

Advanced Machine Learning

Gaussian Processes: From Weight-space to Function-space



$f(x)$



$\sim \mathcal{N}(\mu, \Sigma)$

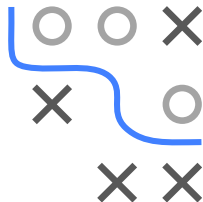
Learning goals

- The difference between weight-space and function-space views

WEIGHT-SPACE VIEW

- Until now we considered a hypothesis space \mathcal{H} of parameterized functions $f(\mathbf{x} \mid \theta)$ (in particular, the space of linear functions).
- Using Bayesian inference, we derived distributions for θ after having observed data \mathcal{D} .
- Prior beliefs about the parameter are expressed via a prior distribution $q(\theta)$, which is updated according to Bayes' rule

$$\underbrace{p(\theta|\mathbf{X}, \mathbf{y})}_{\text{posterior}} = \frac{\overbrace{p(\mathbf{y}|\mathbf{X}, \theta)}^{\text{likelihood}} \overbrace{q(\theta)}^{\text{prior}}}{\underbrace{p(\mathbf{y}|\mathbf{X})}_{\text{marginal}}}.$$



WEIGHT-SPACE VS. FUNCTION-SPACE VIEW

Weight-Space View

Parameterize functions

Example: $f(\mathbf{x} \mid \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{x}$

Define distributions on θ

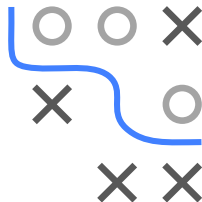
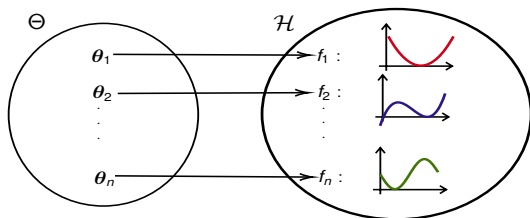
Inference in parameter space Θ

Function-Space View

Define distributions on f

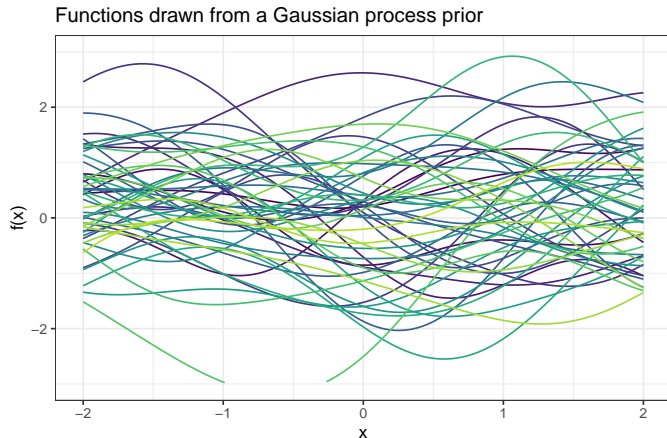
Inference in function space \mathcal{H}

- Directly search in a space of “allowed” functions \mathcal{H} .
- Specify a prior distribution **over functions** instead over a parameter and update it according to the observed data points.

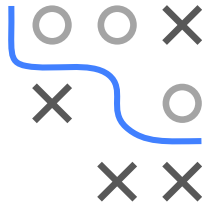


FUNCTION-SPACE VIEW

Intuitively, imagine we could draw a huge number of functions from some prior distribution over functions (*).

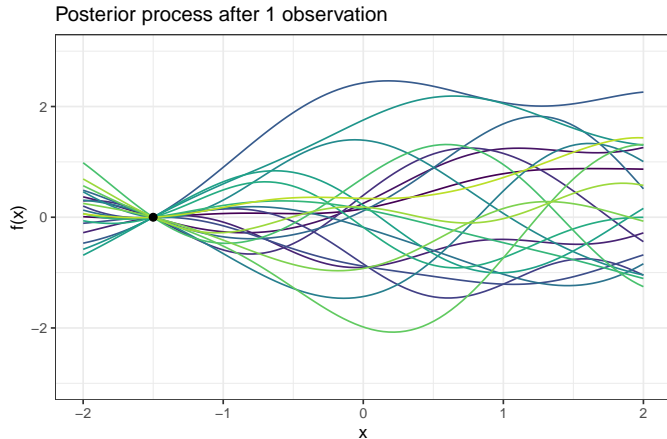


(*) We will see in a minute how distributions over functions can be specified.



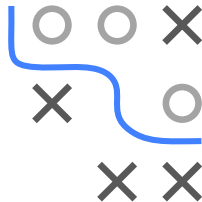
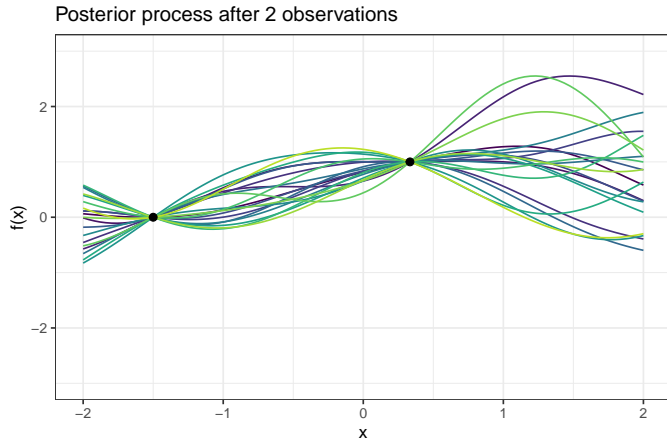
FUNCTION-SPACE VIEW

After observing some data points, we are only allowed to sample those functions, that are consistent with the data.



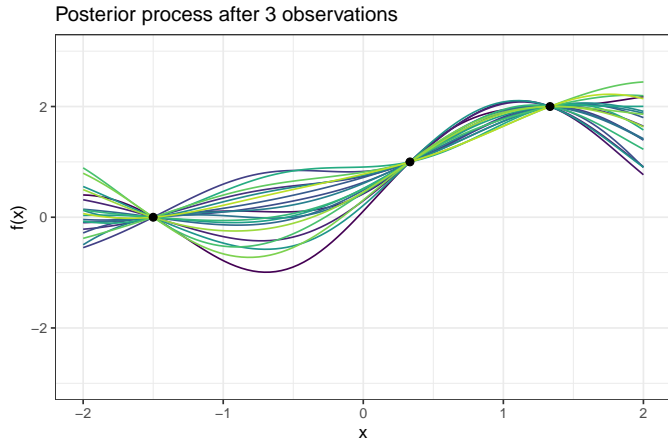
FUNCTION-SPACE VIEW

After observing some data points, we are only allowed to sample those functions, that are consistent with the data.



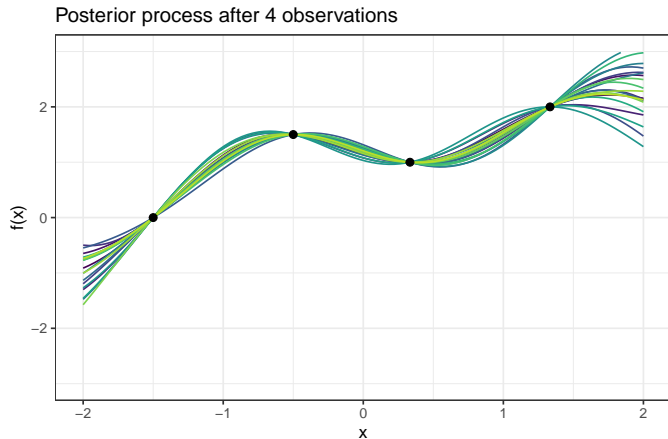
FUNCTION-SPACE VIEW

After observing some data points, we are only allowed to sample those functions, that are consistent with the data.



FUNCTION-SPACE VIEW

As we observe more and more data points, the variety of functions consistent with the data shrinks.



FUNCTION-SPACE VIEW / 2

Intuitively, there is something like “mean” and a “variance” of a distribution over functions.

