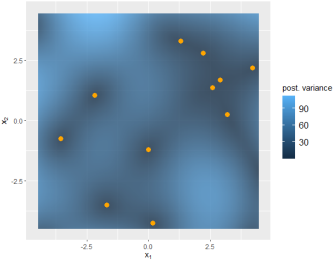
Introduction to Machine Learning

Mean Functions for Gaussian Processes



Learning goals

- XXX
- XXX

THE ROLE OF MEAN FUNCTIONS

- It is common but by no means necessary to consider GPs with a zero-mean function

$$m(\mathbf{x}) \equiv 0$$

- Note that this is not necessarily a drastic limitation, since the mean of the posterior process is not confined to be zero

$$\boldsymbol{f}_*|\mathbf{X}_*, \mathbf{X}, \boldsymbol{f} \sim \mathcal{N}(\mathbf{K}_*^T\mathbf{K}^{-1}\boldsymbol{f}, \mathbf{K}_{**} - \mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{K}_*).$$

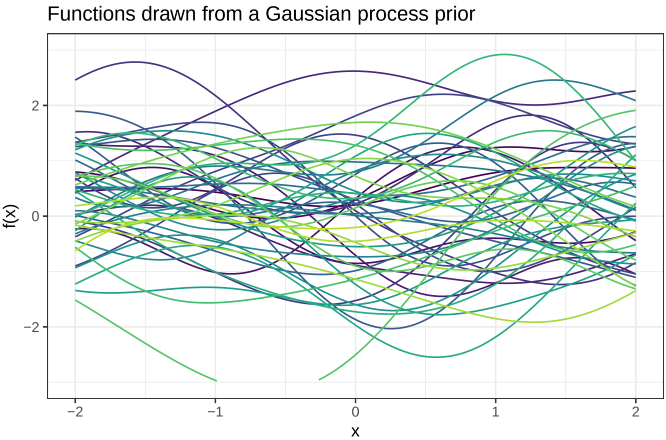
- Yet there are several reasons why one might wish to explicitly model a mean function, including interpretability, convenience of expressing prior informations, ...
- When assuming a non-zero mean GP prior $\mathcal{GP}\left(m(\boldsymbol{x}), k\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right)$ with mean $m(\mathbf{x})$, the predictive mean becomes

$$m(\mathbf{X}_*) + \mathbf{K}_*\mathbf{K}_y^{-1}\left(\boldsymbol{y} - m(\mathbf{X})\right)$$

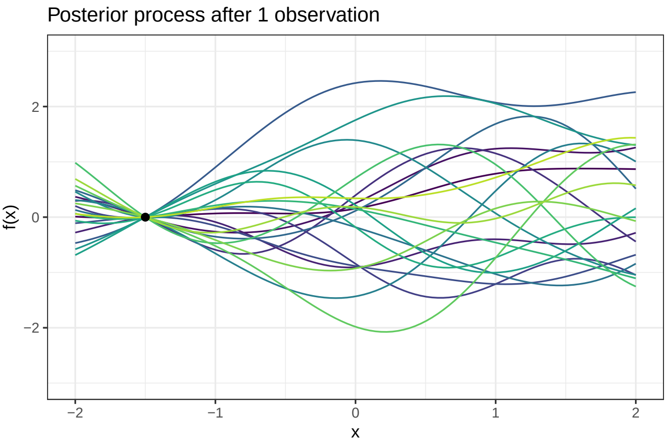
while the predictive variance remains unchanged.

THE ROLE OF MEAN FUNCTIONS

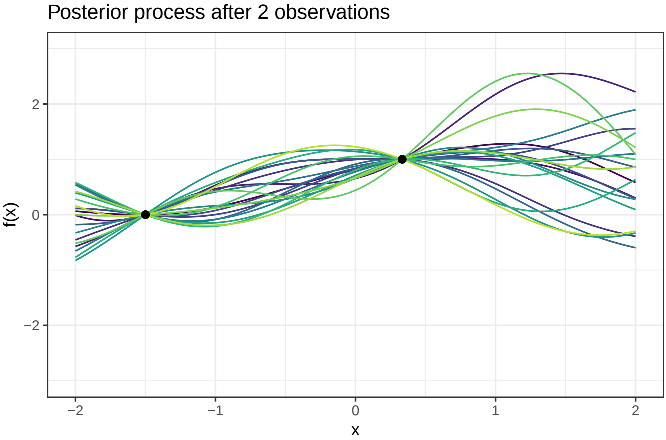
- Gaussian processes with non-zero mean Gaussian process priors are also called Gaussian processes with trend.



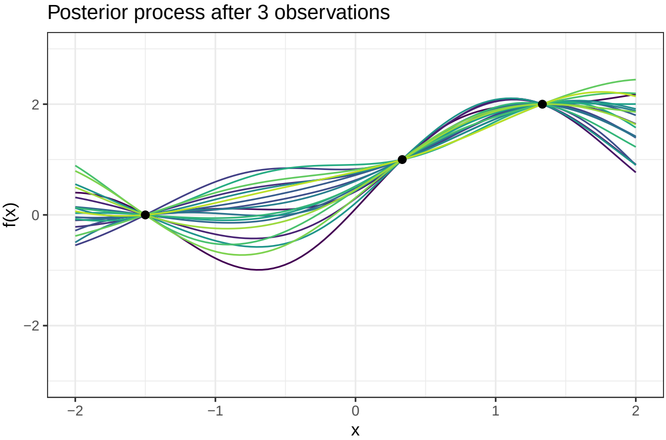
THE ROLE OF MEAN FUNCTIONS



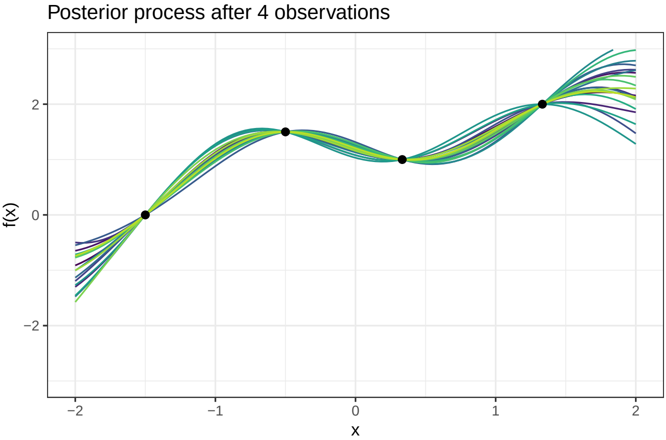
THE ROLE OF MEAN FUNCTIONS



THE ROLE OF MEAN FUNCTIONS



THE ROLE OF MEAN FUNCTIONS



THE ROLE OF MEAN FUNCTIONS

- In practice it can often be difficult to specify a fixed mean function
- In many cases it may be more convenient to specify a few fixed basis functions, whose coefficients, $\boldsymbol{\beta}$, are to be inferred from the data
- Consider

$$g(\mathbf{x}) = b(\mathbf{x})^\top \boldsymbol{\beta} + f(\mathbf{x}), \text{ where } f(\mathbf{x}) \sim \mathcal{GP}\left(0, k(\mathbf{x}, \tilde{\mathbf{x}})\right)$$

- This formulation expresses that the data is close to a global linear model with the residuals being modelled by a GP.
- For the estimation of $g(\mathbf{x})$ please refer to *Rasmussen, Gaussian Processes for Machine Learning, 2006*