Gaussian processes

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

- A common default (e.g. in software) is $m(\mathbf{x}) = 0$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

- A common default (e.g. in software) is $m(\mathbf{x}) = 0$
- One very commonly used covariance function is the *squared exponential* or *radial basis function (RBF)*

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

- A common default (e.g. in software) is $m(\mathbf{x}) = 0$
- One very commonly used covariance function is the *squared exponential* or *radial basis function (RBF)*
  - We'll see a more general form later, but for now we're using: $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2)$

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

- A common default (e.g. in software) is $m(\mathbf{x}) = 0$
- One very commonly used covariance function is the *squared exponential* or *radial basis function (RBF)*
  - We'll see a more general form later, but for now we're using: $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2)$
- For now, assume data is observed without noise

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

$$f \sim \mathcal{GP}(m, k)$$

x could be just about anything, but in this tutorial, we'll assume it's a real vector

  - Mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$
  - Covariance function (a.k.a. *kernel*)

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

- A common default (e.g. in software) is $m(\mathbf{x}) = 0$
- One very commonly used covariance function is the *squared exponential* or *radial basis function (RBF)*
  - We'll see a more general form later, but for now we're using: $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2)$
- For now, assume data is observed without noise [demo1,2]

Note on “dimension” of the input

Note on “dimension” of the input

- Let’s be careful to separate
  two types of “dimension”

Note on “dimension” of the input

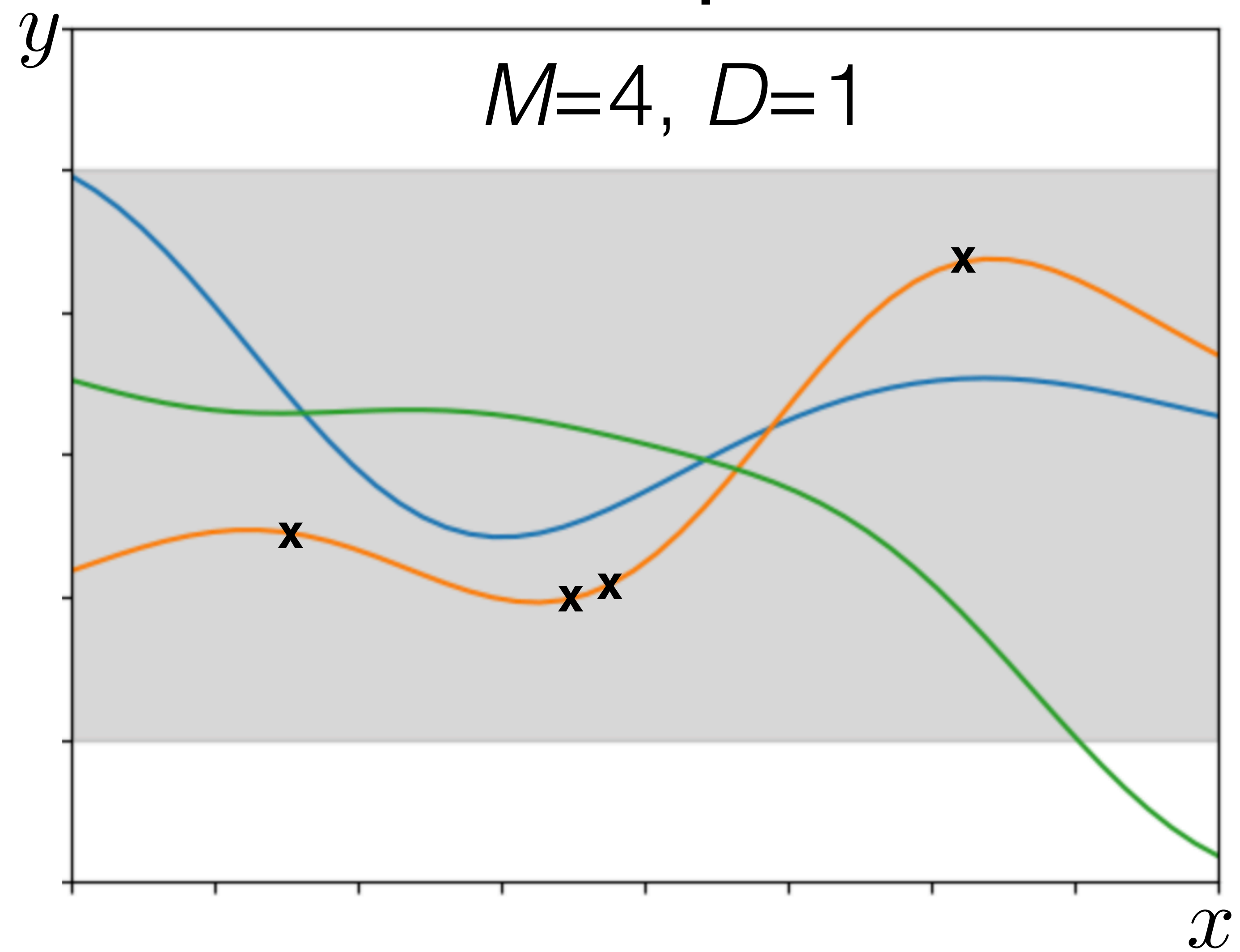
- Let’s be careful to separate two types of “dimension”
  - We’re using a superscript to denote (*M* or *N*) number of points in the space

Note on “dimension” of the input

- Let’s be careful to separate two types of “dimension”
  - We’re using a superscript to denote (M or N) number of points in the space
  - We’ll use a subscript for the (D) different elements of a point’s vector

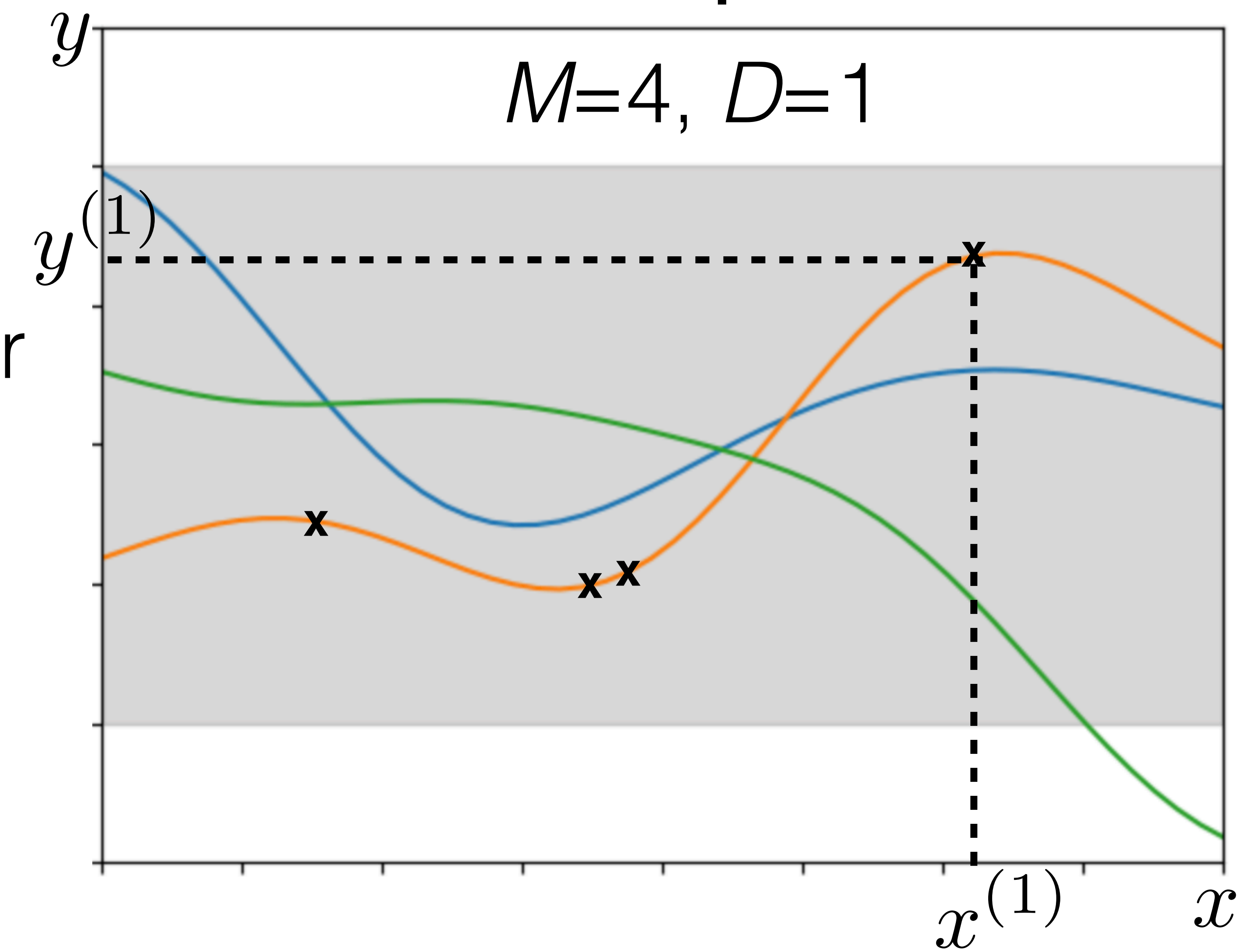
Note on "dimension" of the input

- Let's be careful to separate two types of "dimension"
  - We're using a superscript to denote (*M* or *N*) number of points in the space
  - We'll use a subscript for the (*D*) different elements of a point's vector



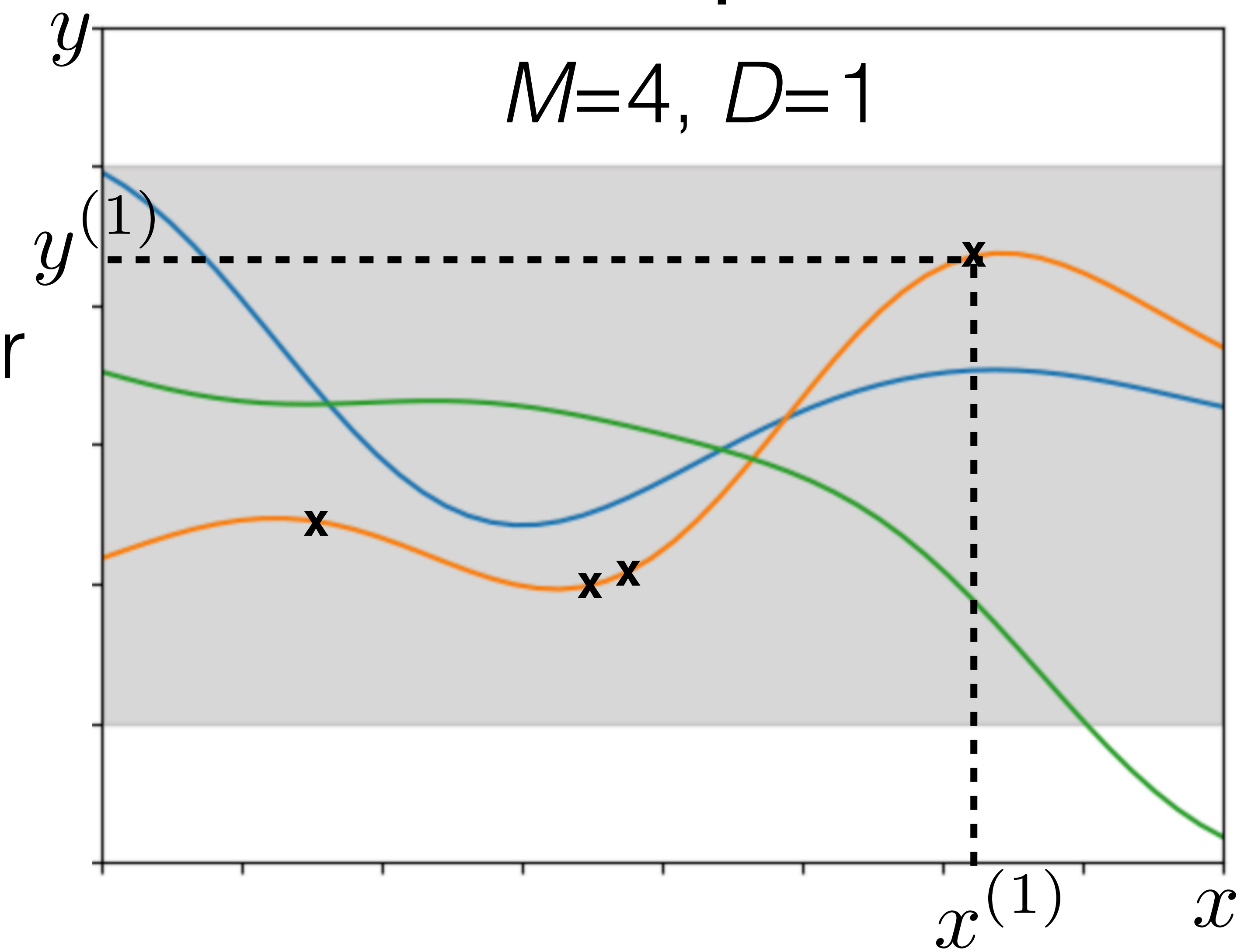
Note on “dimension” of the input

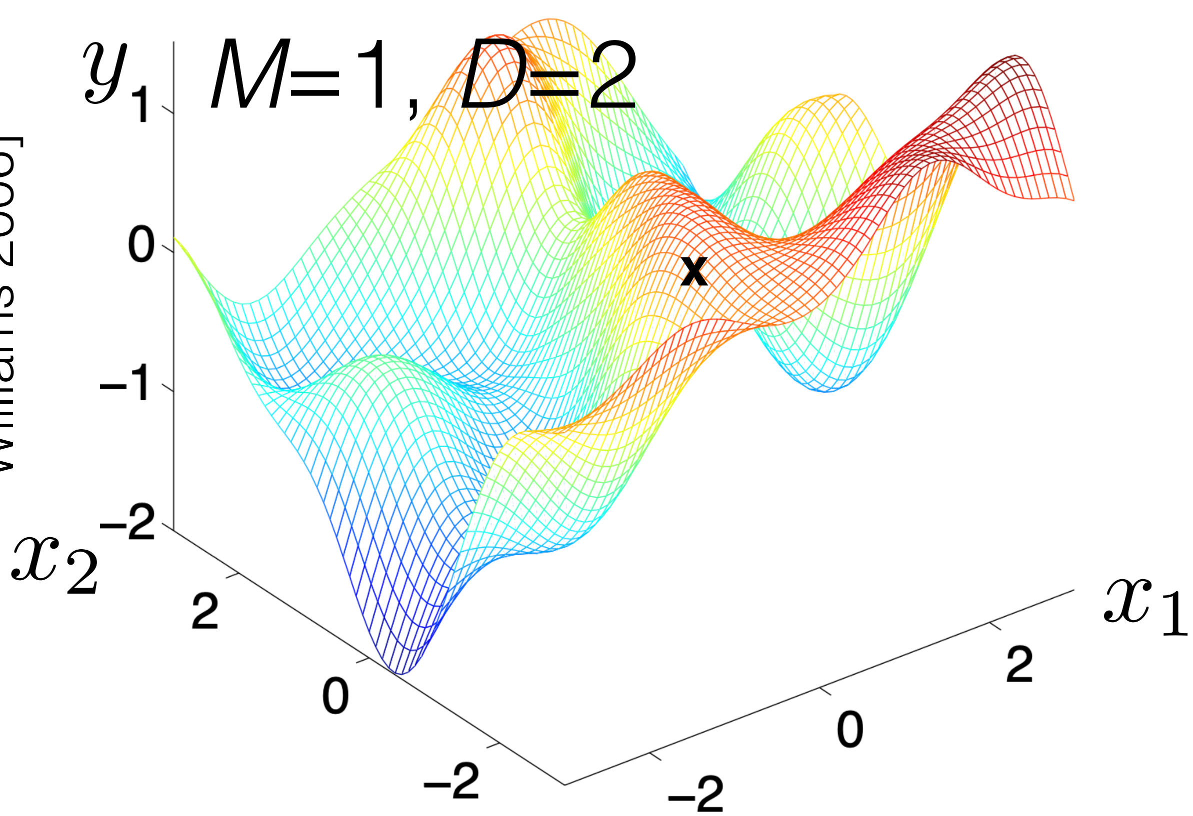
- Let’s be careful to separate two types of “dimension”
  - We’re using a superscript to denote (*M* or *N*) number of points in the space
  - We’ll use a subscript for the (*D*) different elements of a point’s vector



Note on "dimension" of the input

- Let's be careful to separate two types of "dimension"
  - We're using a superscript to denote (*M* or *N*) number of points in the space
  - We'll use a subscript for the (*D*) different elements of a point's vector

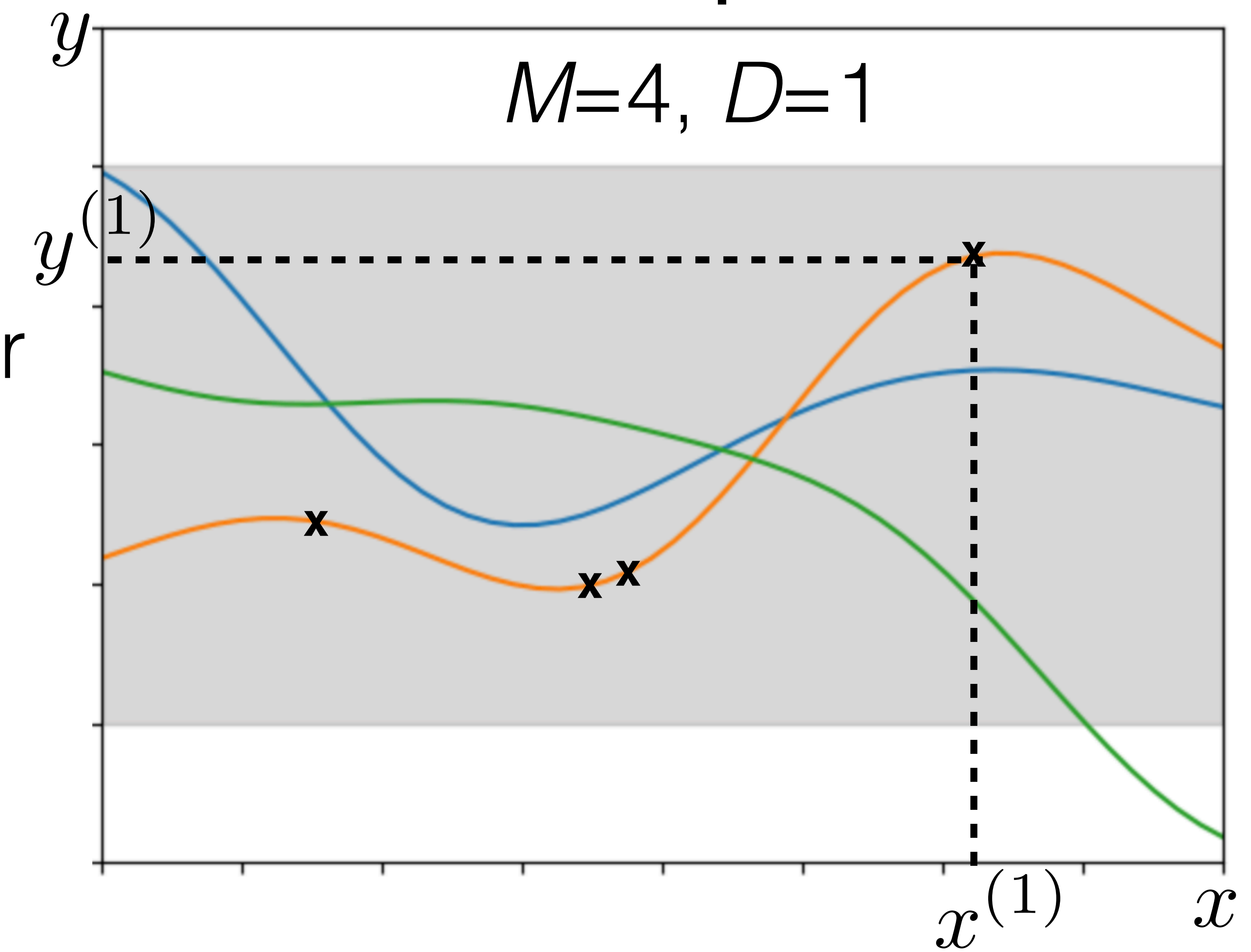


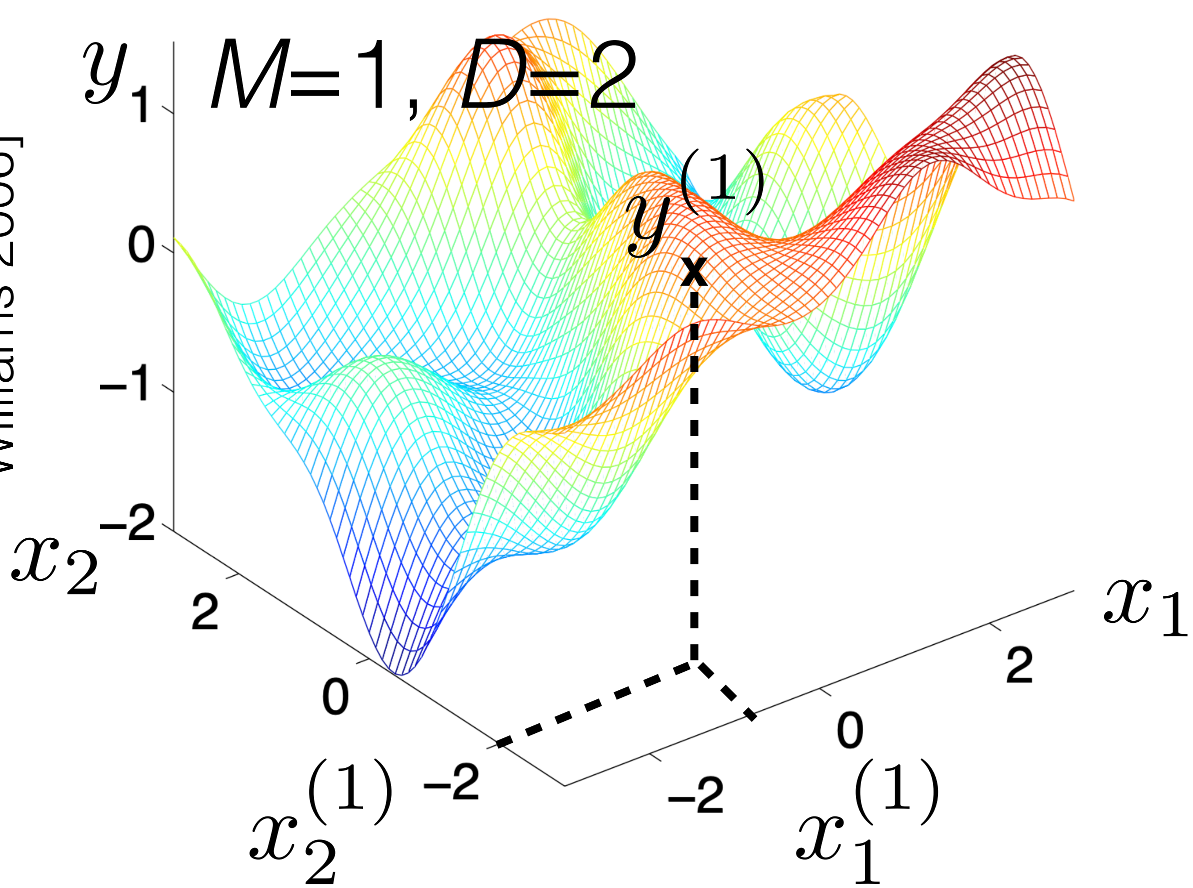


[Rasmussen & Williams 2006]

Note on "dimension" of the input

- Let's be careful to separate two types of "dimension"
  - We're using a superscript to denote (*M* or *N*) number of points in the space
  - We'll use a subscript for the (*D*) different elements of a point's vector

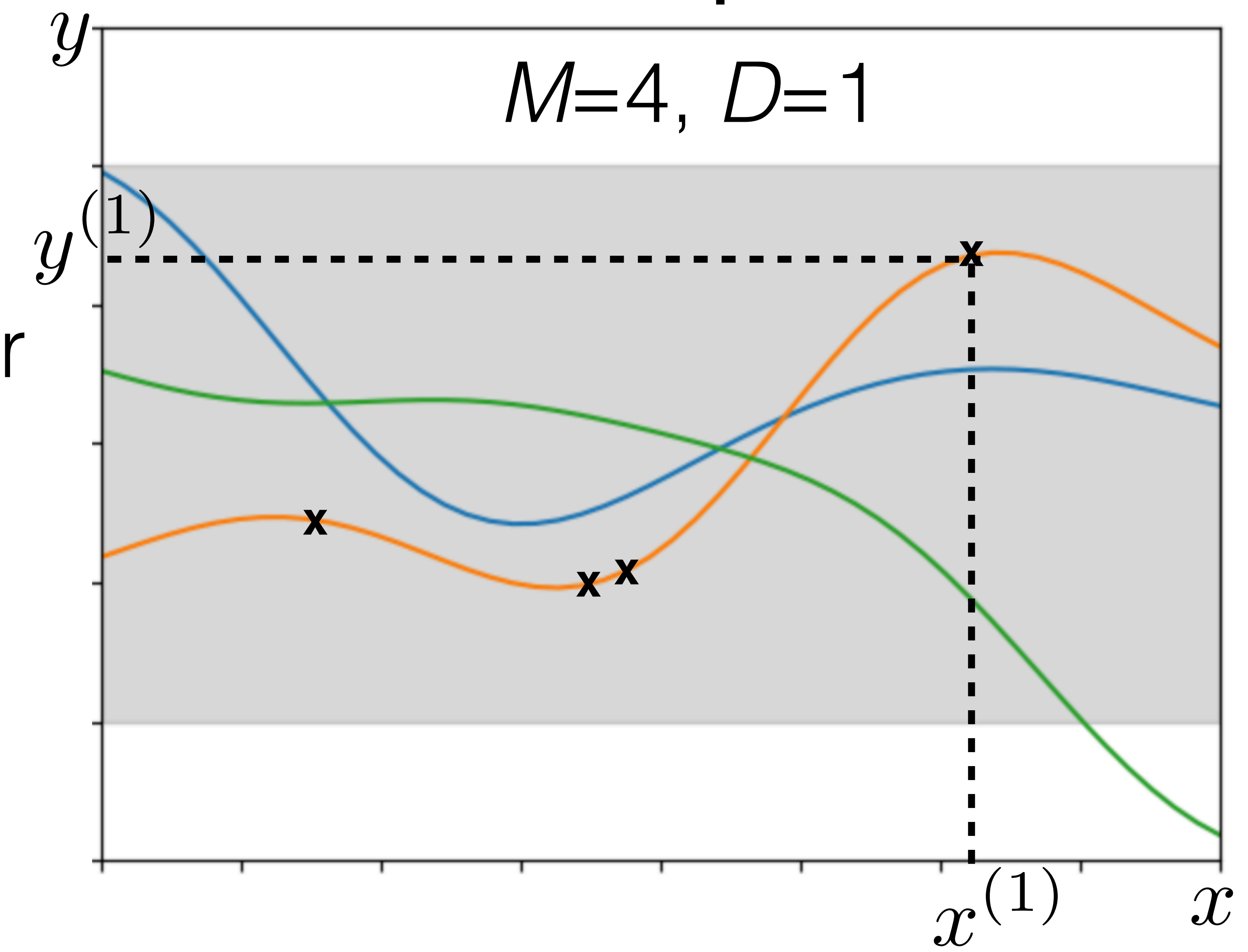


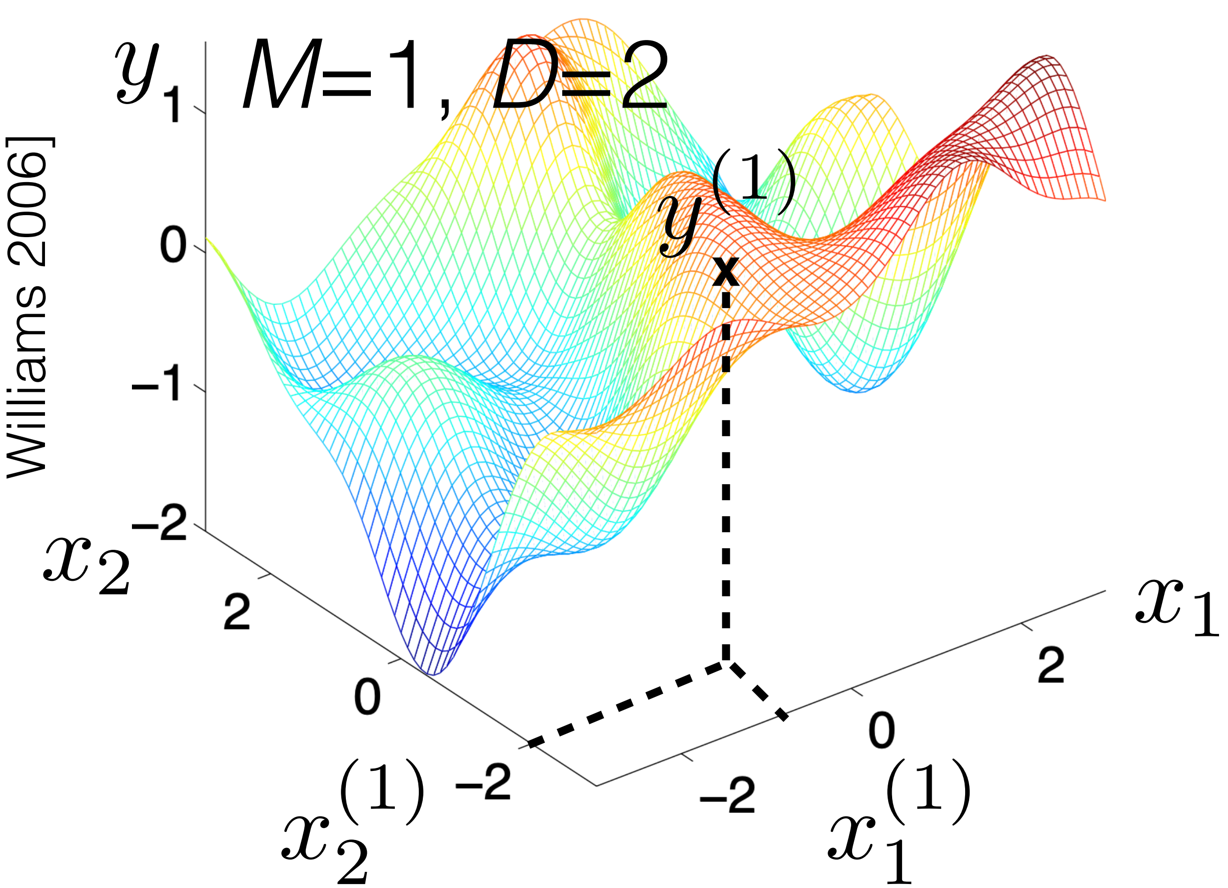


[Rasmussen & Williams 2006]

Note on “dimension” of the input

- Let’s be careful to separate two types of “dimension”
  - We’re using a superscript to denote (*M* or *N*) number of points in the space
  - We’ll use a subscript for the (*D*) different elements of a point’s vector

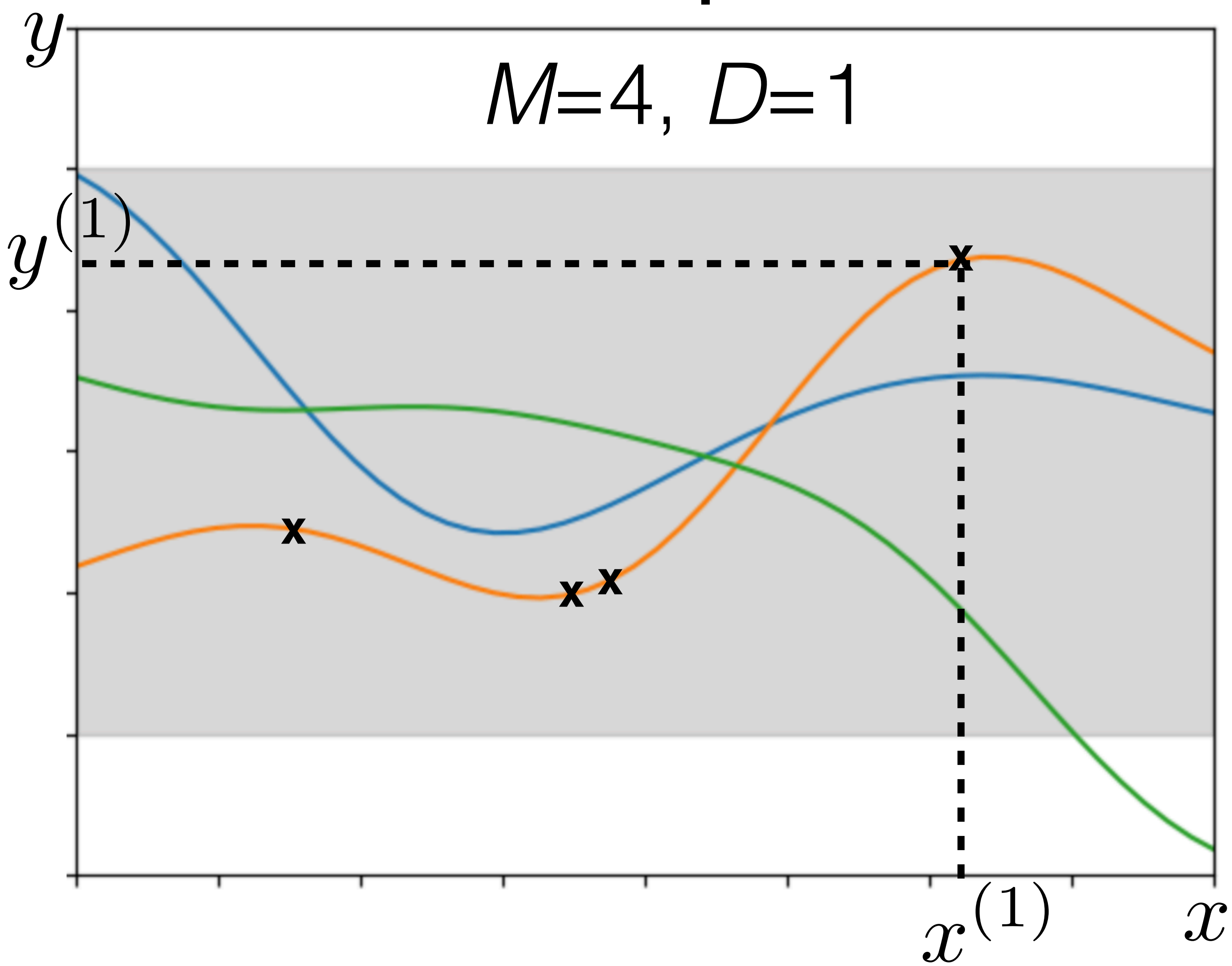


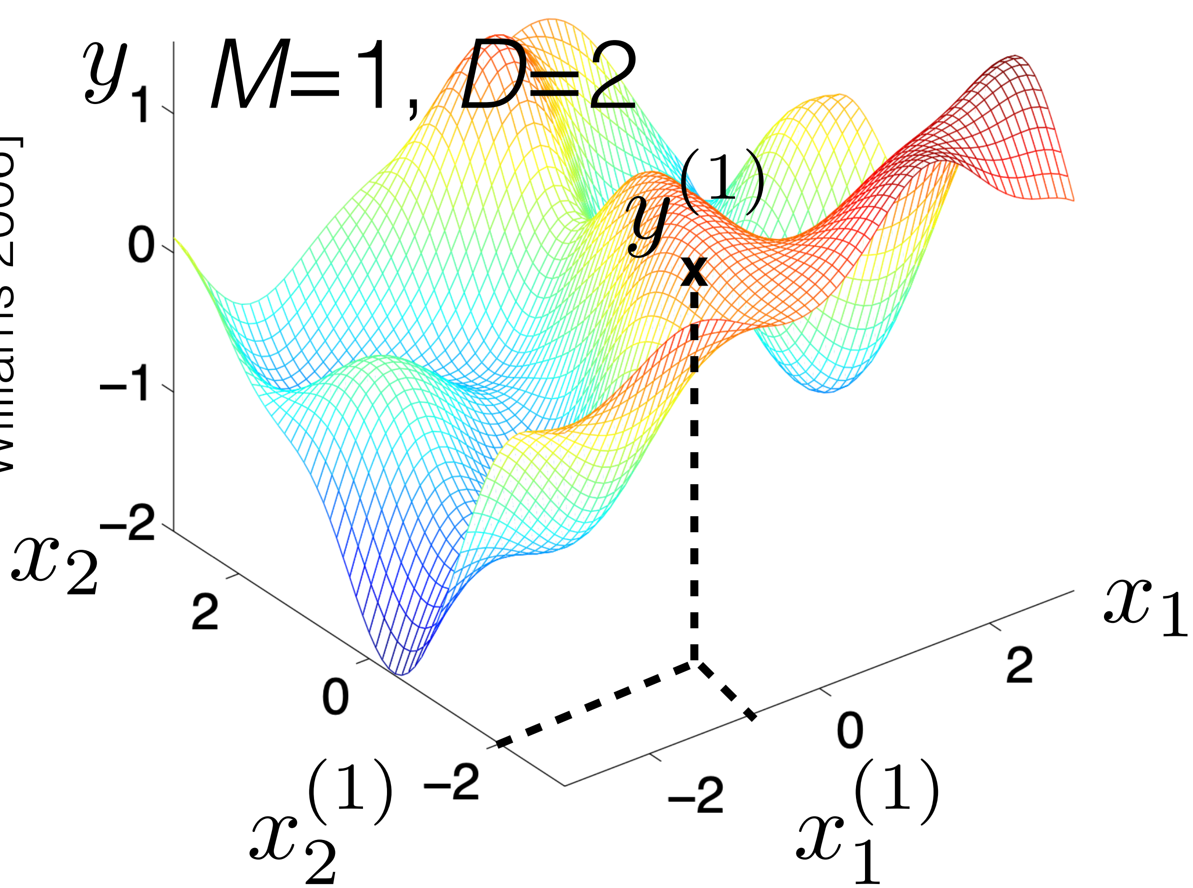


- Note: all of our real-life examples from the start had number of inputs D > 1

Note on “dimension” of the input

- Let’s be careful to separate two types of “dimension”
  - We’re using a superscript to denote (*M* or *N*) number of points in the space
  - We’ll use a subscript for the (*D*) different elements of a point’s vector





[Rasmussen & Williams 2006]

- Note: all of our real-life examples from the start had number of inputs $D > 1$
- $D = 1$ is much easier to visualize, but might not be representative

A Bayesian approach

A Bayesian approach

- p(unknowns | data)

A Bayesian approach

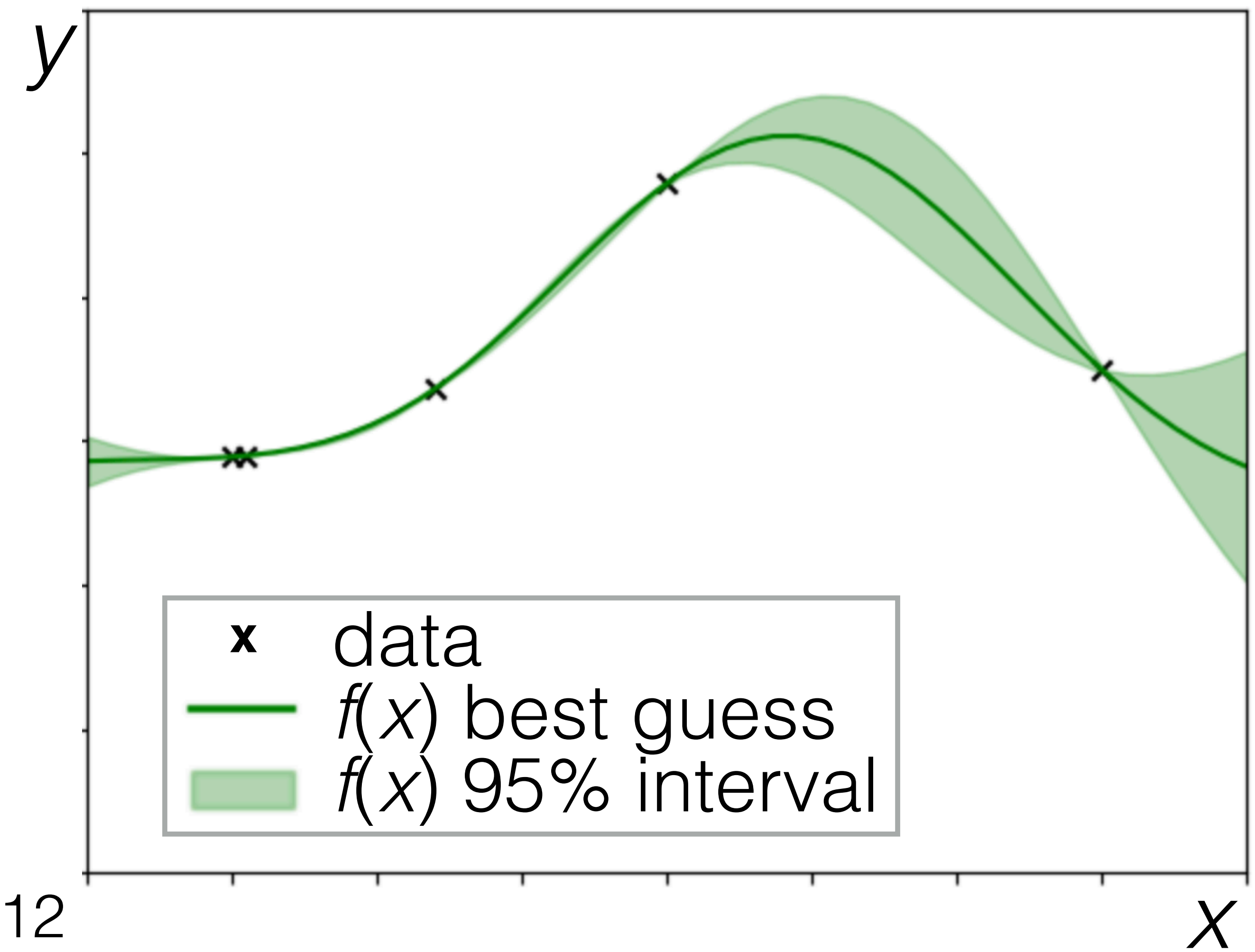
- p(unknowns | data)

Given the data we've seen, what do we know about the underlying function?

A Bayesian approach

- p(unknowns | data)

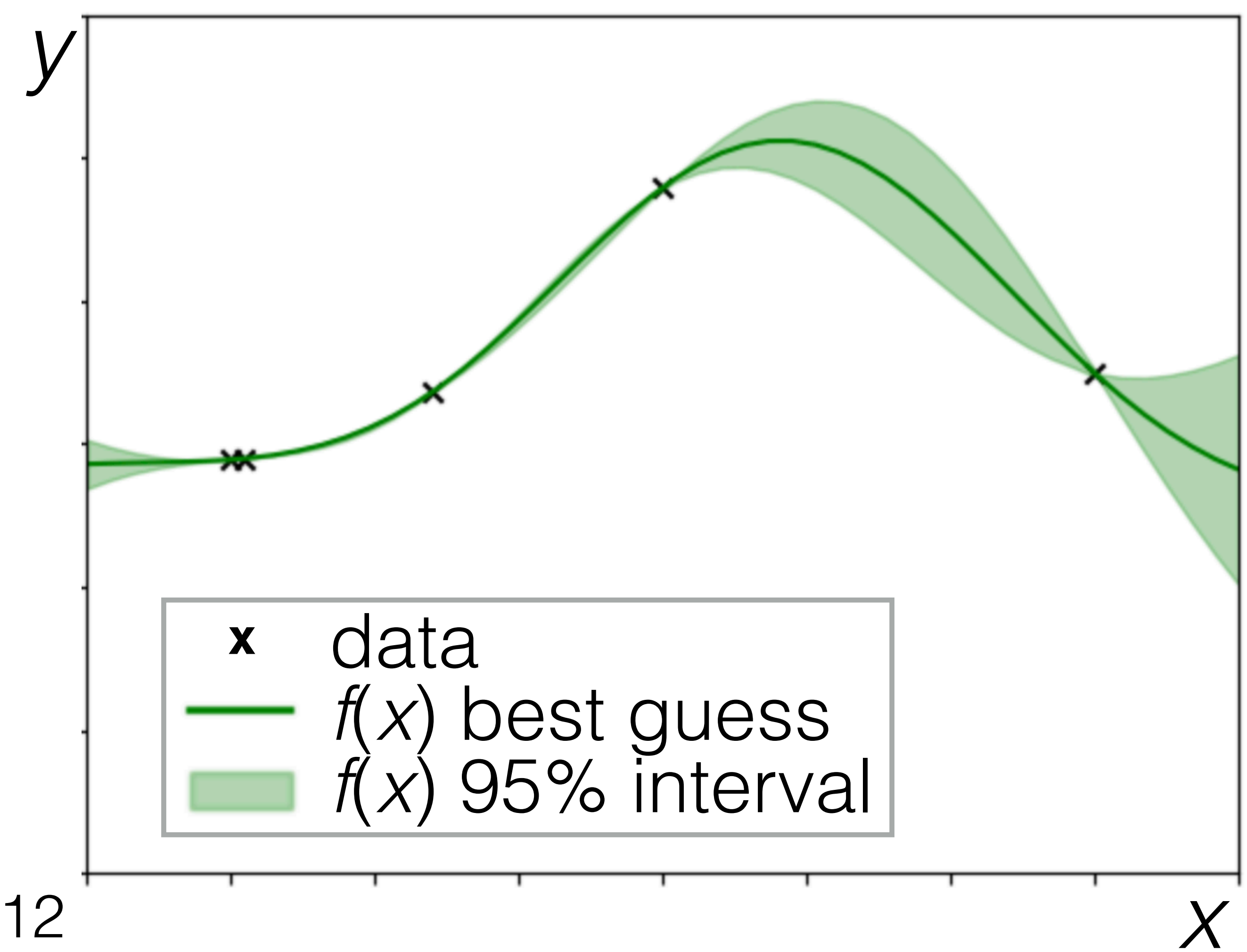
Given the data we've seen, what do we know about the underlying function?



A Bayesian approach

- p(unknowns | data) $\propto$ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

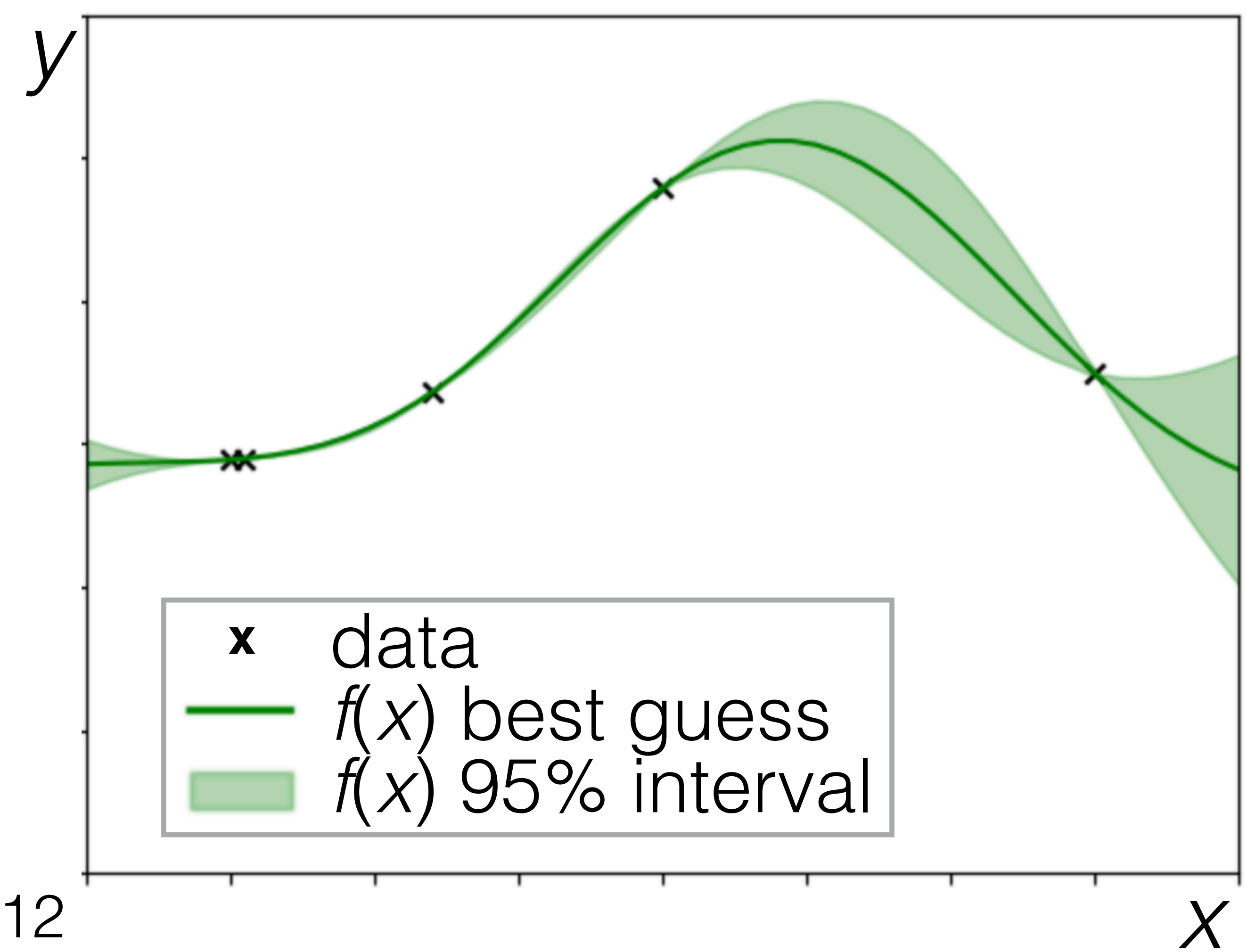


A Bayesian approach

- p(unknowns | data) $\propto$ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

A (statistical) model that can generate functions and data of interest

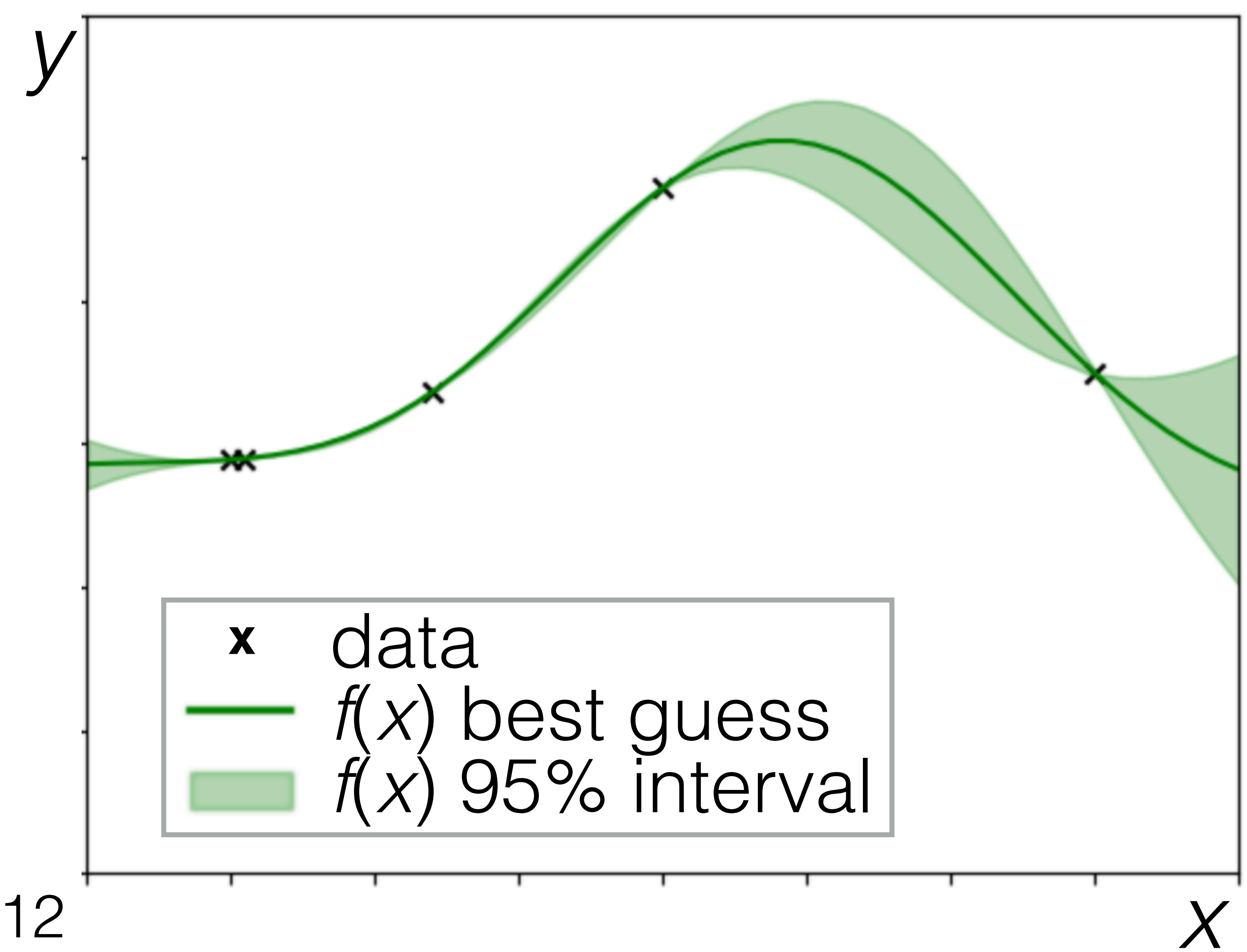


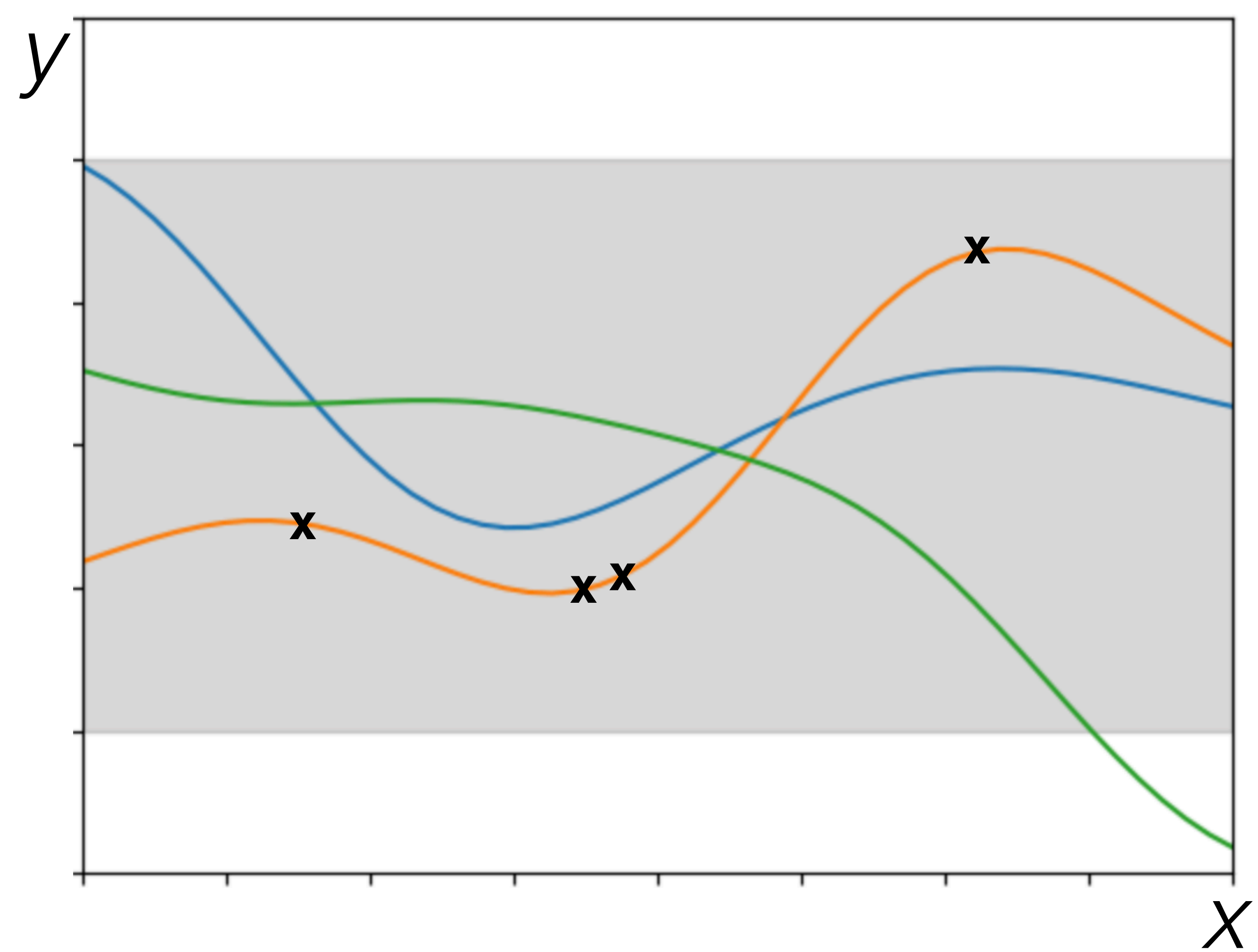
A Bayesian approach

- p(unknowns | data) $\propto$ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

A (statistical) model that can generate functions and data of interest





Inference about unknowns given data

Inference about unknowns given data

- Let X collect the N “training” data points (indexed 1 to N)

Inference about unknowns given data

- Let X collect the N “training” data points (indexed 1 to N)
- Let X' collect the M “test” data points

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function

Inference about unknowns given data

- Let X collect the N “training” data points (indexed 1 to N)
- Let X' collect the M “test” data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*

- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \boxed{?}$$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
    - Where we want to evaluate the function
    - Indexed *N*+1 to *N*+*M*

- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}$$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*

- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\Bigg(\mathbf{0},$$

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0},\right.$$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

A good habit to get into: check the dimensions

$$N\text{x}1 \quad \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

Inference about unknowns given data

- Let *X* collect the *N* "training" data points (indexed 1 to *N*)
- Let *X'* collect the *M* "test" data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X'*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{c} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{c} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

Inference about unknowns given data

- Let X collect the N “training” data points (indexed 1 to N)
- Let X' collect the M “test” data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \boxed{?}$

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $K(X', X) K(X, X)^{-1} f(X)$
  - Covariance: $K(X', X') - K(X', X) K(X, X)^{-1} K(X, X')$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $K(X',X)K(X,X)^{-1}f(X)$ Whole mean: $\boxed{?}$
  - Covariance: $K(X',X') - K(X',X)K(X,X)^{-1}K(X,X')$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{c} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)} \right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $K(X', X) K(X, X)^{-1} f(X)$ Whole mean: *Mx1*
  - Covariance: $K(X', X') - K(X', X) K(X, X)^{-1} K(X, X')$

Inference about unknowns given data

- Let *X* collect the *N* "training" data points (indexed 1 to *N*)
- Let *X'* collect the *M* "test" data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X'*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{c} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)} \right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underset{M\text{x}N}{K(X',X)} K(X,X)^{-1} f(X)$ Whole mean: *Mx1*
  - Covariance: $K(X',X') - K(X',X)K(X,X)^{-1}K(X,X')$

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r}N\text{x}1\\ M\text{x}1\end{array}\begin{bmatrix}f(X)\\ f(X')\end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix}K(X,X) & K(X,X')\\ K(X',X) & K(X',X')\end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{M\text{x}N}\underbrace{K(X,X)}_{N\text{x}N}{}^{-1}f(X)$ Whole mean: Mx1
  - Covariance: $K(X',X') - K(X',X)K(X,X)^{-1}K(X,X')$

Inference about unknowns given data

- Let X collect the N "training" data points (indexed 1 to N)
- Let X' collect the M "test" data points
  - Where we want to evaluate the function
  - Indexed N+1 to N+M
- $K(X,X')$ is the NxM matrix with (n,m) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

A good habit to get into: check the dimensions

$$\begin{array}{c} Nx1 \\ Mx1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)x1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)x(N+M)}\right)$$

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{MxN}\underbrace{K(X,X)^{-1}}_{NxN}\underbrace{f(X)}_{Nx1}$ Whole mean: *Mx1*
  - Covariance: $K(X',X') - K(X',X)K(X,X)^{-1}K(X,X')$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{c} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{M\text{x}N}\underbrace{K(X,X)^{-1}}_{N\text{x}N}\underbrace{f(X)}_{N\text{x}1}$ Whole mean: *Mx1*
  - Covariance: $K(X',X') - K(X',X)K(X,X)^{-1}K(X,X')$

?

Inference about unknowns given data

- Let *X* collect the *N* "training" data points (indexed 1 to *N*)
- Let *X'* collect the *M* "test" data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X'*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

A good habit to get into: check the dimensions

$$\begin{array}{r} Nx1 \\ Mx1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)x1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)x(N+M)}\right)$$

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{MxN}\underbrace{K(X,X)^{-1}}_{NxN}\underbrace{f(X)}_{Nx1}$ Whole mean: *Mx1*
  - Covariance: $\underbrace{K(X',X')}_{MxM} - K(X',X)K(X,X)^{-1}K(X,X')$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)} \right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{M\text{x}N}\underbrace{K(X,X)^{-1}}_{N\text{x}N}\underbrace{f(X)}_{N\text{x}1}$ Whole mean: *M*x*1*
  - Covariance: $\underbrace{K(X',X')}_{M\text{x}M} - K(X',X)K(X,X)^{-1}K(X,X')$

    *M*x*M*

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)} \right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underset{M\text{x}N}{K(X',X)} \underset{N\text{x}N}{K(X,X)^{-1}} \underset{N\text{x}1}{f(X)}$ Whole mean: *Mx1*
  - Covariance: $\underset{M\text{x}M}{K(X',X')} - \underset{M\text{x}M}{\vphantom{K}} \underset{M\text{x}N}{K(X',X)} K(X,X)^{-1} K(X,X')$

Inference about unknowns given data

- Let *X* collect the *N* "training" data points (indexed 1 to *N*)
- Let *X'* collect the *M* "test" data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X'*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

A good habit to get into: check the dimensions

$$\begin{matrix} N\text{x}1 \\ M\text{x}1 \end{matrix}\begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underset{M\text{x}N}{K(X',X)}\underset{N\text{x}N}{K(X,X)^{-1}}\underset{N\text{x}1}{f(X)}$ Whole mean: *M*x*1*
  - Covariance: $\underset{M\text{x}M}{}\ \underset{M\text{x}M}{K(X',X')} - \underset{M\text{x}N}{K(X',X)}\underset{N\text{x}N}{K(X,X)^{-1}}K(X,X')$

Inference about unknowns given data

- Let *X* collect the *N* "training" data points (indexed 1 to *N*)
- Let *X'* collect the *M* "test" data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X'*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underbrace{K(X',X)}_{M\text{x}N}\underbrace{K(X,X)^{-1}}_{N\text{x}N}\underbrace{f(X)}_{N\text{x}1}$ Whole mean: *M*x*1*
  - Covariance: $\underbrace{\underbrace{K(X',X')}_{M\text{x}M} - \underbrace{K(X',X)}_{M\text{x}N}\underbrace{K(X,X)^{-1}}_{N\text{x}N}\underbrace{K(X,X')}_{N\text{x}M}}_{M\text{x}M}$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{l} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

- The conditional satisfies $f(X')|f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underset{M\text{x}N}{K(X',X)}\underset{N\text{x}N}{K(X,X)^{-1}}\underset{N\text{x}1}{f(X)}$ Whole mean: *Mx1*
  - Covariance: $\underset{M\text{x}M}{}\ \underset{M\text{x}M}{K(X',X')} - \underset{M\text{x}N}{K(X',X)}\underset{N\text{x}N}{K(X,X)^{-1}}\underset{N\text{x}M}{K(X,X')}$
- We’ll infer *f*(*X’*) given our simulated data; recall we’re using $k(x,x') = \sigma^2 \exp(-\frac{1}{2}(x-x')^2), \sigma = 1$

Inference about unknowns given data

- Let *X* collect the *N* “training” data points (indexed 1 to *N*)
- Let *X’* collect the *M* “test” data points
  - Where we want to evaluate the function
  - Indexed *N*+1 to *N*+*M*
- *K*(*X*,*X’*) is the *N*x*M* matrix with (*n*,*m*) entry $k(x^{(n)}, x^{(N+m)})$
- Then by our model

$$\begin{array}{r} N\text{x}1 \\ M\text{x}1 \end{array} \begin{bmatrix} f(X) \\ f(X') \end{bmatrix} \sim \mathcal{N}\left(\underbrace{\mathbf{0}}_{(N+M)\text{x}1}, \underbrace{\begin{bmatrix} K(X,X) & K(X,X') \\ K(X',X) & K(X',X') \end{bmatrix}}_{(N+M)\text{x}(N+M)}\right)$$

A good habit to get into: check the dimensions

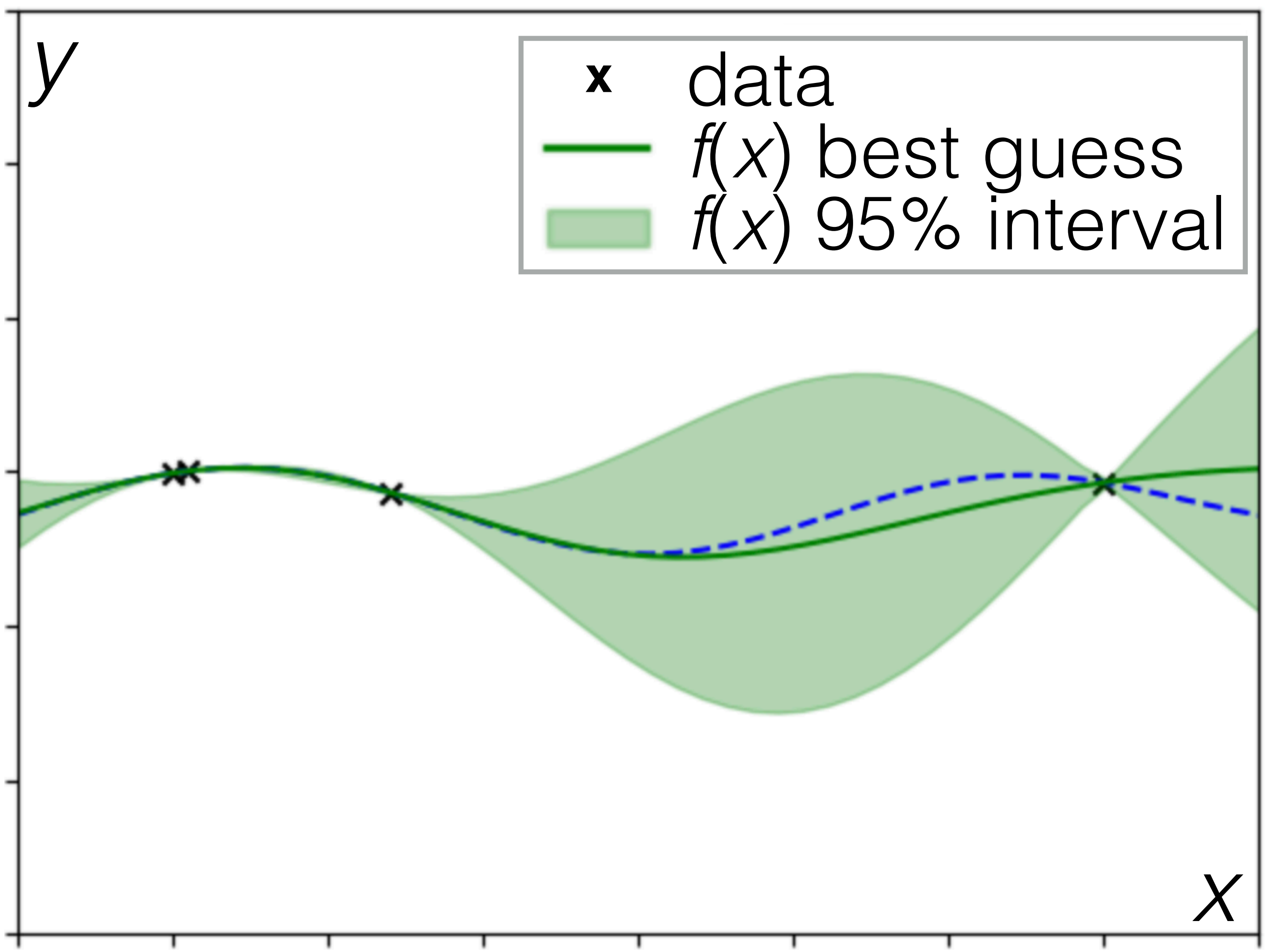
- The conditional satisfies $f(X') | f(X), X, X' \sim \mathcal{N}$ with
  - Mean: $\underset{M\text{x}N}{K(X',X)}\,\underset{N\text{x}N}{K(X,X)^{-1}}\,\underset{N\text{x}1}{f(X)}$ Whole mean: *M*x*1*
  - Covariance: $\underset{M\text{x}M}{}\ \underset{M\text{x}M}{K(X',X')} - \underset{M\text{x}N}{K(X',X)}\,\underset{N\text{x}N}{K(X,X)^{-1}}\,\underset{N\text{x}M}{K(X,X')}$
- We’ll infer *f*(*X’*) given our simulated data; recall we’re using

$$k(x,x') = \sigma^2 \exp(-\tfrac{1}{2}(x-x')^2), \sigma = 1$$

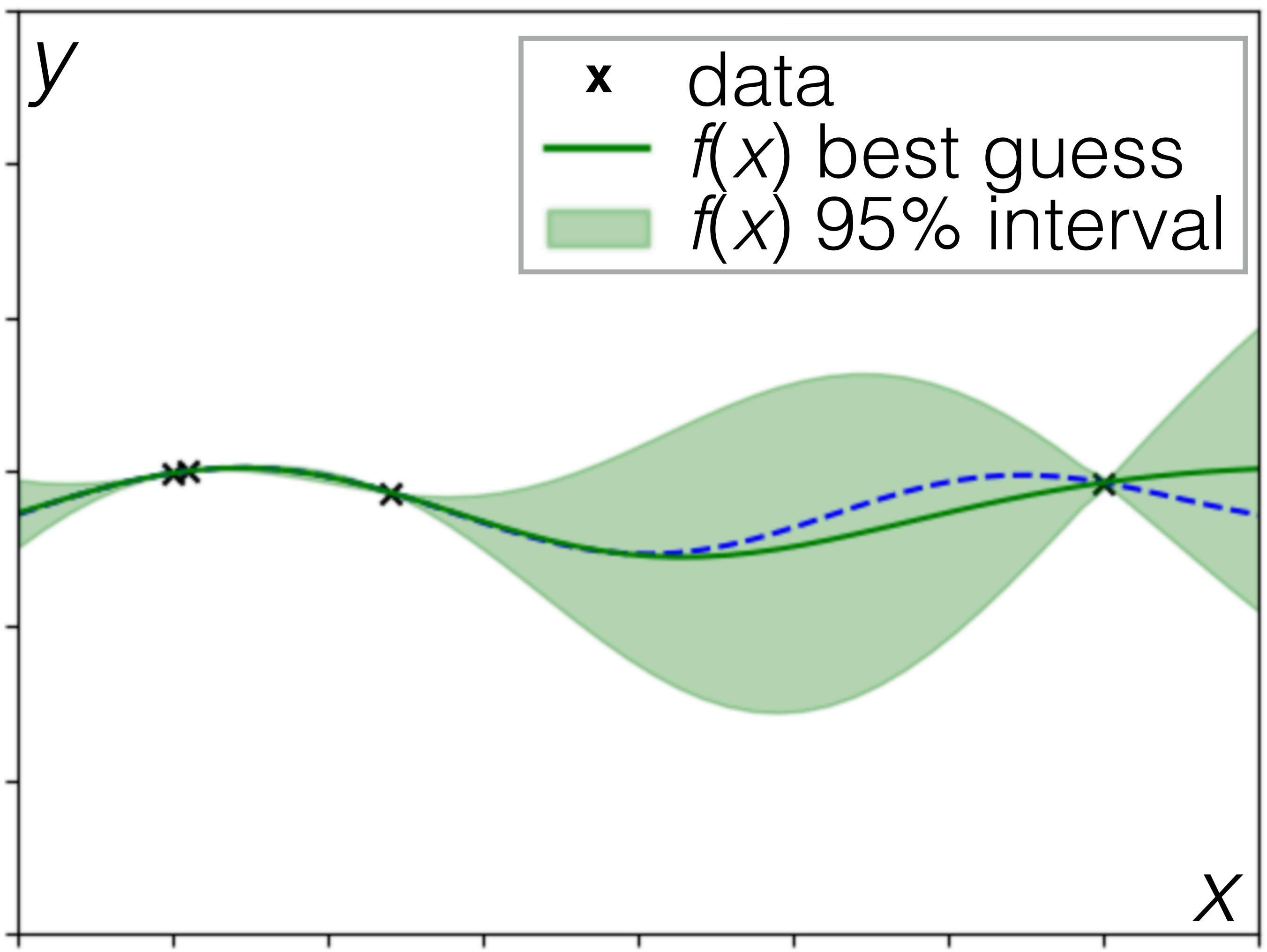
[demo]

Closer look at the uncertainty interval

Closer look at the uncertainty interval

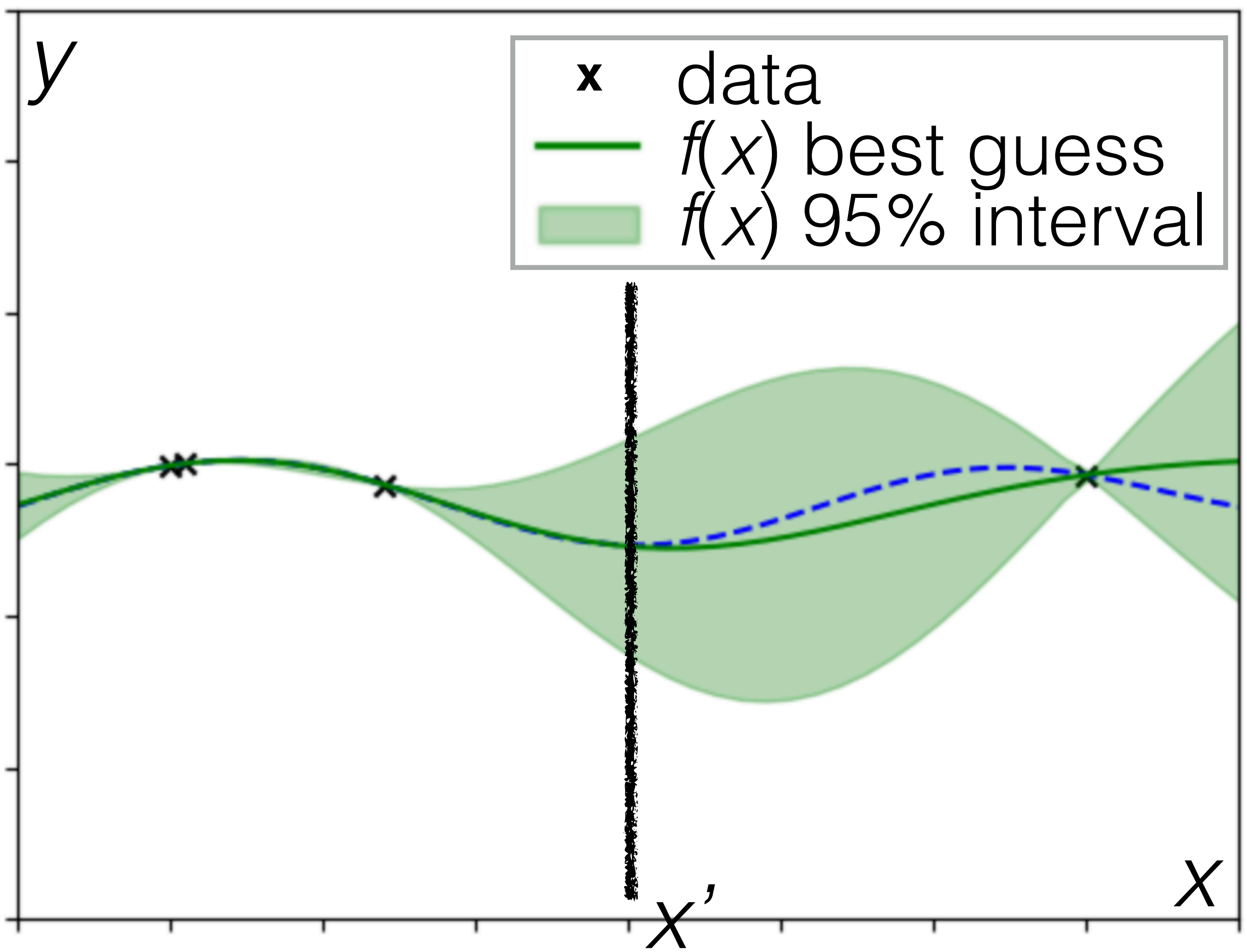


Closer look at the uncertainty interval



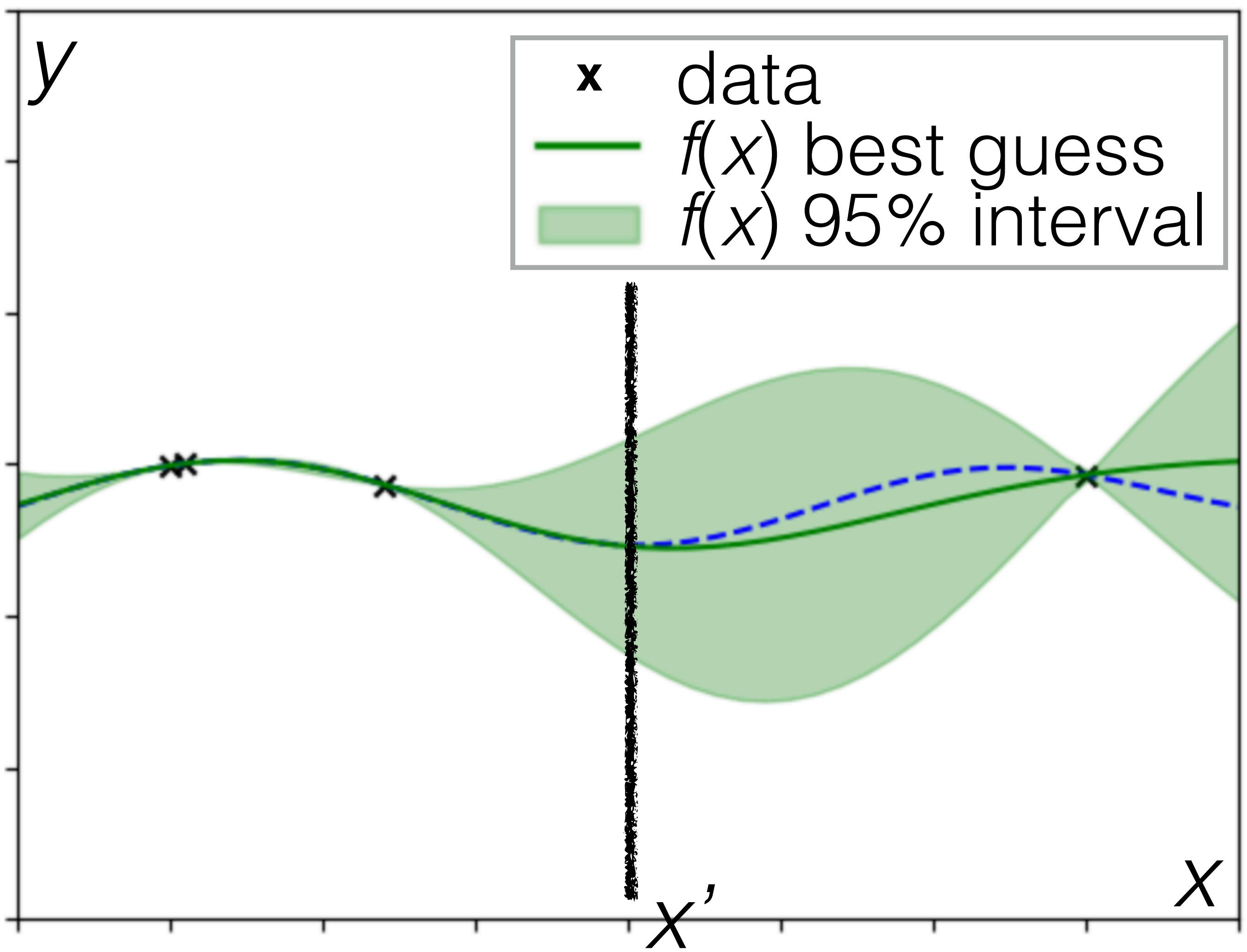
- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian

Closer look at the uncertainty interval



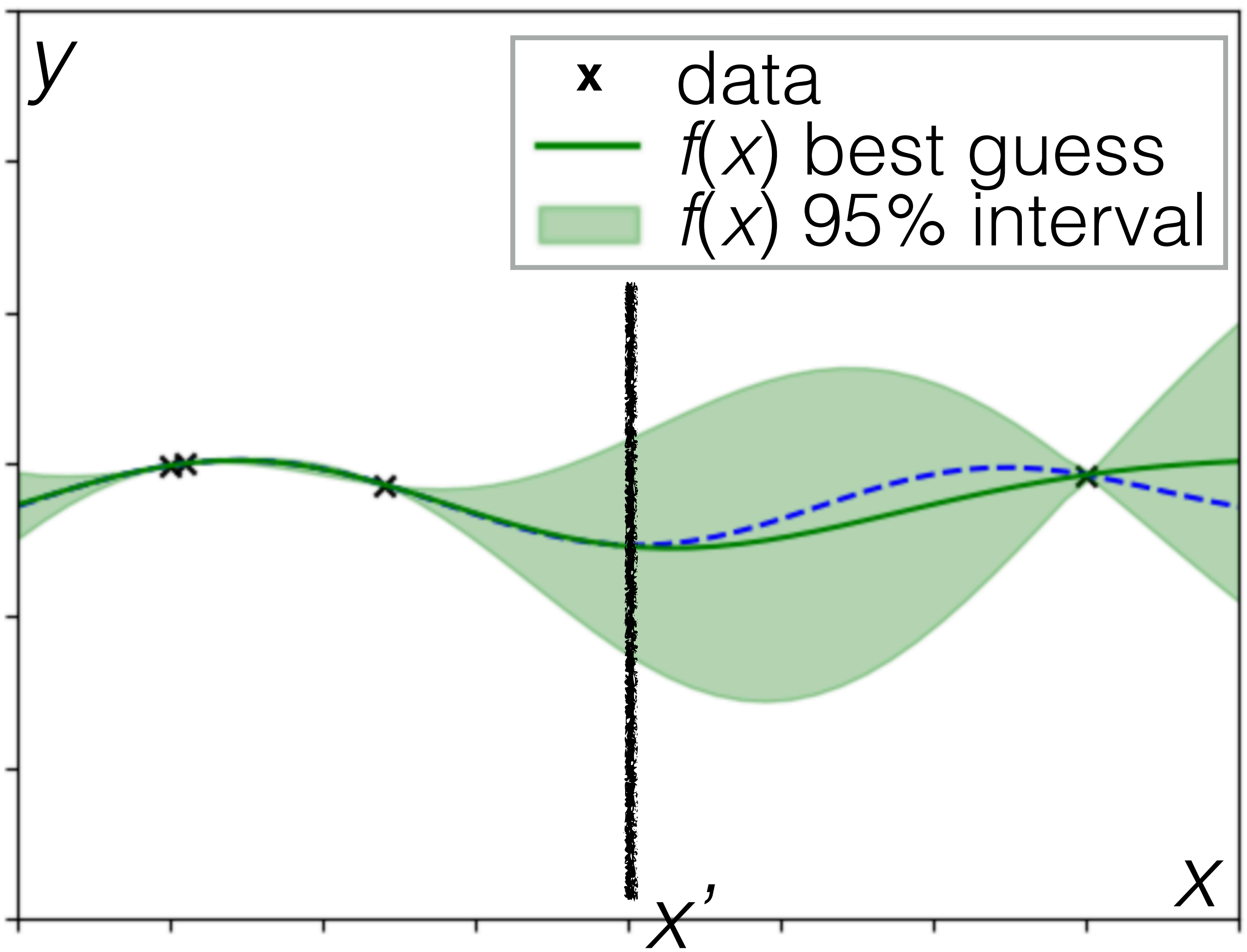
- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian

Closer look at the uncertainty interval



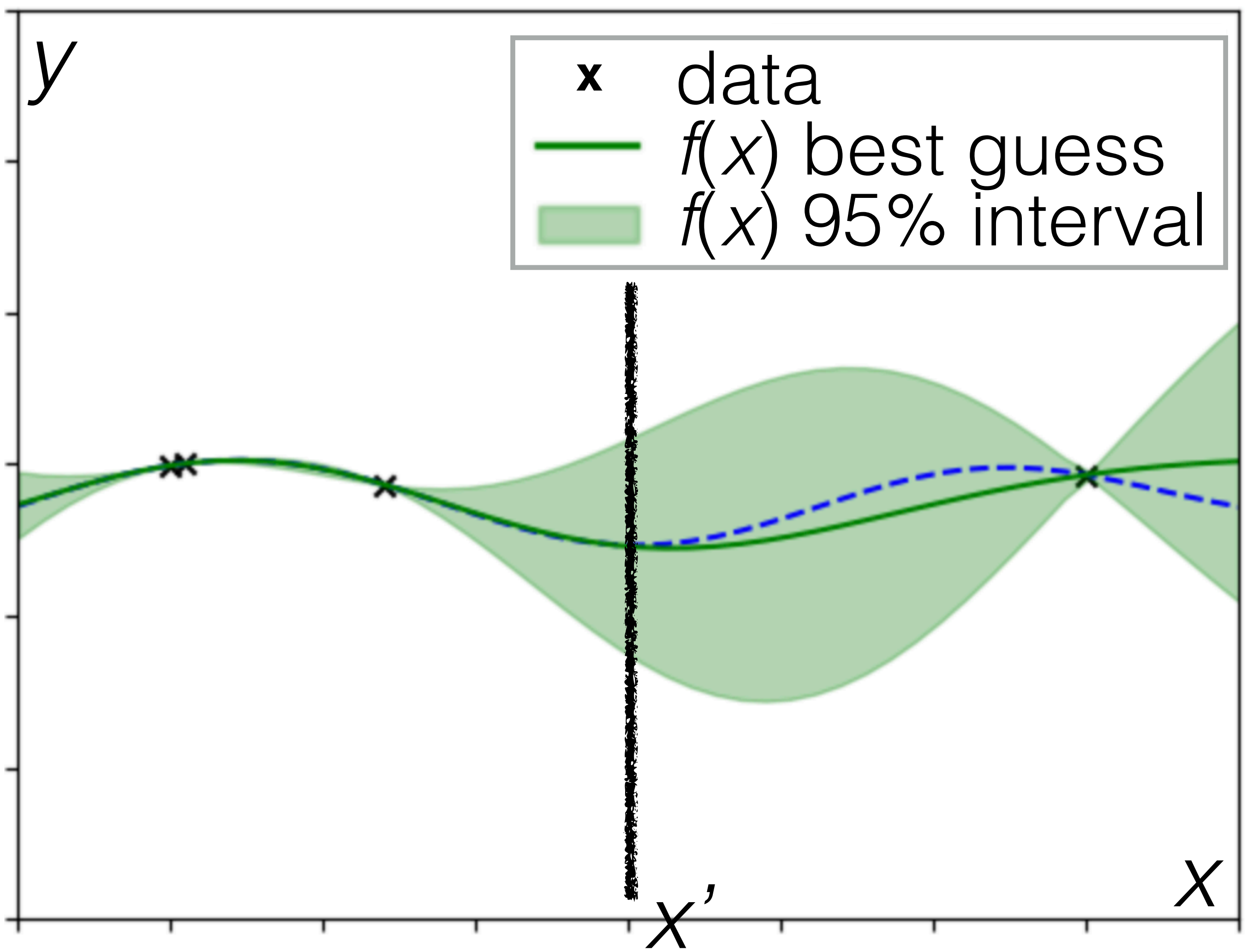
- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian
- The green line at point *x'* is the mean of that Gaussian

Closer look at the uncertainty interval



- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian
- The green line at point *x'* is the mean of that Gaussian
- The green interval at that point: mean +/- 2 std devs

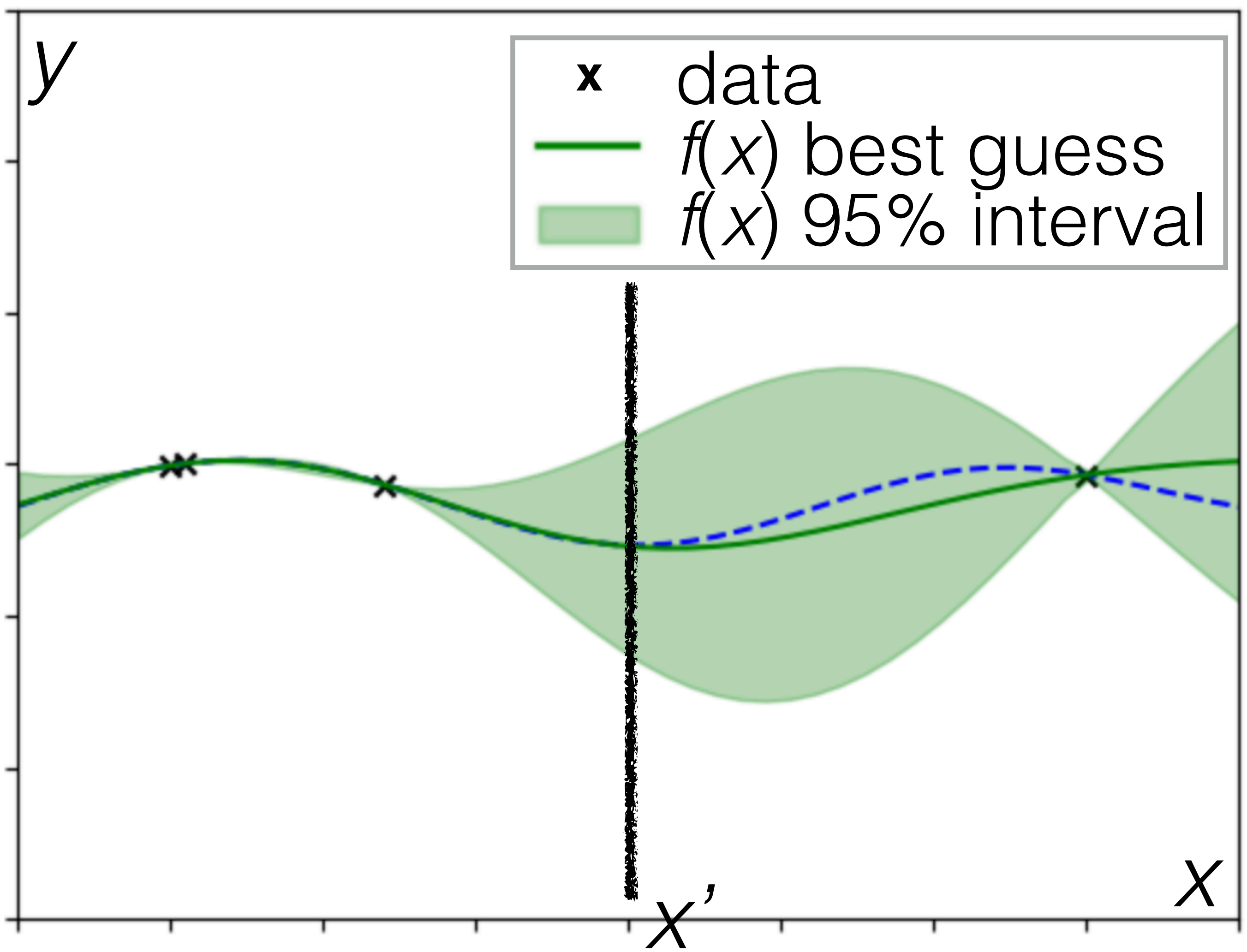
Closer look at the uncertainty interval



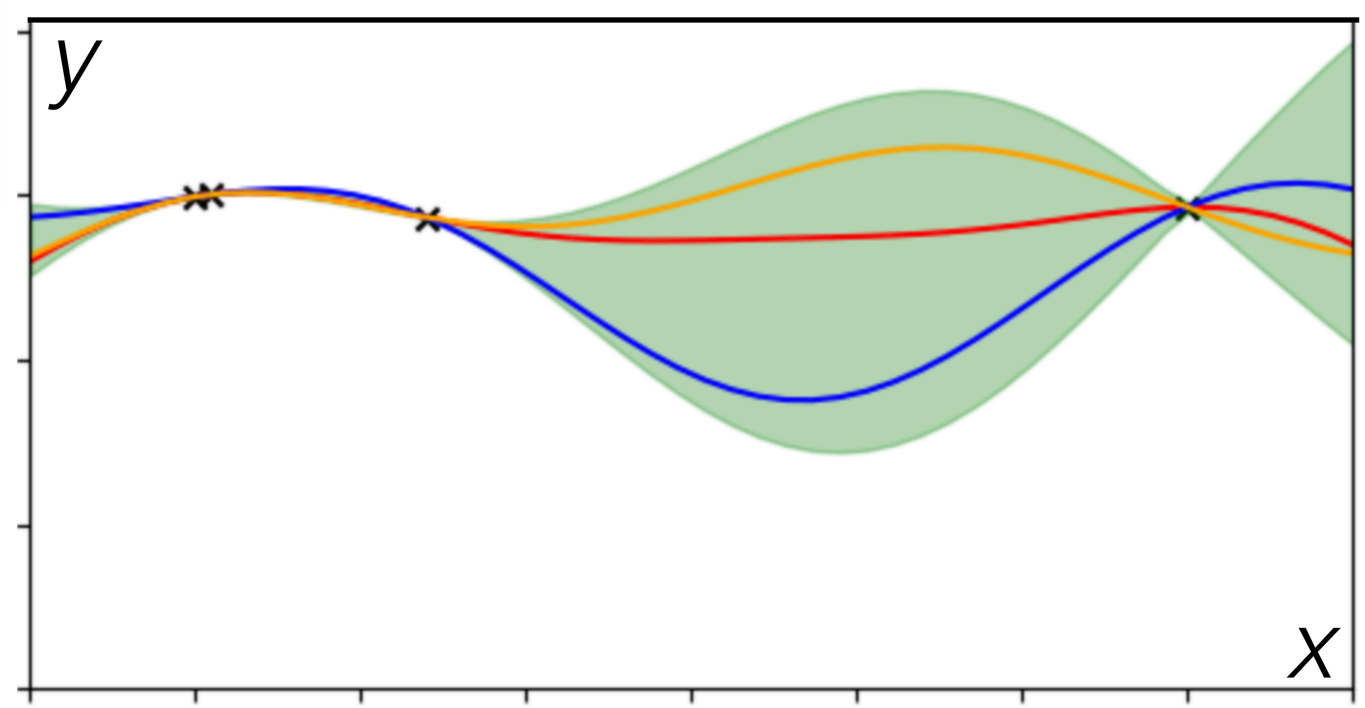
- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian
- The green line at point *x'* is the mean of that Gaussian
- The green interval at that point: mean +/- 2 std devs

- Draw random *f* conditional on the training data

Closer look at the uncertainty interval

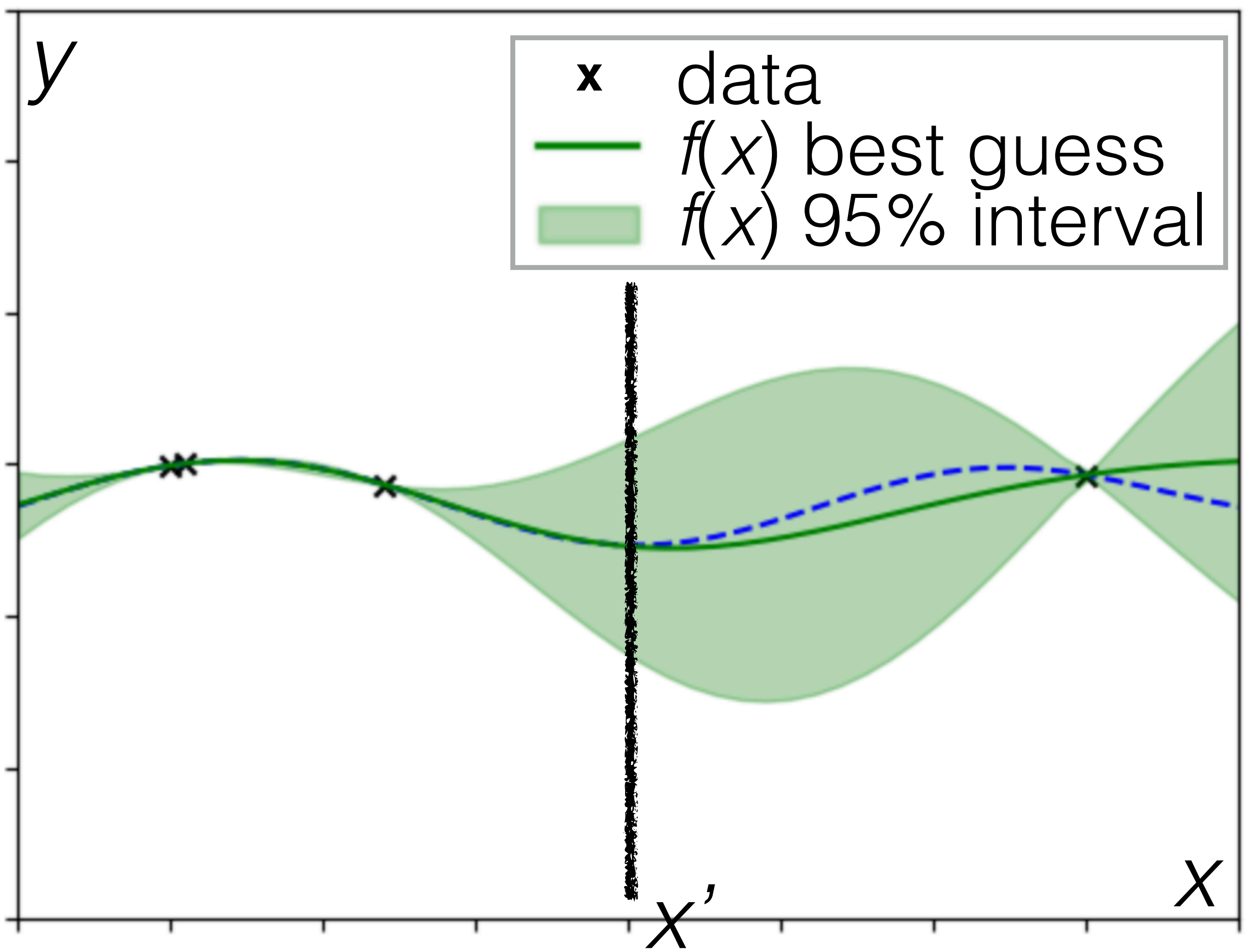


- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian
- The green line at point *x'* is the mean of that Gaussian
- The green interval at that point: mean +/- 2 std devs

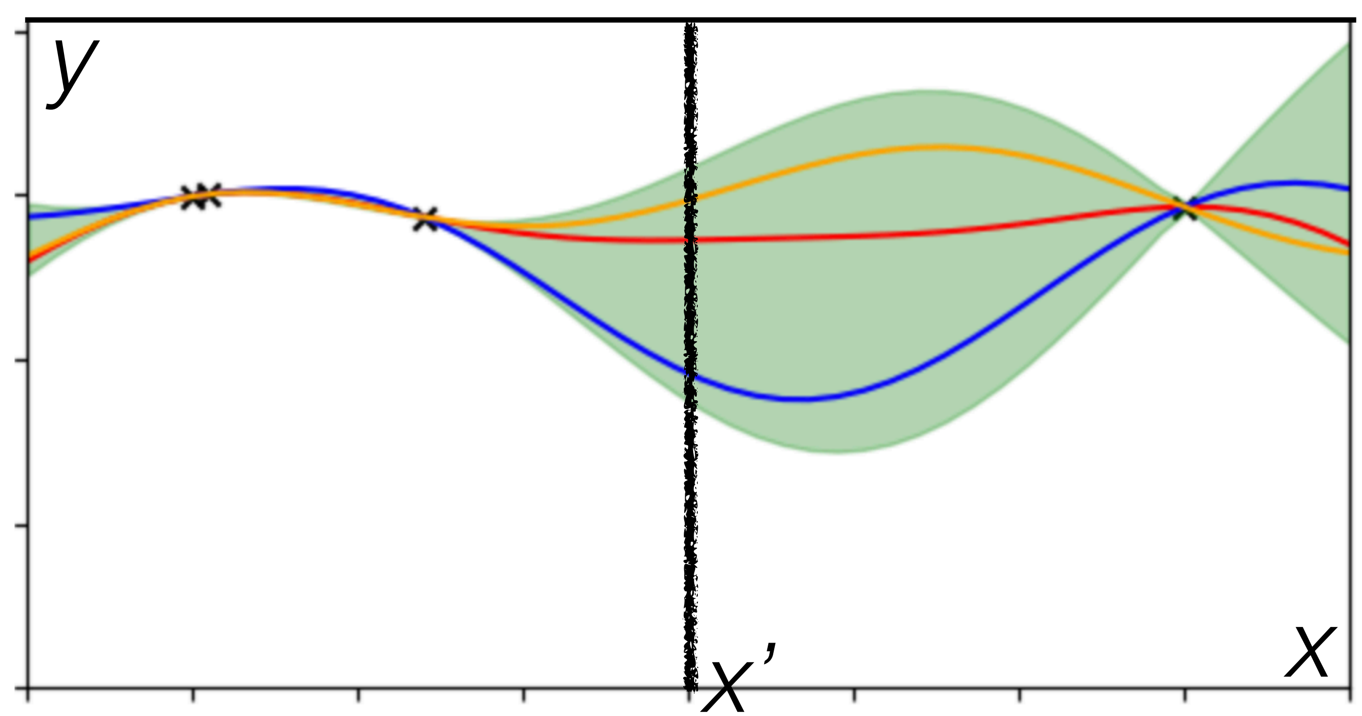


- Draw random *f* conditional on the training data

Closer look at the uncertainty interval



- Under GP, $f(x')|f(X), X, x'$ at a point *x'* is marginally Gaussian
- The green line at point *x'* is the mean of that Gaussian
- The green interval at that point: mean +/- 2 std devs



- Draw random *f* conditional on the training data