Chapter 1

Introduction

1.1 Regression

1.2 Gaussian process models

Figure 1.1 shows a Gaussian process posterior. Typically, it's rendered with the mean and +- 2SD, but there's nothing special about mean.

1.3 Latent Variable Models

1.4 Derivation of Component Marginal Variance

In this section, we derive the posterior marginal variance and covariance of the additive components of a gp. These formulas let us plot the marginal variance of each component separately. These formulas can also be used to examine the posterior covariance between pairs of components.

Let us assume that our function $\mathbf{f}$ is a sum of two functions, $\mathbf{f}_1$ and $\mathbf{f}_2$, where $\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2$.

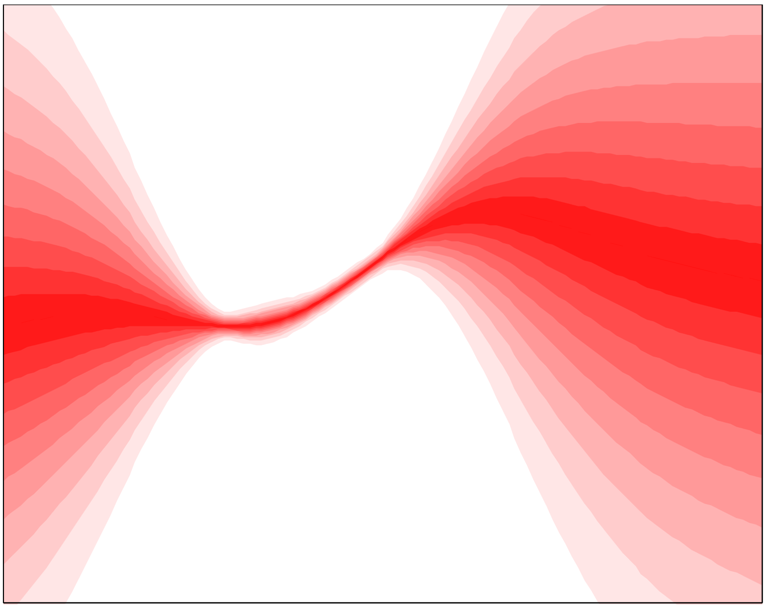

Fig. 1.1 A visual representation of a Gaussian process posterior.

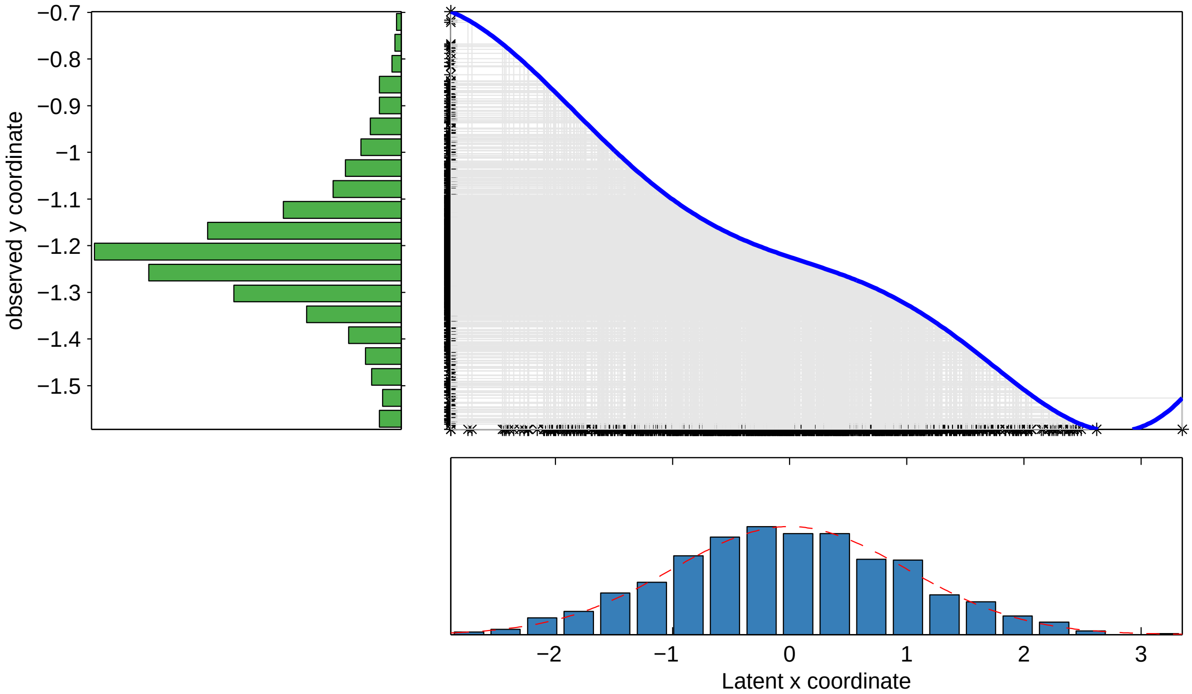

Fig. 1.2 A visual representation of the Gaussian process latent variable model. Bottom: density and samples from a 1D Gaussian, specifying the distribution $p(\mathbf{X})$ in the latent space. Top Right: A function drawn from a GP prior. Left: A nonparametric density defined by warping the latent density through the function drawn from a GP prior.

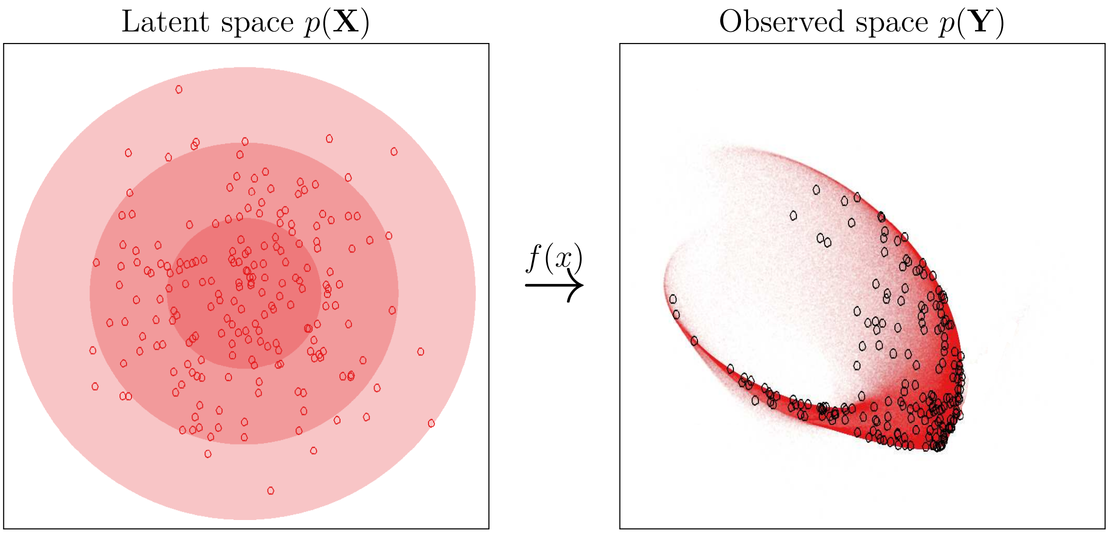

Fig. 1.3 A visual representation of the Gaussian process latent variable model. Left: Isocontours and samples from a 2D Gaussian, specifying the distribution $p(\mathbf{X})$ in the latent space. Right: Density and samples from a nonparametric density defined by warping the latent density through a function drawn from a GP prior.

If $\mathbf{f}_1$ and $\mathbf{f}_2$ are a priori independent, and $\mathbf{f}_1 \sim \mathrm{gp}(\mu_1, k_1)$ and $\mathbf{f}_2 \sim \mathrm{gp}(\mu_2, k_2)$, then

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1^\star \\ \mathbf{f}_2 \\ \mathbf{f}_2^\star \\ \mathbf{f} \\ \mathbf{f}^\star \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_1^\star \\ \mu_2 \\ \mu_2^\star \\ \mu_1 + \mu_2 \\ \mu_1^\star + \mu_2^\star \end{bmatrix}, \begin{bmatrix} \mathbf{k}_1 & \mathbf{k}_1^\star & 0 & 0 & \mathbf{k}_1 & \mathbf{k}_1^\star \\ \mathbf{k}_1^\star & \mathbf{k}_1^{\star\star} & 0 & 0 & \mathbf{k}_1^\star & \mathbf{k}_1^{\star\star} \\ 0 & 0 & \mathbf{k}_2 & \mathbf{k}_2^\star & \mathbf{k}_2 & \mathbf{k}_2^\star \\ 0 & 0 & \mathbf{k}_2^\star & \mathbf{k}_2^{\star\star} & \mathbf{k}_2^\star & \mathbf{k}_2^{\star\star} \\ \mathbf{k}_1 & \mathbf{k}_1^\star & \mathbf{k}_2 & \mathbf{k}_2^\star & \mathbf{k}_1 + \mathbf{k}_2 & \mathbf{k}_1^\star + \mathbf{k}_2^\star \\ \mathbf{k}_1^\star & \mathbf{k}_1^{\star\star} & \mathbf{k}_2^\star & \mathbf{k}_2^{\star\star} & \mathbf{k}_1^\star + \mathbf{k}_2^\star & \mathbf{k}_1^{\star\star} + \mathbf{k}_2^{\star\star} \end{bmatrix}\right) \tag{1.1}$$

where $\mathbf{k}_1 = k_1(\mathbf{X}, \mathbf{X})$ and $\mathbf{k}_1^\star = k_1(\mathbf{X}^\star, \mathbf{X})$.

By the formula for Gaussian conditionals:

$$\mathbf{x}_A | \mathbf{x}B \sim \mathcal{N}(\mu_A + \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}(\mathbf{x}_B - \mu_B), \mathbf{\Sigma}_{AA} - \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}\mathbf{\Sigma}_{BA}), \tag{1.2}$$

we get that the conditional variance of a Gaussian conditioned on its sum with another Gaussian is given by

$$\mathbf{f}_1(\mathbf{x}^\star) | \mathbf{f}(\mathbf{x}) \sim \mathcal{N}\Big(\mu_1^\star + \mathbf{k}_1^{\star\mathsf{T}}(\mathbf{K}_1 + \mathbf{K}_2)^{-1}(\mathbf{f} - \mu_1 - \mu_2), \tag{1.3}$$

$$\mathbf{k}_1^{\star\star} - \mathbf{k}_1^{\star\mathsf{T}}(\mathbf{K}_1 + \mathbf{K}_2)^{-1}\mathbf{k}_1^\star\Big). \tag{1.4}$$

The covariance between the two components, conditioned on their sum is given by:

$$\mathrm{cov}\left[\mathbf{f}_1(\mathbf{x}^\star), \mathbf{f}_2(\mathbf{x}^\star) | \mathbf{f}(\mathbf{x})\right] = -\mathbf{k}_1(\mathbf{x}^\star, \mathbf{x})\left[\mathbf{K}_1(\mathbf{x}, \mathbf{x}) + \mathbf{K}_2(\mathbf{x}, \mathbf{x})\right]^{-1}\mathbf{k}_1(\mathbf{x}, \mathbf{x}^\star) \tag{1.5}$$

These formulae express the posterior model uncertainty about different components of the signal, integrating over the possible configurations of the other components.

1.5 Covariance functions

Kernels specify similarity between function values of two objects, not between similarity of objects