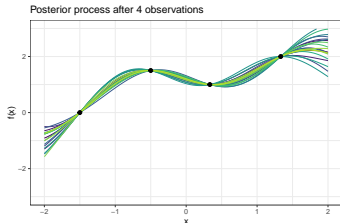
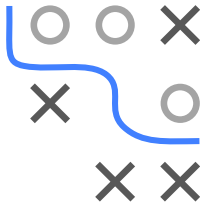


# Introduction to Machine Learning

## Gaussian Processes

### Mean functions for GPs



### Learning goals

- Trends can be modeled via specification of the mean function

# THE ROLE OF MEAN FUNCTIONS

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

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

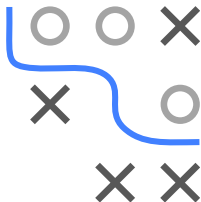
- Note that this is not necessarily a drastic limitation, since the mean of the posterior process is not confined to be zero

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

- 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  $\mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  with mean  $m(\mathbf{x})$ , the predictive mean becomes

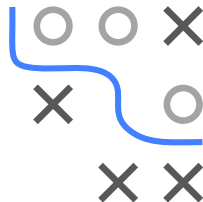
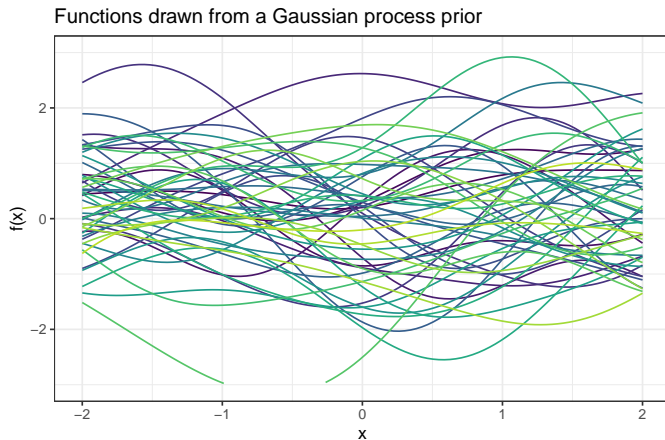
$$m(\mathbf{X}_*) + \mathbf{K}_* \mathbf{K}_y^{-1} (\mathbf{y} - m(\mathbf{X}))$$

while the predictive variance remains unchanged.

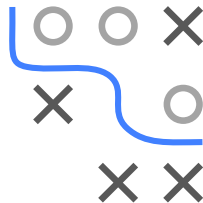
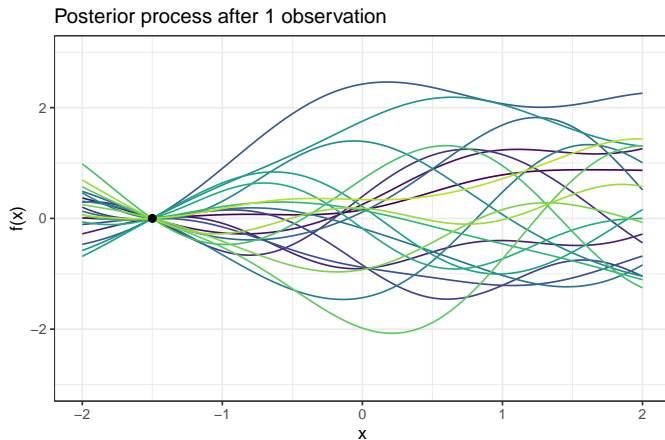


# THE ROLE OF MEAN FUNCTIONS / 2

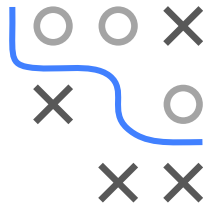
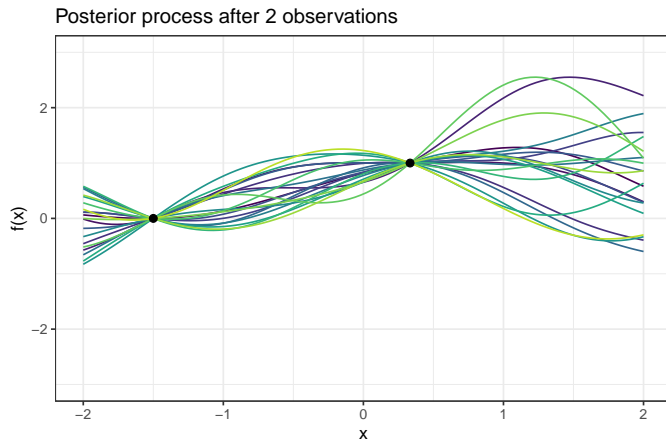
- Gaussian processes with non-zero mean Gaussian process priors are also called Gaussian processes with trend.



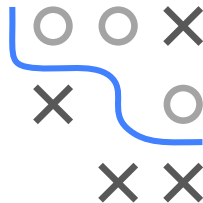
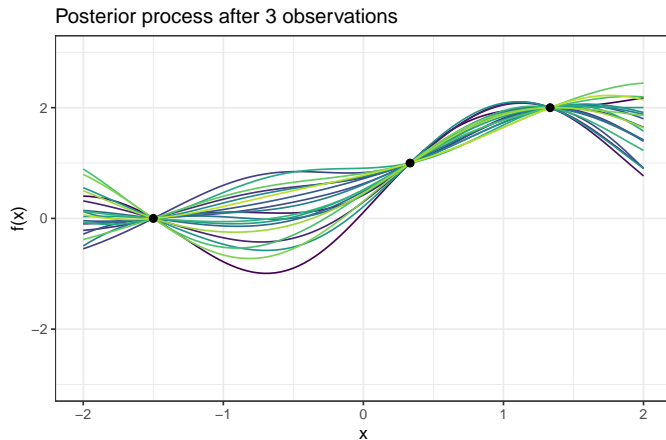
# THE ROLE OF MEAN FUNCTIONS / 3



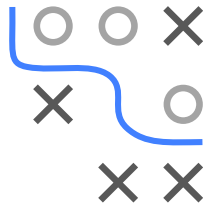
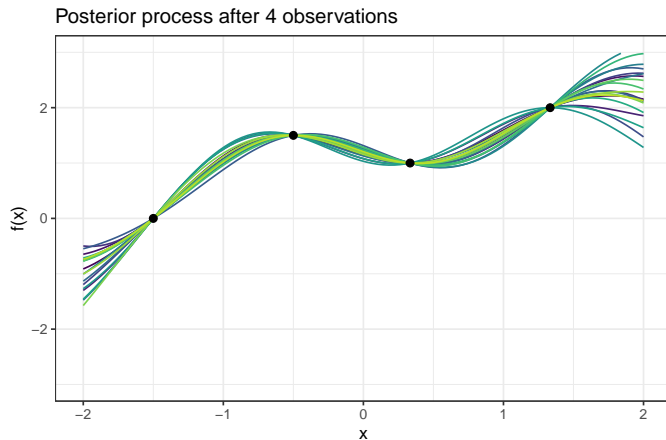
# THE ROLE OF MEAN FUNCTIONS / 4



# THE ROLE OF MEAN FUNCTIONS / 5



# THE ROLE OF MEAN FUNCTIONS / 6



# THE ROLE OF MEAN FUNCTIONS / 7

- 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,  $\beta$ , are to be inferred from the data
- Consider

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

- 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*

