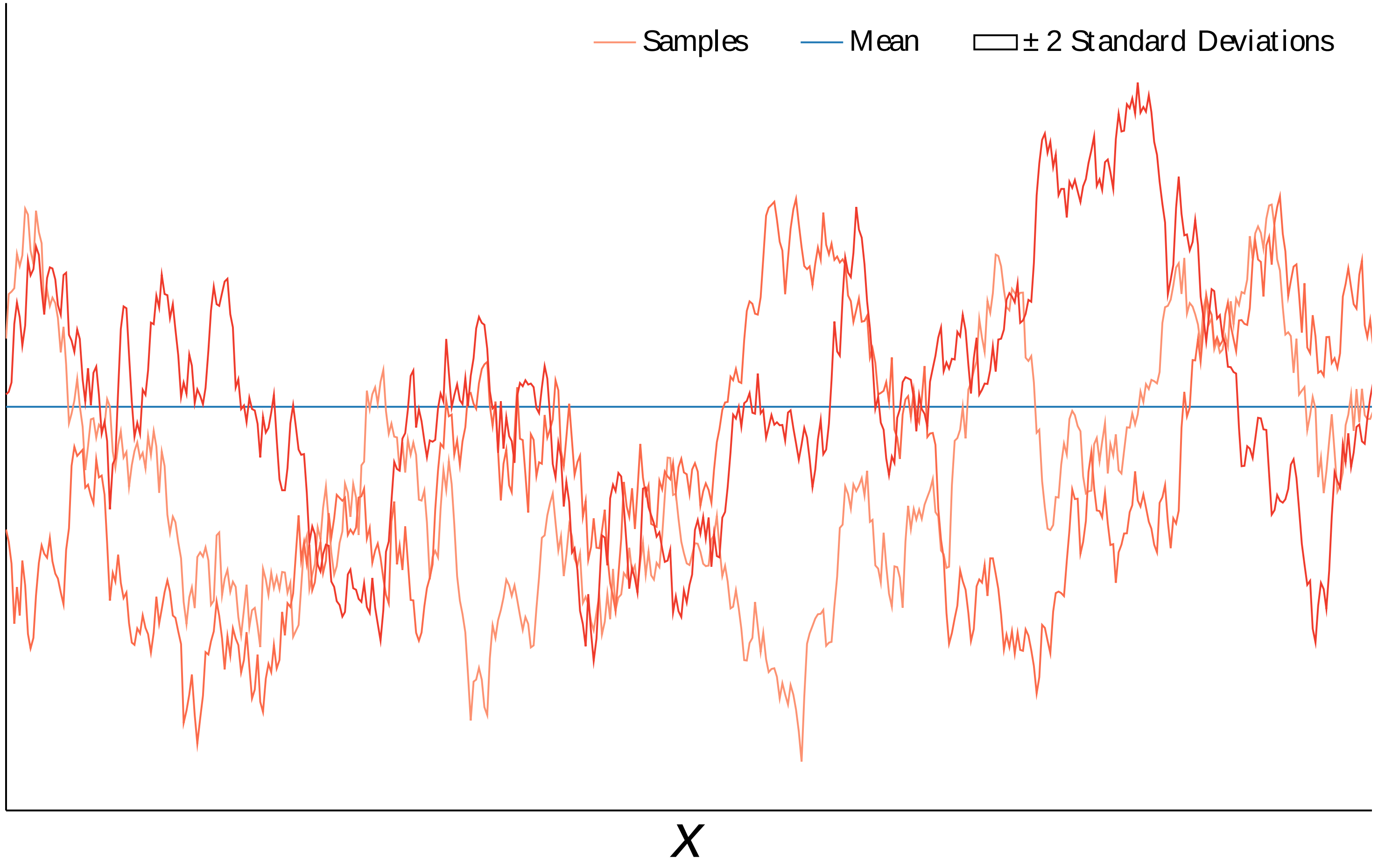
$$f : \mathbb{R}^p \mapsto \mathbb{R} \sim \mathcal{GP}\left(f; \mu(x) = 0, K(x, x') = \exp\left(-\frac{(x - x')^2}{2}\right)\right)$$



$$f \sim \mathcal{GP}(\mu, \Sigma) \rightarrow f(x) \sim \mathcal{N}\big(\mu(x), \Sigma(x, x)\big)$$

Gaussian Process (GP)

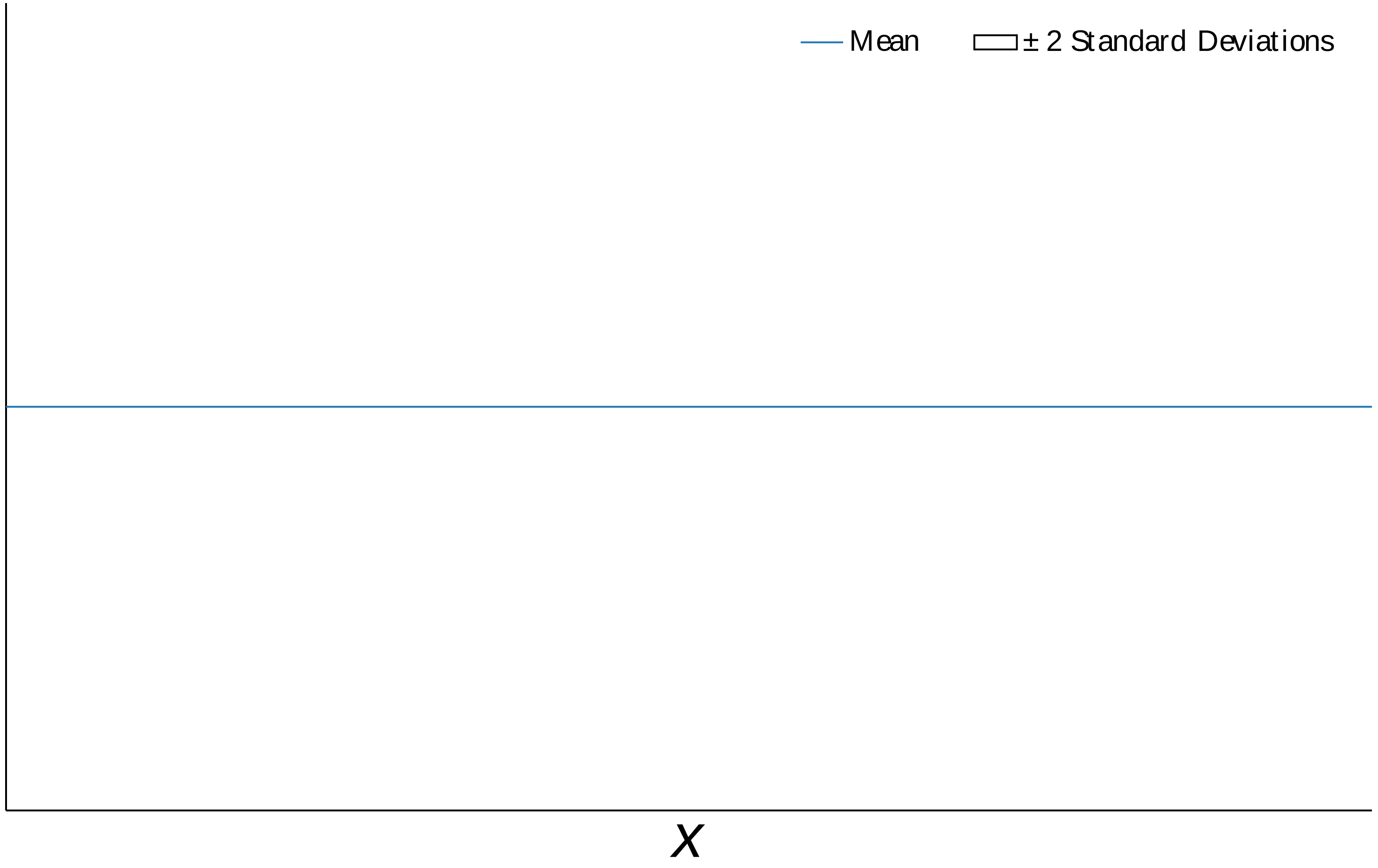
$$f : \mathbb{R}^p \mapsto \mathbb{R} \sim \mathcal{GP}(f; \mu(x) = 0, K(x, x') = \exp(-|x - x'|))$$



$$f \sim \mathcal{GP}(\mu, \Sigma) \rightarrow f(x) \sim \mathcal{N}\big(\mu(x), \Sigma(x, x)\big)$$

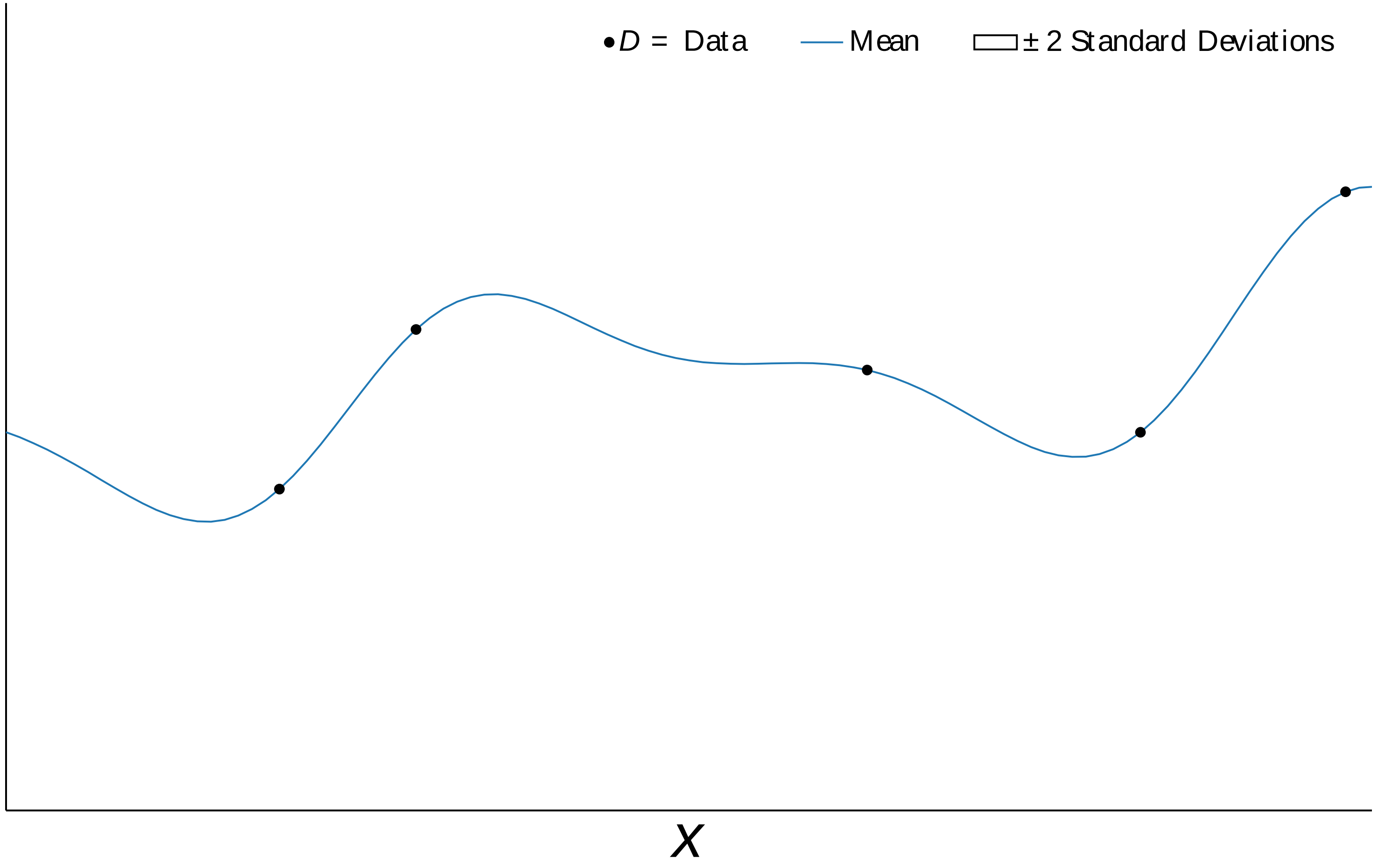
GP Prior

$$f: \mathbb{R}^p \mapsto \mathbb{R} \sim \mathcal{GP}\left(f; \mu(x) = 0, K(x, x') = \exp\left(-\frac{(x - x')^2}{2}\right)\right)$$



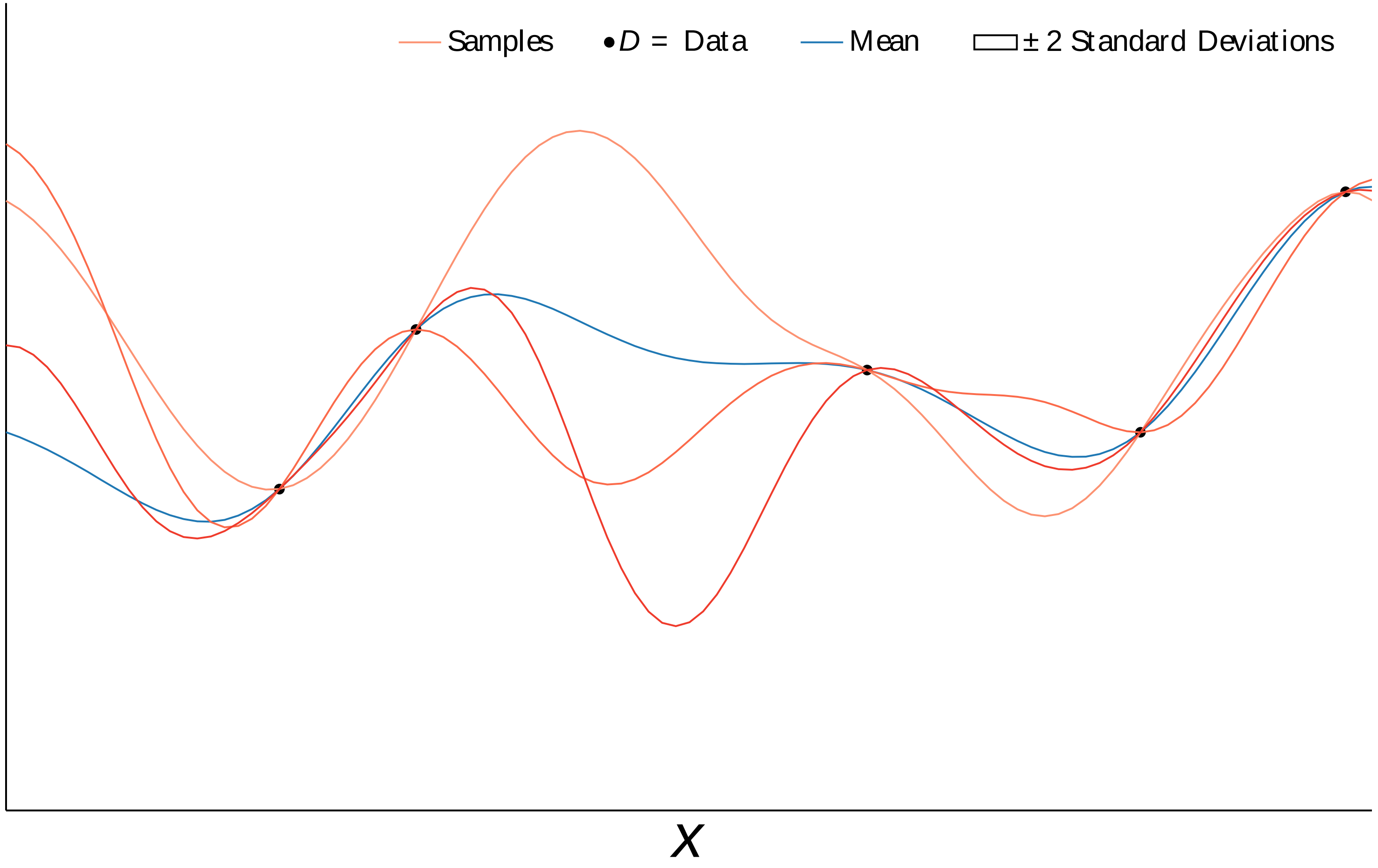
GP Posterior

$$f \mid \mathcal{D} \sim \mathcal{GP}(f; \mu_{\mathcal{D}}, K_{\mathcal{D}})$$



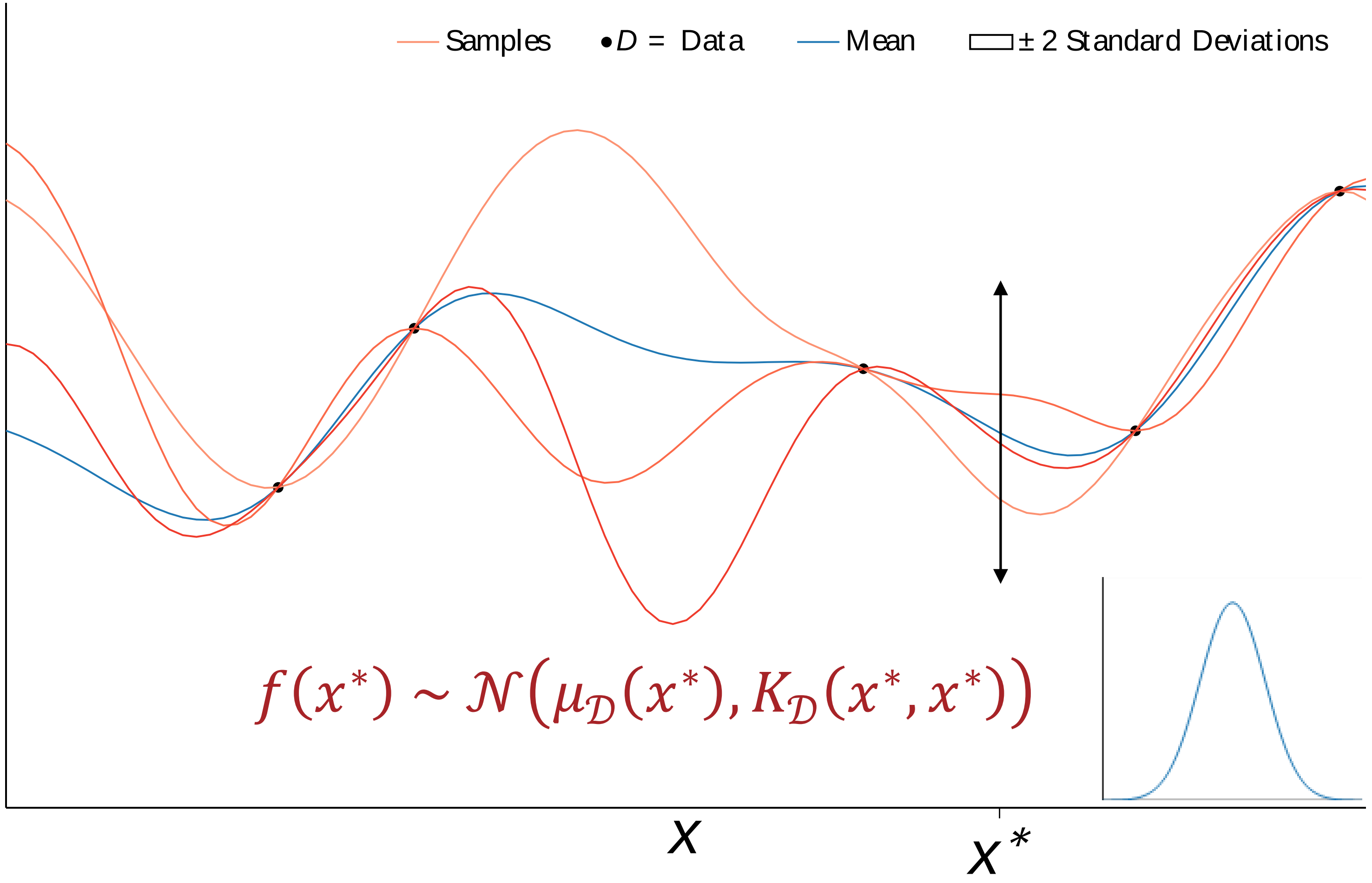
GP Posterior

$$f \mid \mathcal{D} \sim \mathcal{GP}(f; \mu_{\mathcal{D}}, K_{\mathcal{D}})$$

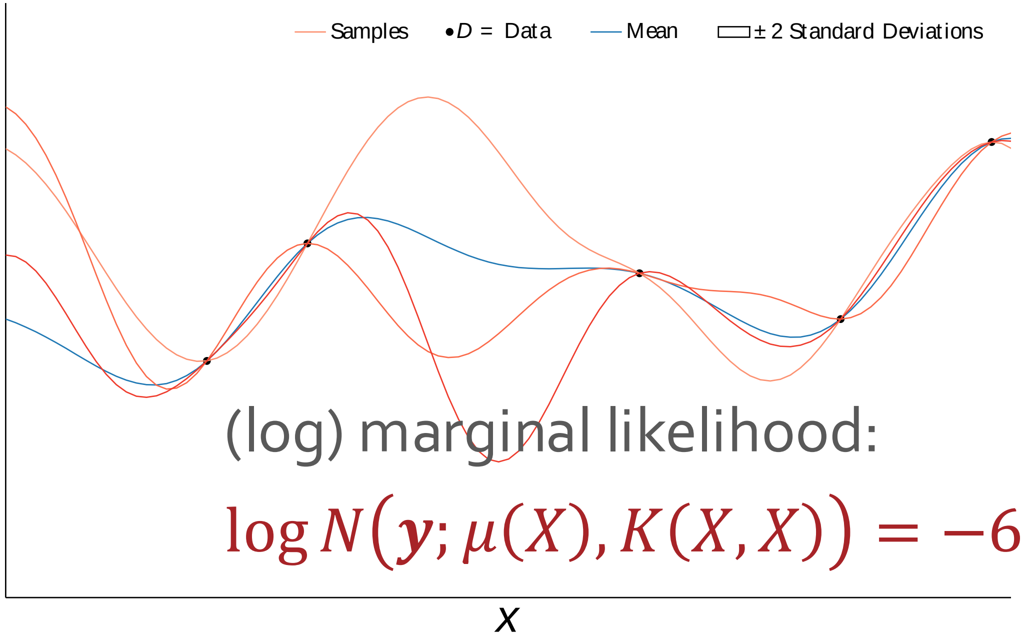


GP Posterior

$$f \mid \mathcal{D} \sim \mathcal{GP}(f; \mu_{\mathcal{D}}, K_{\mathcal{D}})$$



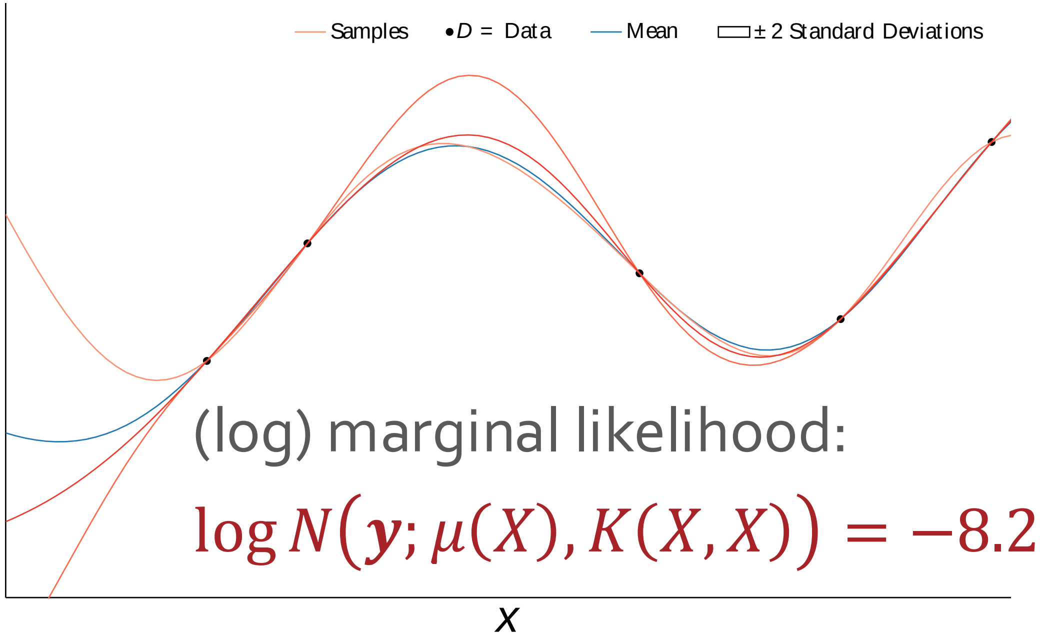
$$f(x^*) \sim \mathcal{N}\big(\mu_{\mathcal{D}}(x^*), K_{\mathcal{D}}(x^*, x^*)\big)$$



(log) marginal likelihood:

$$\log N\left(\boldsymbol{y}; \mu(X), K(X, X)\right) = -6.82$$

$$f \sim \mathcal{GP}\left(f;\ 0, (1^2) \exp\left(-\frac{(x - x')^2}{2(1^2)}\right)\right)$$

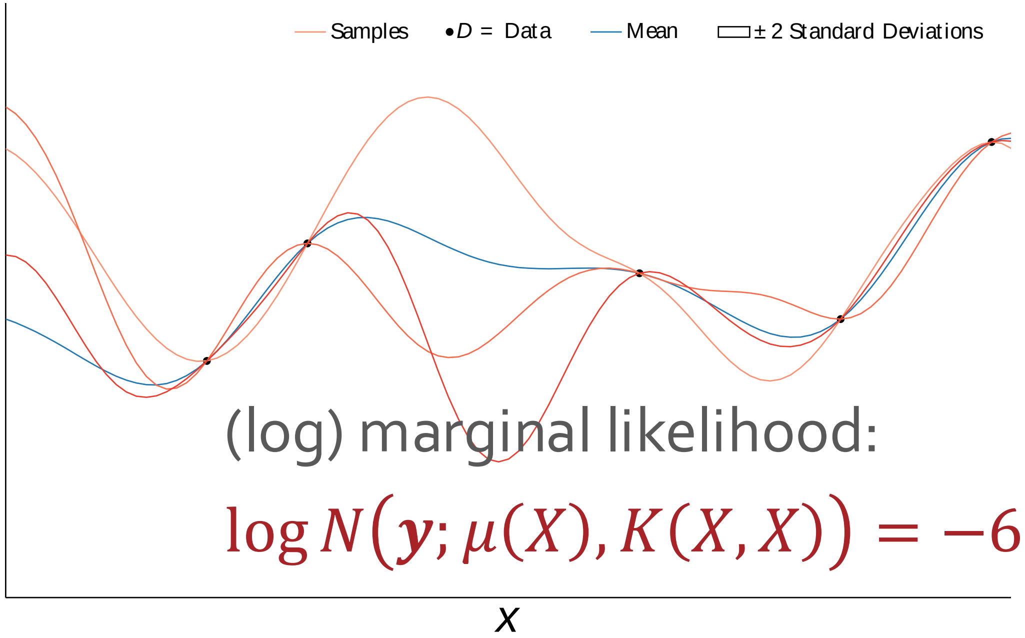


(log) marginal likelihood:

$$\log N\left(\boldsymbol{y}; \mu(X), K(X, X)\right) = -8.26$$

$$f \sim \mathcal{GP}\left(f;\ 0, (2^2) \exp\left(-\frac{(x - x')^2}{2(2^2)}\right)\right)$$

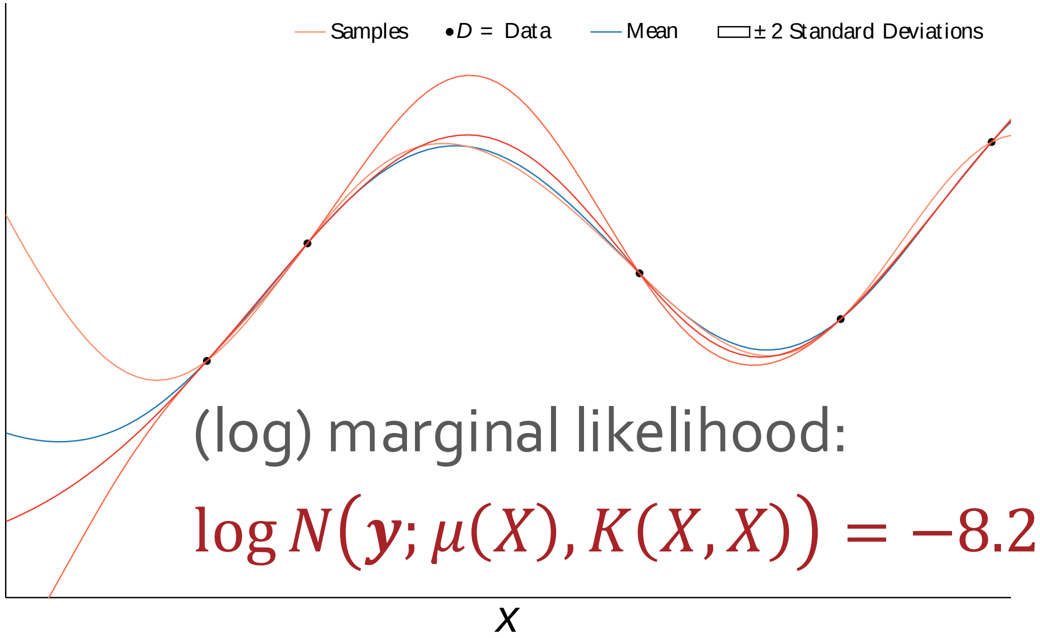
Kernel Hyperparameters



(log) marginal likelihood:

$$\log N\big(\boldsymbol{y}; \mu(X), K(X,X)\big) = -6.82$$

$$f \sim \mathcal{GP}\left(f;\ 0, (1^2)\exp\left(-\frac{(x-x')^2}{2(1^2)}\right)\right)$$

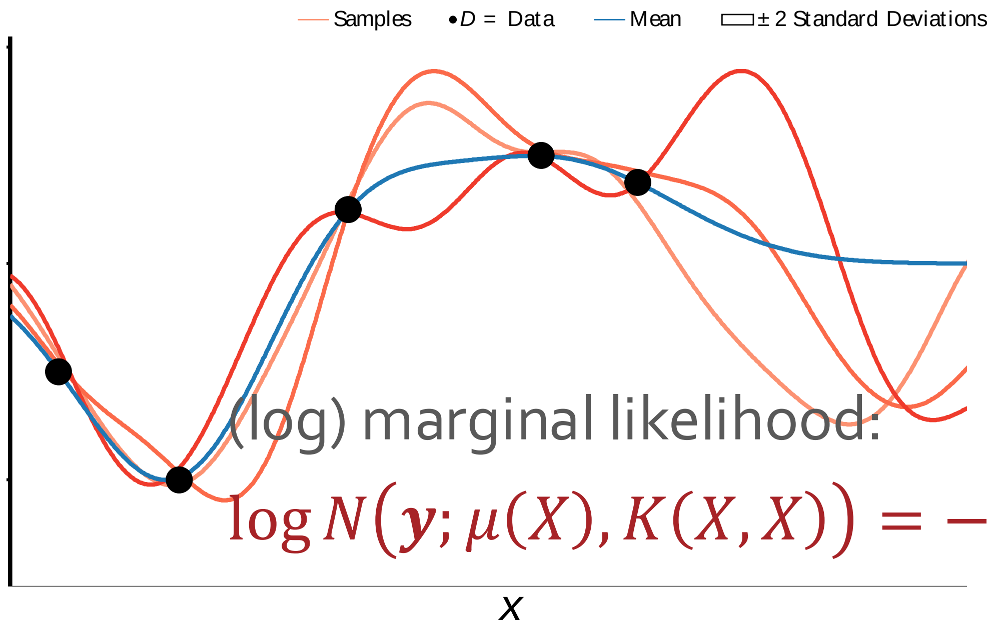


(log) marginal likelihood:

$$\log N\big(\boldsymbol{y}; \mu(X), K(X,X)\big) = -8.26$$

$$f \sim \mathcal{GP}\left(f;\ 0, (2^2)\exp\left(-\frac{(x-x')^2}{2(2^2)}\right)\right)$$

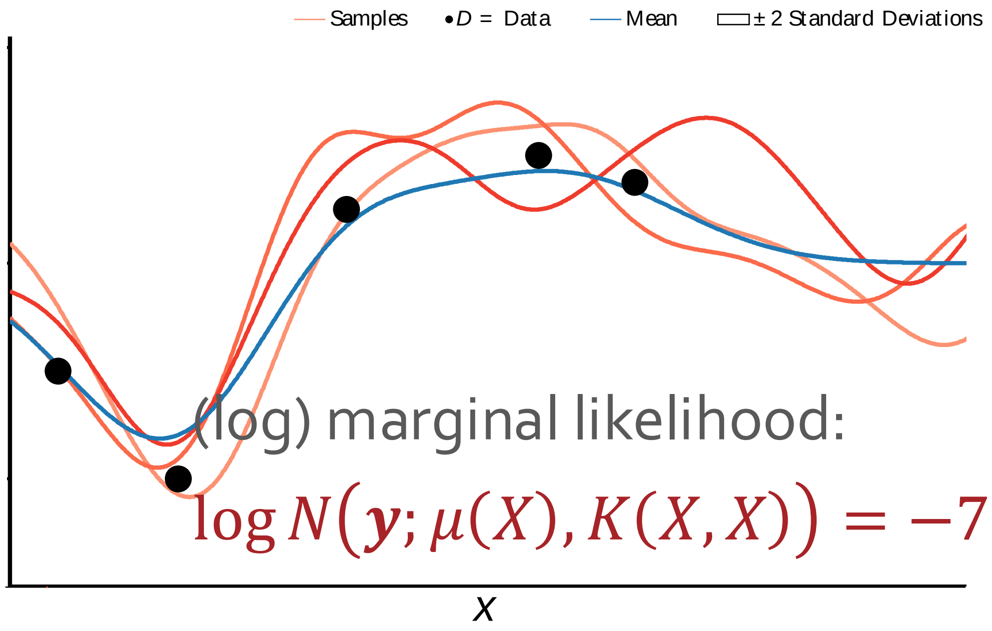
Wait, does this model always have zero training error rate???



(log) marginal likelihood:

$\log N\left(\boldsymbol{y}; \mu(X), K(X,X)\right) = -5.11$

$\sigma^2 = 0.1$



(log) marginal likelihood:

$\log N\left(\boldsymbol{y}; \mu(X), K(X,X)\right) = -7.84$

$\sigma^2 = 0.5$

Noise