Tutoring Session 8

Gaussian Processes

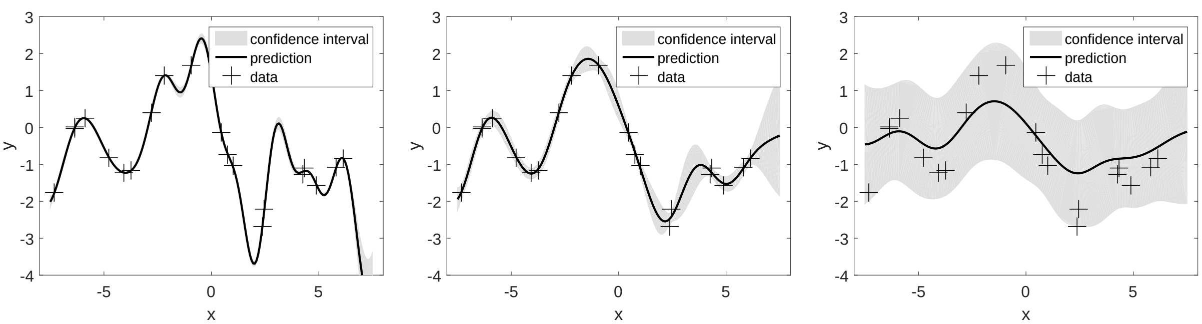
1 Simple Regression

We have a data set $\boldsymbol{X} \in \mathbb{R}^1$. You are given Gaussian processes $f \sim \mathcal{GP}(m, K)$ with mean function $m(x) = 0$, covariance function $K(x, x')$, and a noisy observation $\epsilon \sim \mathcal{N}(0, \sigma_y^2)$.

Problem 1: Assume we use $K(x, x') = (xx' + 1)^2$ as our covariance function and furthermore have observations $x_1 = -\frac{1}{2}, x_2 = 2$. Write down the distribution of $p(f(x_1), f(x_2))$. What is the relationship of $f(x_1)$ and $f(x_2)$?

Problem 2: Now lets assume we have values $y_1 = 4, y_2 = -1$ and unknown value f_* for $x_* = 1$. We also set $\sigma_y^2 = 1$, What is the conditional distribution $\boldsymbol{f}_*|\boldsymbol{y}, \boldsymbol{X}, x_*$

Problem 3: We have a squared exponential kernel. With different values of σ_y^2, the GP models are shown in the figures below. Which model is best? What causes the other two to be not good? Explain your answer.



2 Weight Space and Function Space GPs are Equivalent

(Using the notation of Rasmussen (Gaussian Processes for Machine Learning (Rasmussen, Williams), (free download at www.gaussianprocess.org)).

Assume we have a dataset $\{(x_i, y_i)\}_{I=1}^n$ with $x_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}^1$, and a feature map $\phi : \mathbb{R}^D \to \mathbb{R}^N$. We can then form a $D \times n$ pattern matrix X and a $N \times n$ design matrix ϕ.

From Bayesian Linear Regression we know that the full Bayesian approach predictive distribution for the function value f_* of some new x_* is given by

$$f_*|x_*, X, y \sim \mathcal{N}(\sigma_n^{-2}\phi(x_*)^T A^{-1}\phi y, \phi(x_*)^T A^{-1}\phi(x_*))$$

with

$$A = \sigma_n^{-2}\phi\phi^T + \Sigma_p^{-1}$$

where σ_n is from $y = f + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma_n)$ and Σ_p is from the w-prior $w \sim \mathcal{N}(0, \Sigma_p)$.

Problem 4: Kernelize this expression by rearranging it so that any dependence on X or x_* is in terms of $\phi(\ldots)^T\Sigma_p^{-1}\phi(\ldots)$ and replacing $\phi(x)^T\Sigma_p^{-1}\phi(x') = \phi(x)^T\Sigma_p^{-T/2}\Sigma_p^{-1/2}\phi(x') = \psi(x)^T\psi(x') = k(x, x')$! Show this is equivalent to Gaussian Processes with noise!