WEIGHT-SPACE VIEW

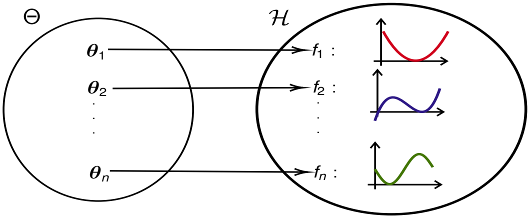
- Until now we considered a hypothesis space $\mathcal{H}$ of parameterized functions $f(\mathbf{x} \mid \boldsymbol{\theta})$ (in particular, the space of linear functions).
- Using Bayesian inference, we derived distributions for $\boldsymbol{\theta}$ after having observed data $\mathcal{D}$.
- Prior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

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

WEIGHT-SPACE VS. FUNCTION-SPACE VIEW

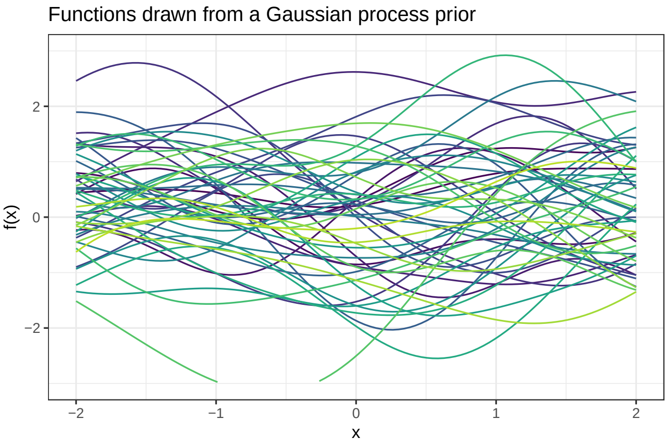
Weight-Space View	**Function-Space View**
Parameterize functions Example: $f(\mathbf{x} \mid \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{x}$	
Define distributions on $\boldsymbol{\theta}$	Define distributions on f
Inference in parameter space Θ	Inference in function space $\mathcal{H}$

- 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.

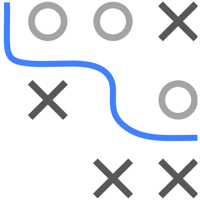


FUNCTION-SPACE VIEW

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

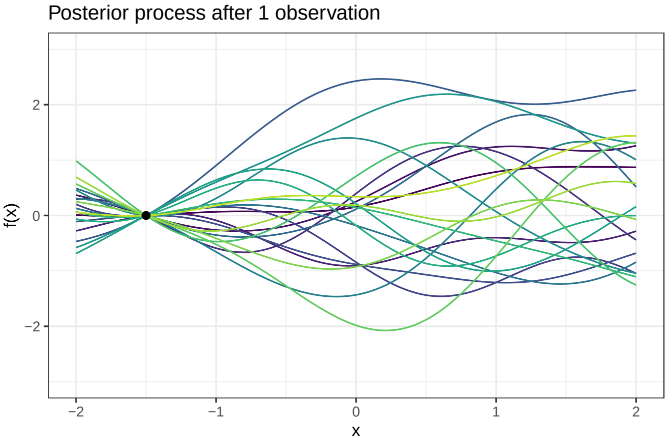


$^{(*)}$ We will see in a minute how distributions over functions can be specified.



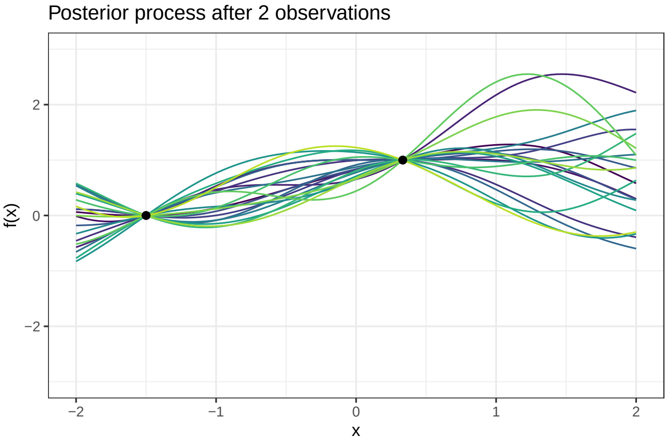
FUNCTION-SPACE VIEW

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



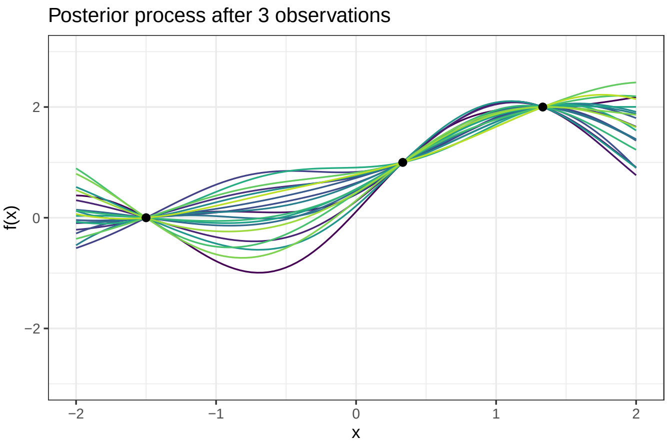
FUNCTION-SPACE VIEW

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.