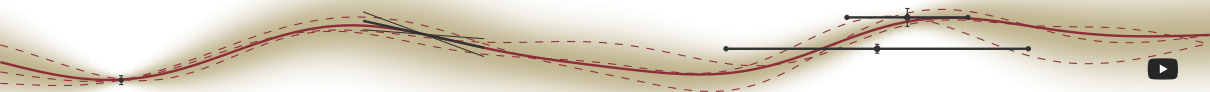
Probabilistic Inference and Learning
Lecture 07
Parametric Regression

Philipp Hennig

11 May 2020

Eberhard Karls
Universität Tübingen

Faculty of Science
Department of Computer Science
Chair for the Methods of Machine Learning



#	date	content	Ex
1	20.04.	Introduction	1
2	21.04.	Reasoning under Uncertainty	
3	27.04.	Continuous Variables	2
4	28.04.	Monte Carlo	
5	04.05.	Markov Chain Monte Carlo	3
6	05.05.	Gaussian Distributions	
7	11.05.	**Parametric Regression**	4
8	12.05.	Understanding Deep Learning	
9	18.05.	Gaussian Processes	5
10	19.05.	An Example for GP Regression	
11	25.05.	Understanding Kernels	6
12	26.05.	Gauss-Markov Models	
13	08.06.	GP Classification	7
14	09.06.	Logistic Regression	8
15	15.06.	Exponential Families	
16	16.06.	Graphical Models	9
17	22.06.	Factor Graphs	
18	23.06.	The Sum-Product Algorithm	10
19	29.06.	Example: Topic Models	
20	30.06.	Mixture Models	11
21	06.07.	EM	
22	07.07.	Variational Inference	12
23	13.07.	Example: Topic Models	
24	14.07.	Example: Inferring Topics	13
25	20.07.	Example: Kernel Topic Models	
26	21.07.	Revision	

Recap: Gaussian Distributions

Gaussian distributions provide the linear algebra of inference

- products of Gaussians are Gaussians $\quad C := (A^{-1} + B^{-1})^{-1} \quad c := C(A^{-1}a + B^{-1}b)$

$$\mathcal{N}(x; a, A)\mathcal{N}(x; b, B) = \mathcal{N}(x; c, C)\mathcal{N}(a; b, A + B)$$

- marginals of Gaussians are Gaussians

$$\int \mathcal{N}\left[\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right] dy = \mathcal{N}(x; \mu_x, \Sigma_{xx})$$

- (linear) conditionals of Gaussians are Gaussians

$$p(x \mid y) = \frac{p(x, y)}{p(y)} = \mathcal{N}\left(x; \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)$$

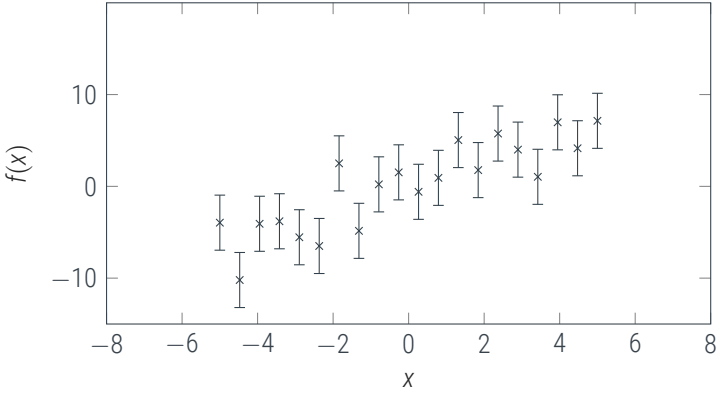
- linear projections of Gaussians are Gaussians

$$p(z) = \mathcal{N}(z; \mu, \Sigma) \quad \Rightarrow \quad p(Az) = \mathcal{N}(Az, A\mu, A\Sigma A^\intercal)$$

Bayesian inference becomes linear algebra $\qquad p(x) = \mathcal{N}(x; \mu, \Sigma) \qquad p(y \mid x) = \mathcal{N}(y; A^\intercal x + b, \Lambda)$

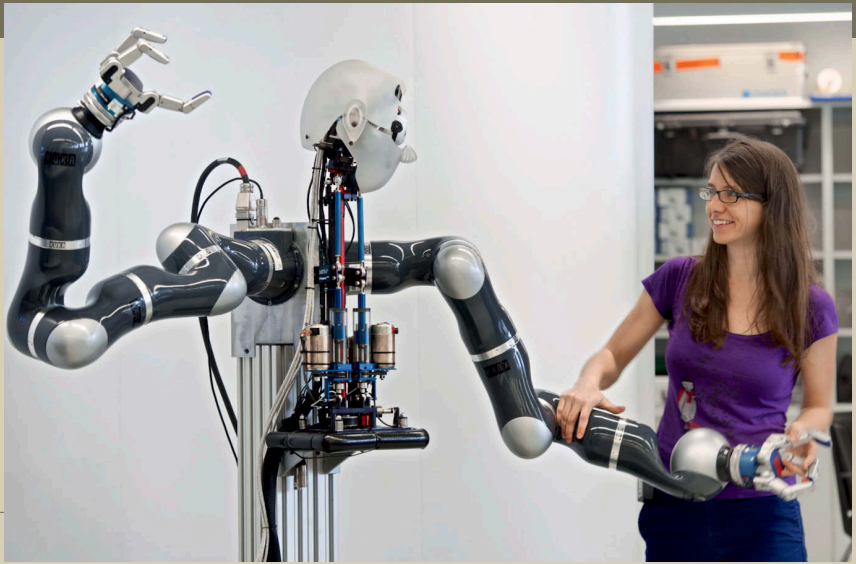
$$p(B^\intercal x + c \mid y) = \mathcal{N}[B^\intercal x + c; B^\intercal \mu + c + B^\intercal \Sigma A(A^\intercal \Sigma A + \Lambda)^{-1}(y - A^\intercal \mu - b), B^\intercal \Sigma B - B^\intercal \Sigma A(A^\intercal \Sigma A + \Lambda)^{-1} A^\intercal \Sigma B]$$

given: $\boldsymbol{y} \in \mathbb{R}^N, \quad p(\boldsymbol{y} \mid f) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{f}(\boldsymbol{x}), \sigma^2 I_N)$. What is f?

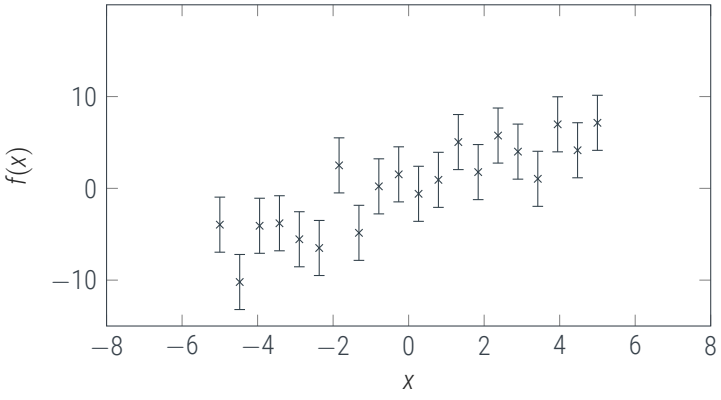


Supervised Regression

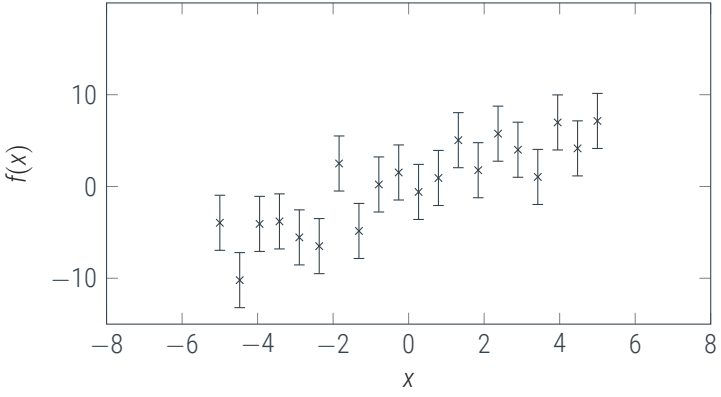
A data set

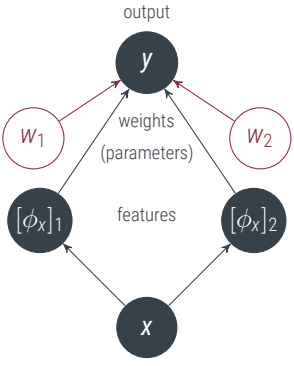


given: $\boldsymbol{y} \in \mathbb{R}^N, \quad p(\boldsymbol{y} \mid f) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{f}(\boldsymbol{x}), \sigma^2 I_N)$. What is f?



Assume **linear** function $f(x) = w_1 + w_2 x = \phi_x^\intercal \boldsymbol{w}$ with **features** $\phi(x) = \begin{bmatrix} 1 \\ x \end{bmatrix} =: \phi_x$

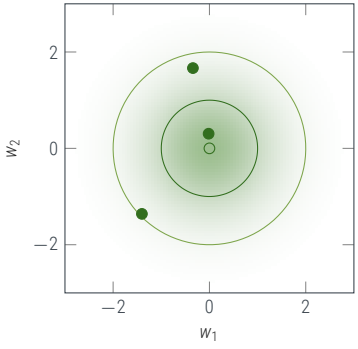




A linear **generative** model

if every variable is Gaussian and every relationship is linear, all marginals and conditionals are also Gaussian

$$f(x) = w_1 + w_2 x = \phi_x^\intercal w$$
$$p(w) = \mathcal{N}(w; \mu, \Sigma)$$

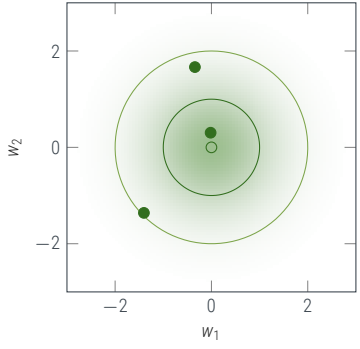


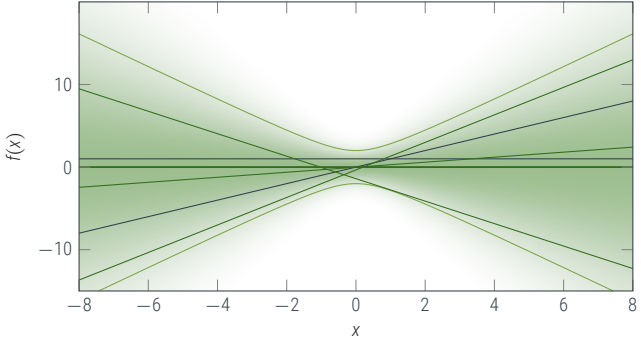
A linear **generative** model

if every variable is Gaussian and every relationship is linear, all marginals and conditionals are also Gaussian

$$f(x) = w_1 + w_2 x = \phi_x^\intercal w$$
$$p(w) = \mathcal{N}(w; \mu, \Sigma)$$

$$p(f) = \mathcal{N}(f; \phi_x^\intercal \mu, \phi_x^\intercal \Sigma \phi_x)$$

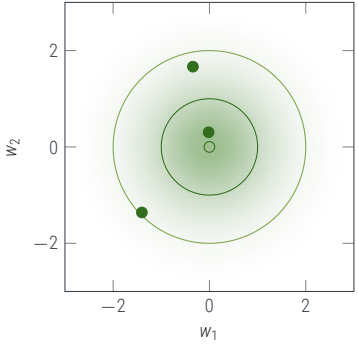




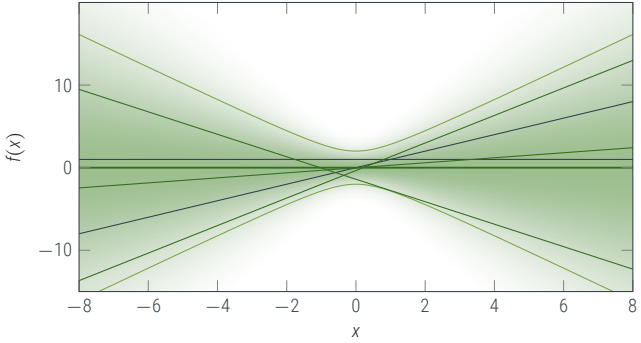
A linear **generative** model

if every variable is Gaussian and every relationship is linear, all marginals and conditionals are also Gaussian

$$f(x) = w_1 + w_2 x = \phi_x^\intercal w$$
$$p(w) = \mathcal{N}(w; \mu, \Sigma)$$



$$p(f) = \mathcal{N}(f; \phi_x^\intercal \mu, \phi_x^\intercal \Sigma \phi_x)$$



Notation

this will become exceedingly helpful later on

Dataset: $X := \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{X}^N, \boldsymbol{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N$. We will use the following very sloppy notation, sloppily

$$\phi_x := \phi(x) = \begin{bmatrix} 1 \\ x \end{bmatrix} \in \mathbb{R}^F \qquad \phi_X := \begin{bmatrix} \phi(x_1) & \phi(x_2) & \cdots & \phi(x_N) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \in \mathbb{R}^{F \times N}$$

$$f_x := f(x) \in \mathbb{R} \qquad f_X := \phi_X^\intercal \boldsymbol{w} = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) \\ \phi_1(x_2) & \phi_2(x_2) \\ \vdots & \vdots \\ \phi_1(x_N) & \phi_2(x_N) \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \phi_{x_1}^\intercal \boldsymbol{w} \\ \phi_{x_2}^\intercal \boldsymbol{w} \\ \vdots \\ \phi_{x_N}^\intercal \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{bmatrix} \in \mathbb{R}^N$$

Think of f as an infinitely long vector, indexed by x:

$$\boldsymbol{v} \in \mathbb{R}^N, I \in \mathbb{N}^d \Rightarrow v_I := [v_{I_1}, \ldots, v_{I_d}] \in \mathbb{R} \quad \Longleftrightarrow \quad f \in \mathbb{R}^\infty, X \in \mathbb{R}^N \Rightarrow f_X := [f_{x_1}, \ldots, f_{x_N}] \in \mathbb{R}^N.$$

Gaussian Inference on a linear function

weight space / function space

$$\text{prior} \quad p(w) = \mathcal{N}(w; \mu, \Sigma) \quad \Rightarrow \quad p(f) = \mathcal{N}(f_X; \phi_X^\intercal \mu, \phi_X \Sigma \phi_X)$$

$$\text{likelihood} \quad p(y \mid w, \phi_X) = \mathcal{N}(y; \phi_X^\intercal w, \sigma^2 I) = \mathcal{N}(y; f_X, \sigma^2 I)$$

Gaussian Inference on a linear function

weight space / function space

$$
\begin{aligned}
\text{prior} \quad p(w) &= \mathcal{N}(w; \mu, \Sigma) \quad \Rightarrow \quad p(f) = \mathcal{N}(f_X; \phi_X^\intercal \mu, \phi_X \Sigma \phi_X) \\
\text{likelihood} \quad p(y \mid w, \phi_X) &= \mathcal{N}(y; \phi_X^\intercal w, \sigma^2 I) = \mathcal{N}(y; f_X, \sigma^2 I) \\
\text{posterior on } \boldsymbol{w} \quad p(w \mid \boldsymbol{y}, \phi_X) &= \mathcal{N}(w; \mu + \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal \mu), \\
&\qquad \Sigma - \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}\phi_X^\intercal \Sigma) \\
&= \mathcal{N}\Big(w; (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1} \left(\Sigma^{-1}\mu + \sigma^{-2}\phi_X \boldsymbol{y}\right), \\
&\qquad (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\Big)
\end{aligned}
$$

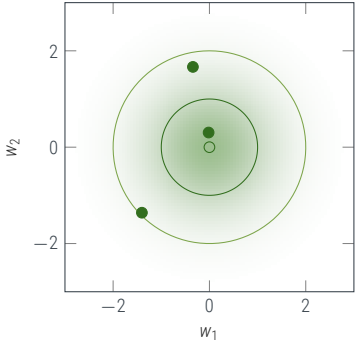
Gaussian Inference on a linear function

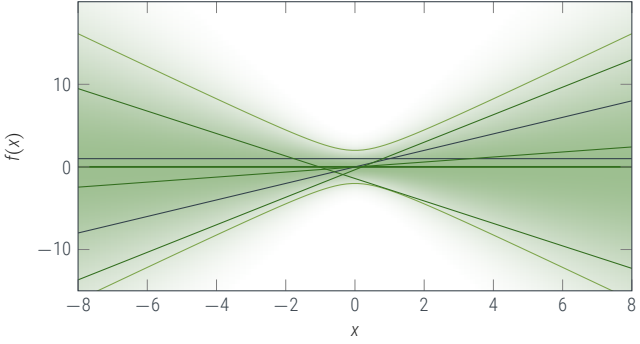
weight space / function space

$$
\begin{aligned}
\text{prior} \quad p(w) &= \mathcal{N}(w; \mu, \Sigma) \quad \Rightarrow \quad p(f) = \mathcal{N}(f_X; \phi_X^\intercal \mu, \phi_X \Sigma \phi_X) \\
\text{likelihood} \quad p(y \mid w, \phi_X) &= \mathcal{N}(y; \phi_X^\intercal w, \sigma^2 I) = \mathcal{N}(y; f_X, \sigma^2 I) \\
\text{posterior on } \boldsymbol{w} \quad p(w \mid \boldsymbol{y}, \phi_X) &= \mathcal{N}(w; \mu + \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal \mu), \\
&\qquad \Sigma - \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}\phi_X^\intercal \Sigma) \\
&= \mathcal{N}\Big(w; (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\left(\Sigma^{-1}\mu + \sigma^{-2}\phi_X \boldsymbol{y}\right), \\
&\qquad (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\Big) \\
\text{posterior on } f \quad p(f_x \mid y, \phi_X) &= \mathcal{N}(f_x; \phi_x^\intercal \mu + \phi_x^\intercal \Sigma \phi_X (\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal \mu), \\
&\qquad \phi_x^\intercal \Sigma \phi_x - \phi_x^\intercal \Sigma \phi_X (\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}\phi_X^\intercal \Sigma \phi_x) \\
&\qquad \mathcal{N}\Big(f_x; \phi_x(\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\left(\Sigma^{-1}\mu + \sigma^{-2}\phi_X \boldsymbol{y}\right), \\
&\qquad \phi_x(\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\phi_x^\intercal\Big)
\end{aligned}
$$

$$p(w \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(w; \mu + \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal\mu), \Sigma - \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)$$

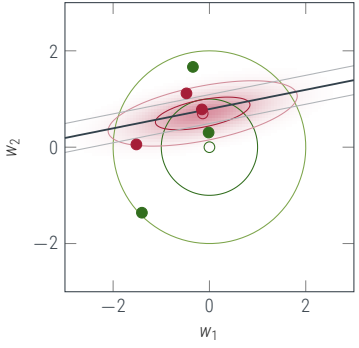
$$p(f_x \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(f_x; \phi_x^\intercal\mu + \phi_x^\intercal\Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal\mu), \phi_x^\intercal\left(\Sigma - ^\intercal\ \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)\phi_x\right)$$

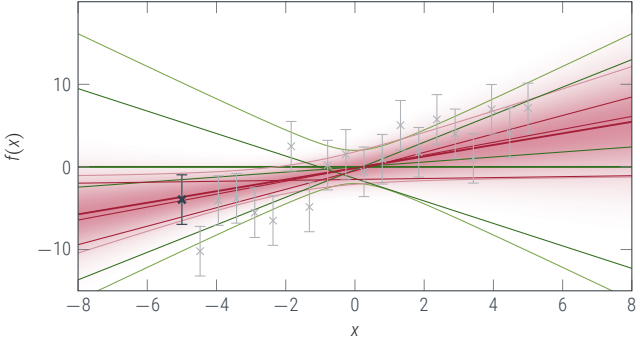




$$p(w \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(w; \mu + \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal\mu), \Sigma - \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)$$

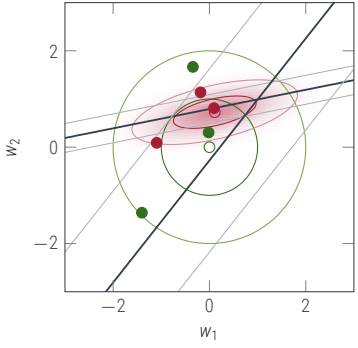
$$p(f_x \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(f_x; \phi_x^\intercal\mu + \phi_x^\intercal\Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal\mu), \phi_x^\intercal\left(\Sigma - ^\intercal\ \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)\phi_x\right)$$

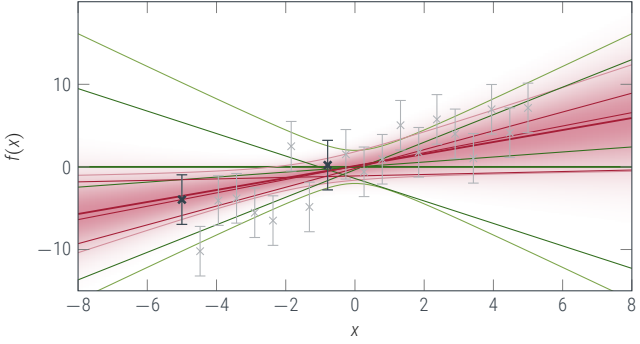




$$p(w \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(w; \mu + \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal\mu), \Sigma - \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)$$

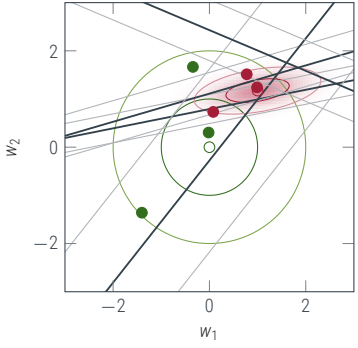
$$p(f_x \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(f_x; \phi_x^\intercal\mu + \phi_x^\intercal\Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal\mu), \phi_x^\intercal\left(\Sigma - ^\intercal\ \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)\phi_x\right)$$

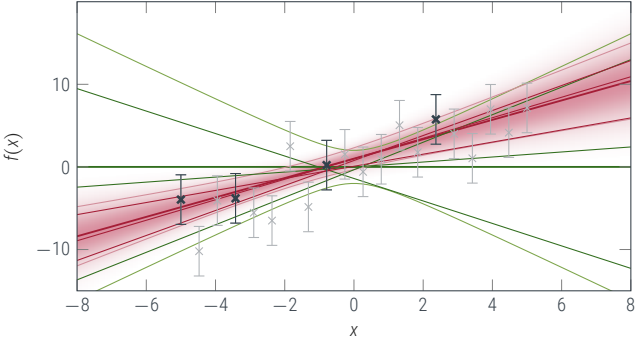




$$p(w \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(w; \mu + \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal\mu), \Sigma - \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)$$

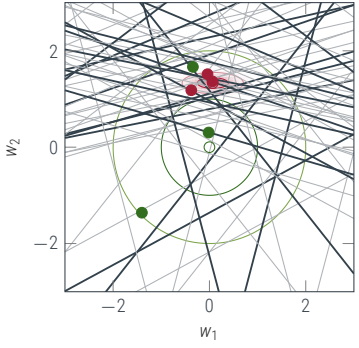
$$p(f_x \mid \boldsymbol{y}, \phi_X) = \mathcal{N}\left(f_x; \phi_x^\intercal\mu + \phi_x^\intercal\Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal\mu), \phi_x^\intercal\left(\Sigma - ^\intercal\ \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)\phi_x\right)$$

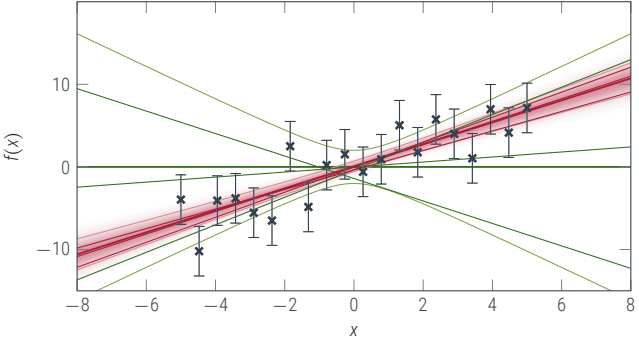




$$p(w \mid \mathbf{y}, \phi_X) = \mathcal{N}\left(w; \mu + \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(\mathbf{y} - \phi_X^\intercal\mu), \Sigma - \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)$$

$$p(f_x \mid \mathbf{y}, \phi_X) = \mathcal{N}\left(f_x; \phi_x^\intercal\mu + \phi_x^\intercal\Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal\mu), \phi_x^\intercal\left(\Sigma - ^\intercal\ \Sigma\phi_X(\phi_X^\intercal\Sigma\phi_X + \sigma^2 I)^{-1}\phi_X^\intercal\Sigma\right)\phi_x\right)$$





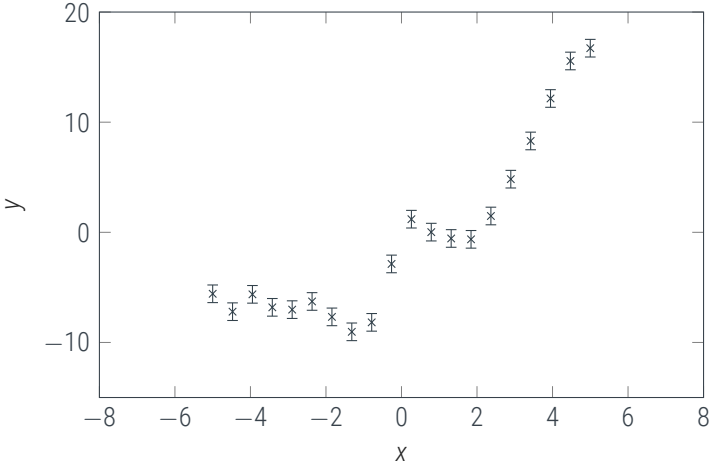
$$
\begin{aligned}
\text{prior} \quad p(w) &= \mathcal{N}(w; \mu, \Sigma) \quad \Rightarrow \quad p(f) = \mathcal{N}(f_x; \phi_x^\intercal \mu, \phi_x \Sigma \phi_x) \\
\text{likelihood} \quad p(y \mid w, \phi_x) &= \mathcal{N}(y; \phi_X^\intercal w, \sigma^2 I) = \mathcal{N}(y; f_X, \sigma^2 I) \\
\text{posterior on } \boldsymbol{w} \quad p(w \mid \boldsymbol{y}, \phi_X) &= \mathcal{N}(w; \mu + \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}(\boldsymbol{y} - \phi_X^\intercal \mu), \\
&\qquad \Sigma - \Sigma\phi_X(\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}\phi_X^\intercal \Sigma) \\
&= \mathcal{N}\Big(w; (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\left(\Sigma^{-1}\mu + \sigma^{-2}\phi_X \boldsymbol{y}\right), \\
&\qquad (\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\Big) \\
\text{posterior on } f \quad p(f_x \mid y, \phi_X) &= \mathcal{N}(f_x; \phi_x^\intercal \mu + \phi_x^\intercal \Sigma \phi_X (\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}(y - \phi_X^\intercal \mu), \\
&\qquad \phi_x^\intercal \Sigma \phi_x - \phi_x^\intercal \Sigma \phi_X (\phi_X^\intercal \Sigma \phi_X + \sigma^2 I)^{-1}\phi_X^\intercal \Sigma \phi_x) \\
&\qquad \mathcal{N}\Big(f_x; \phi_x(\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\left(\Sigma^{-1}\mu + \sigma^{-2}\phi_X \boldsymbol{y}\right), \\
&\qquad \phi_x(\Sigma^{-1} + \sigma^{-2}\phi_X^\intercal \phi_X)^{-1}\phi_x^\intercal\Big)
\end{aligned}
$$

Code

Gaussian_Linear_Regression.ipynb

A more Realistic Dataset

General linear regression

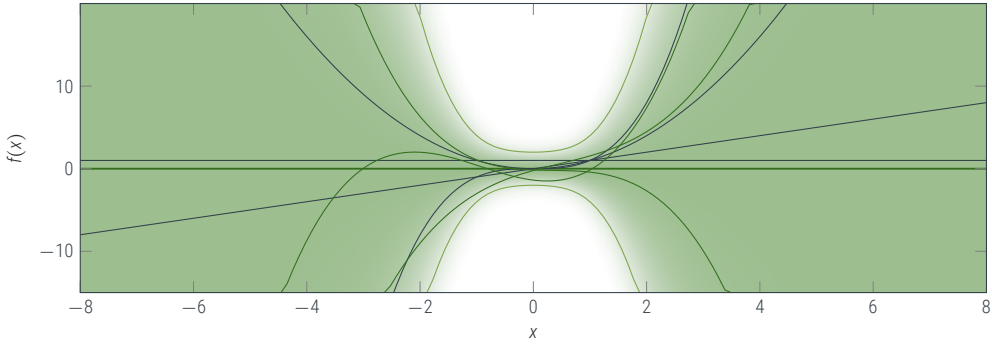


$$f(x) = w_1 + w_2 x = \phi_x^\intercal w$$

$$\phi_x := \begin{bmatrix} 1 \\ x \end{bmatrix}$$

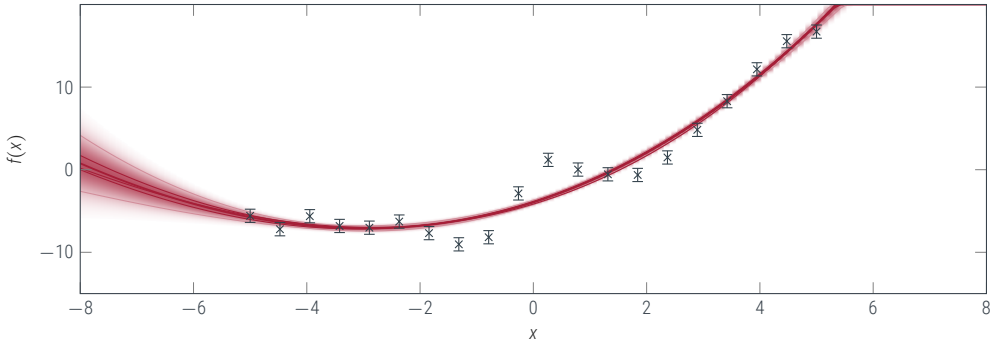
Cubic Regression

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix} 1 & x & x.^2 & x.^3 \end{bmatrix}^\intercal$$



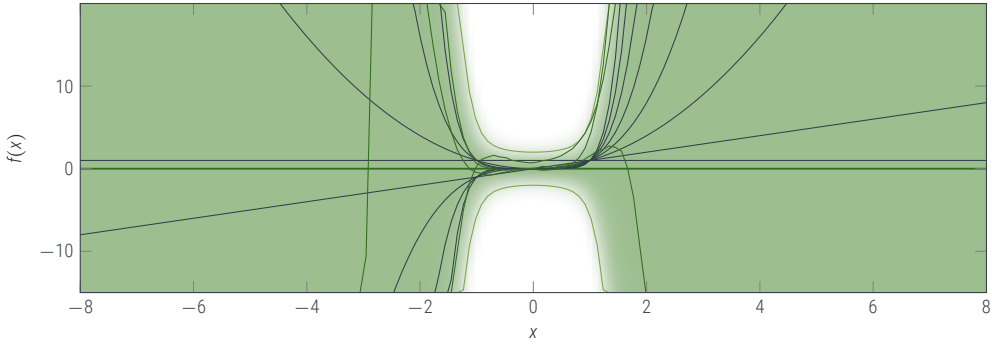
Cubic Regression

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix} 1 & x & x.^2 & x.^3 \end{bmatrix}^\intercal$$



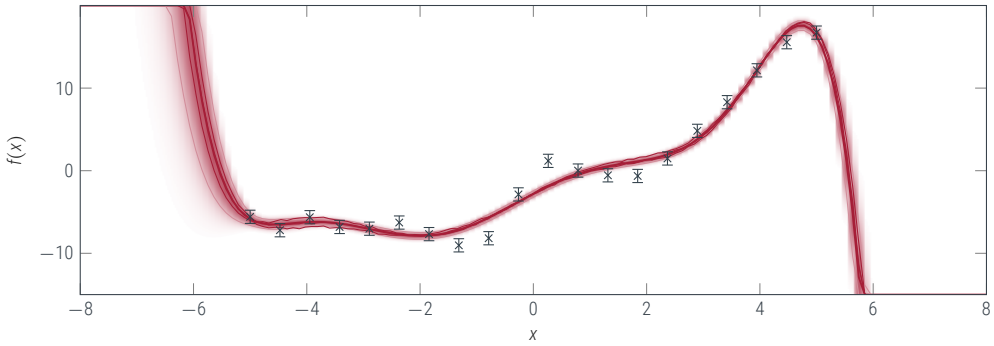
Septic Regression (?)

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix} 1 & x & x.^2 & \cdots & x.^7 \end{bmatrix}^\intercal$$



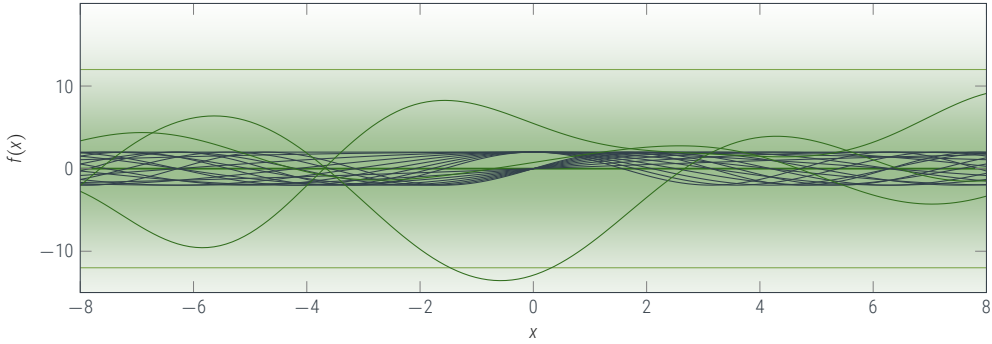
Septic Regression (?)

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix} 1 & x & x.^2 & \cdots & x.^7 \end{bmatrix}^\intercal$$



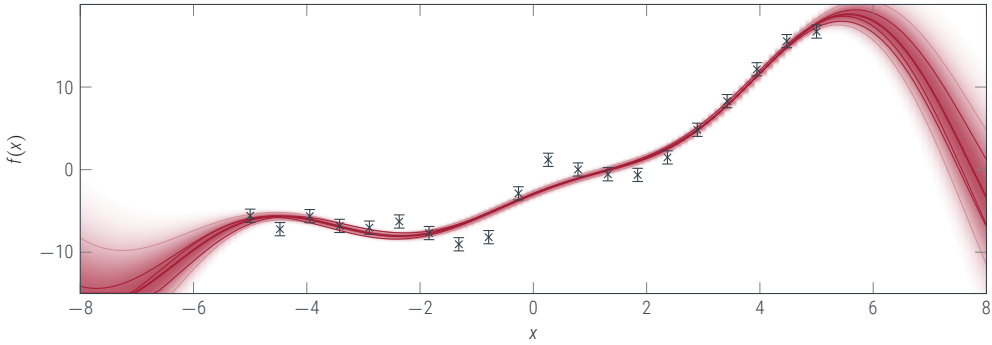
Fourier Regression

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix}\cos(x) & \cos(2x) & \cos(3x) & \dots & \sin(x) & \sin(2x) & \dots\end{bmatrix}^\intercal$$



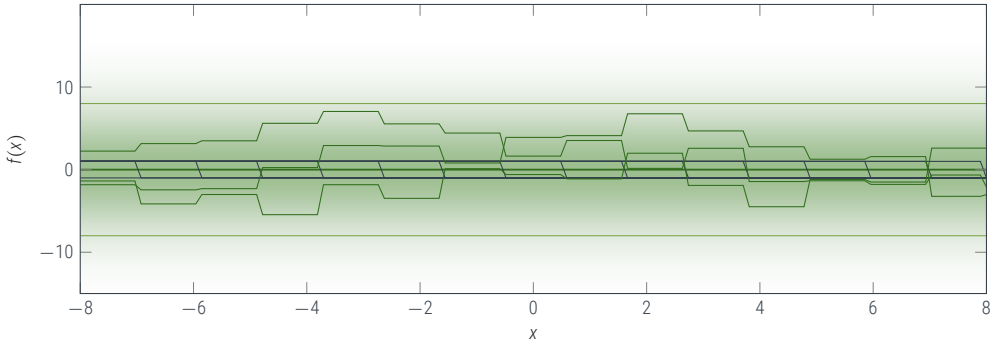
Fourier Regression

$$f(x) = \phi(x)^\intercal w \qquad \phi(x) = \begin{bmatrix}\cos(x) & \cos(2x) & \cos(3x) & \dots & \sin(x) & \sin(2x) & \dots\end{bmatrix}^\intercal$$



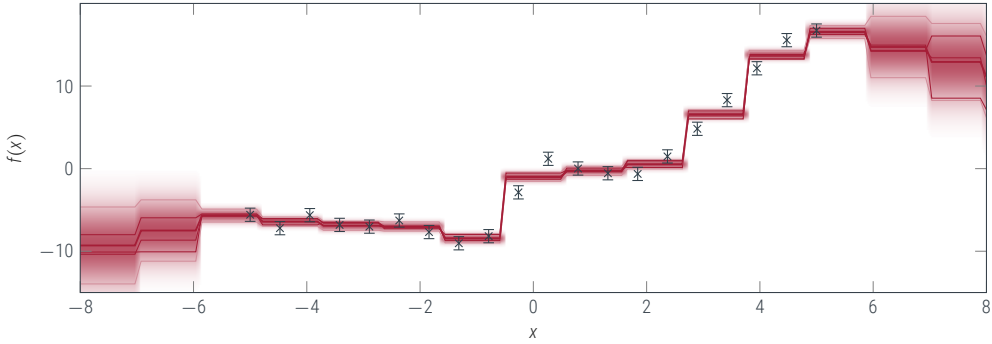
Pixel Regression

$$\phi(x) = -1 + 2\begin{bmatrix}\theta(x-8) & \theta(8-x) & \theta(x-7) & \theta(7-x) & \dots\end{bmatrix}^\intercal$$



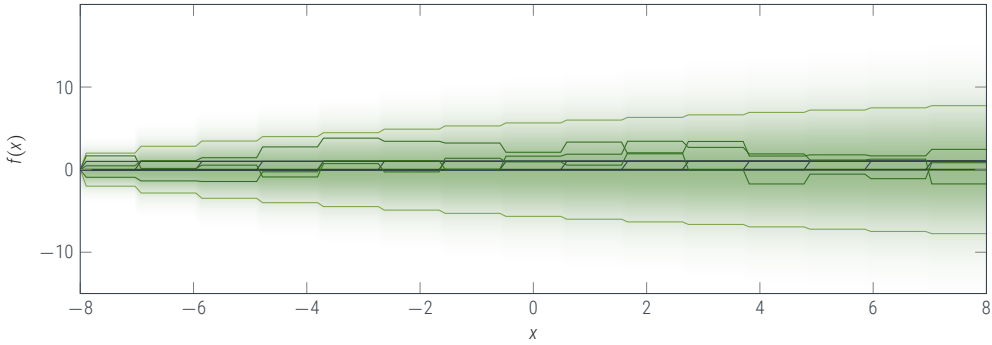
Pixel Regression

$$\phi(x) = -1 + 2\begin{bmatrix}\theta(x-8) & \theta(8-x) & \theta(x-7) & \theta(7-x) & \dots\end{bmatrix}^\intercal$$



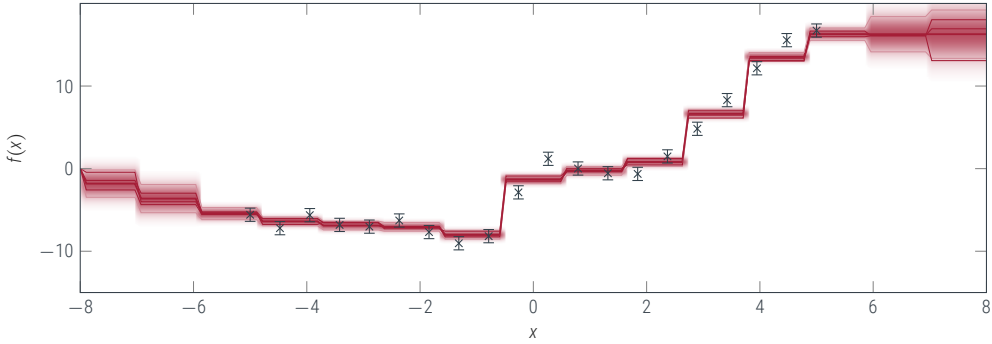
Switch Regression

$$\phi(x) = \begin{bmatrix} \theta(x-8) & \theta(8-x) & \theta(x-7) & \theta(7-x) & \dots \end{bmatrix}^\intercal$$



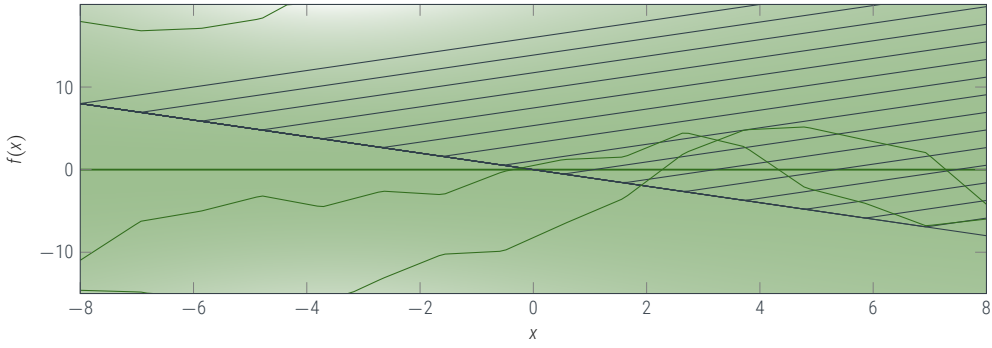
Switch Regression

$$\phi(x) = \begin{bmatrix} \theta(x-8) & \theta(8-x) & \theta(x-7) & \theta(7-x) & \dots \end{bmatrix}^\intercal$$



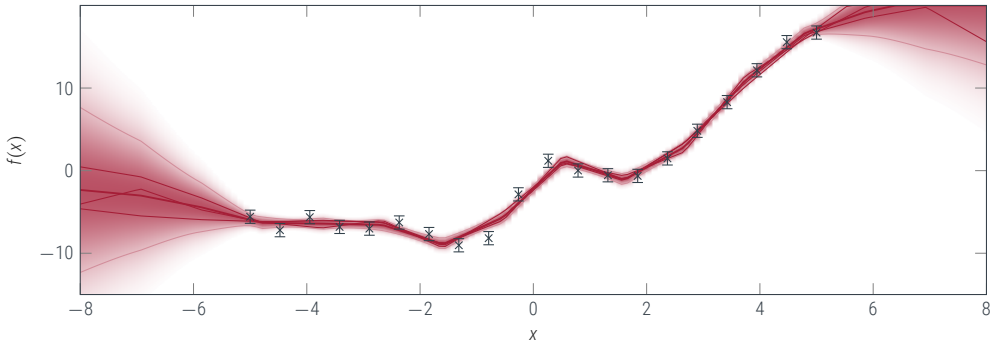
V Regression

$$\phi(x) = \begin{bmatrix} |x-8|-8 & |x-7|-7 & |x-6|-6 & \dots \end{bmatrix}^\intercal$$



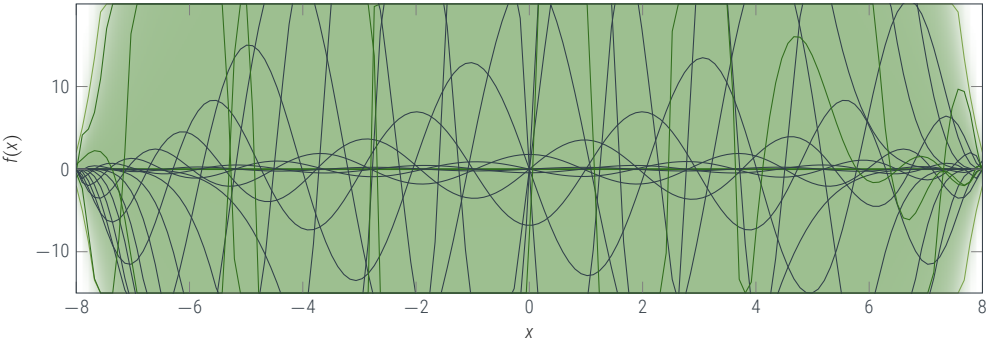
V Regression

$$\phi(x) = \begin{bmatrix} |x-8|-8 & |x-7|-7 & |x-6|-6 & \dots \end{bmatrix}^\intercal$$



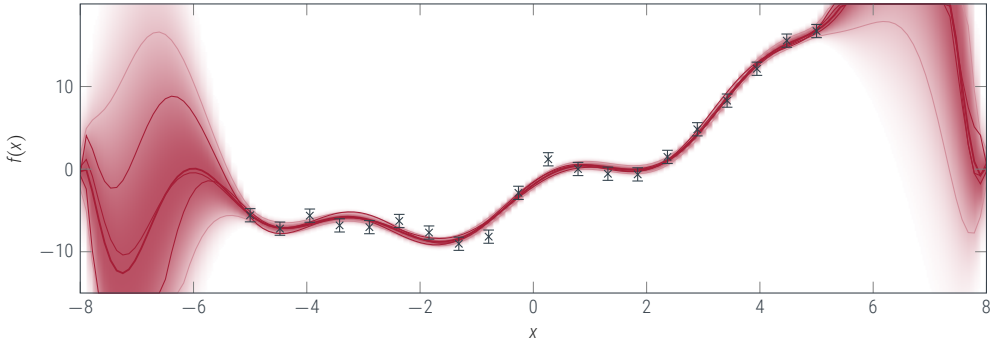
Legendre Regression

$$\phi(x) = \left[b^0 P_0(x), b^1 P_1(x), \ldots, b^{13} P_{13}(x)\right]^\intercal \qquad P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n$$



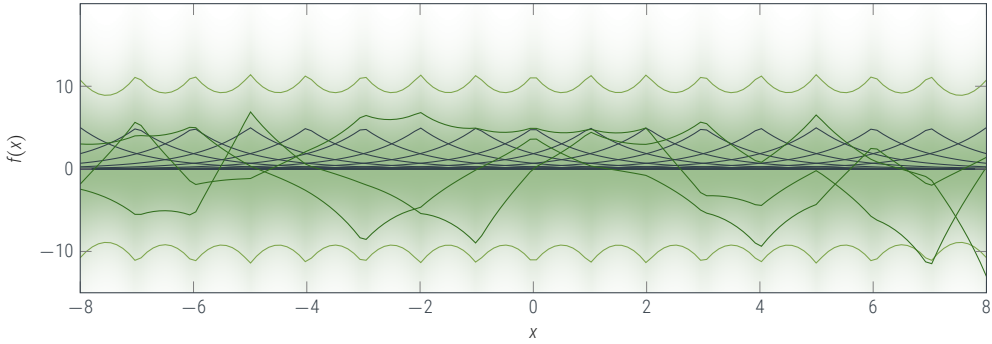
Legendre Regression

$$\phi(x) = \left[b^0 P_0(x), b^1 P_1(x), \ldots, b^{13} P_{13}(x)\right]^\intercal \qquad P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n$$



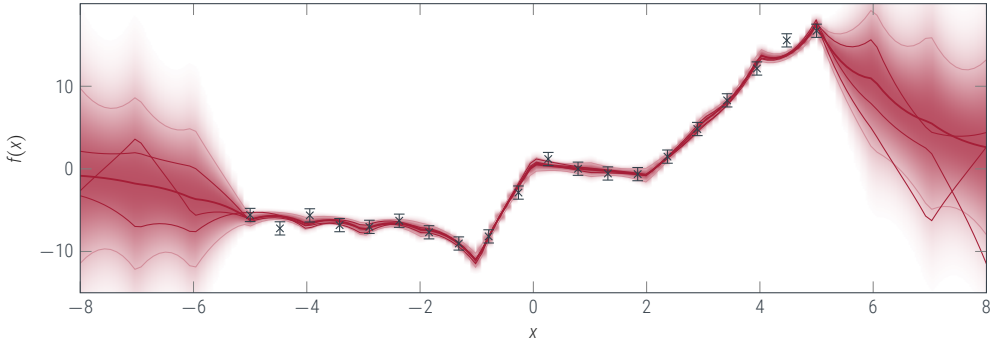
Eiffel Tower Regression

$$\phi(x) = \begin{bmatrix} e^{-|x-8|} & e^{-|x-7|} & e^{-|x-6|} & \dots \end{bmatrix}^\intercal$$



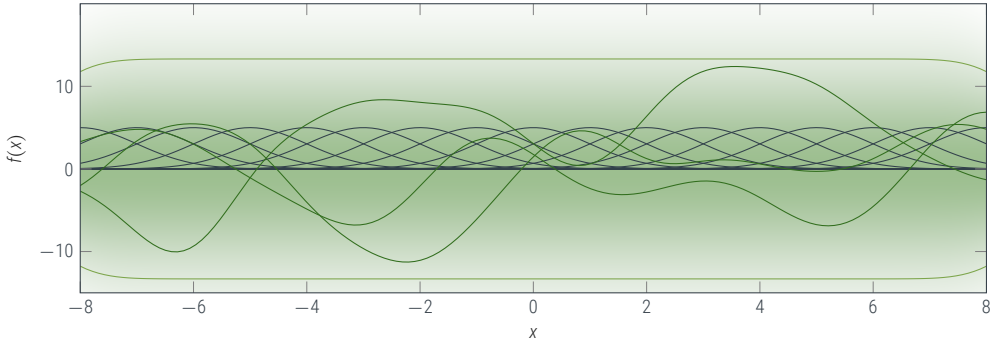
Eiffel Tower Regression

$$\phi(x) = \begin{bmatrix} e^{-|x-8|} & e^{-|x-7|} & e^{-|x-6|} & \dots \end{bmatrix}^\intercal$$



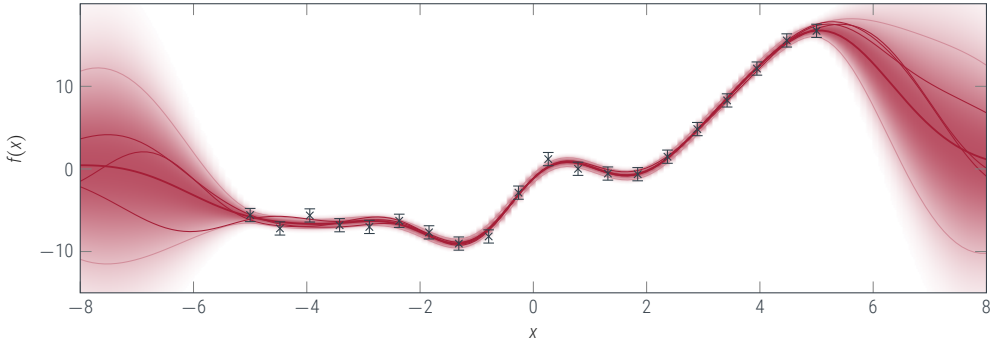
Bell Curve Regression

$$\phi(x) = \begin{bmatrix} e^{-\frac{1}{2}(x-8)^2} & e^{-\frac{1}{2}(x-7)^2} & e^{-\frac{1}{2}(x-6)^2} & \dots \end{bmatrix}^\intercal$$



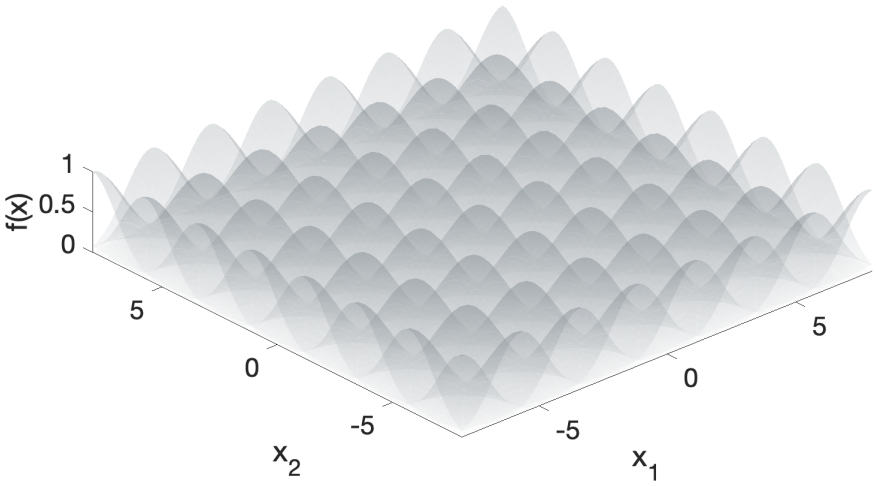
Bell Curve Regression

$$\phi(x) = \begin{bmatrix} e^{-\frac{1}{2}(x-8)^2} & e^{-\frac{1}{2}(x-7)^2} & e^{-\frac{1}{2}(x-6)^2} & \dots \end{bmatrix}^\intercal$$



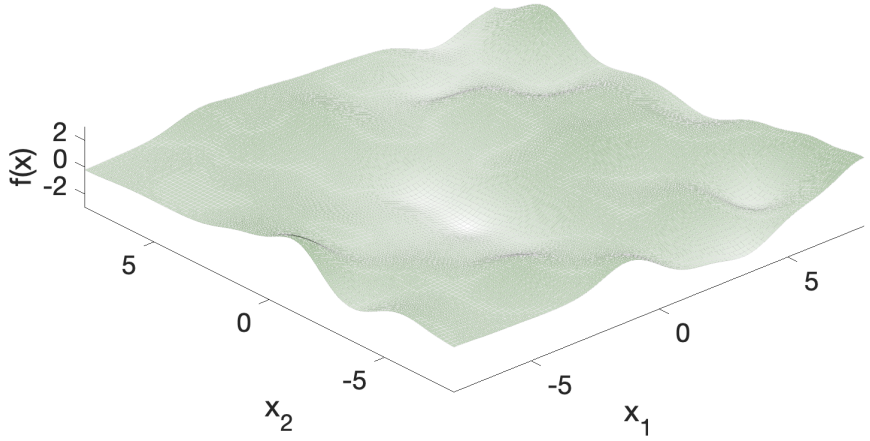
Multiple Inputs

Input domain $\mathbb{X}$ can be anything



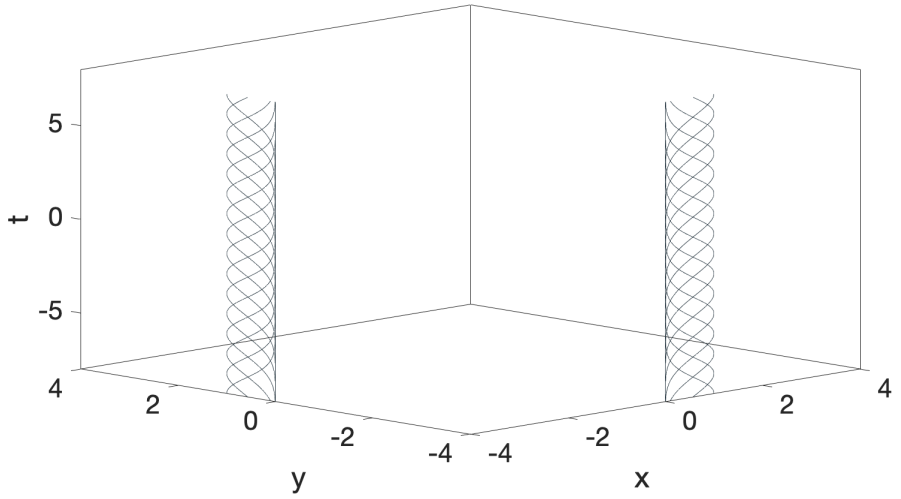
Multiple Inputs

Input domain $\mathbb{X}$ can be anything



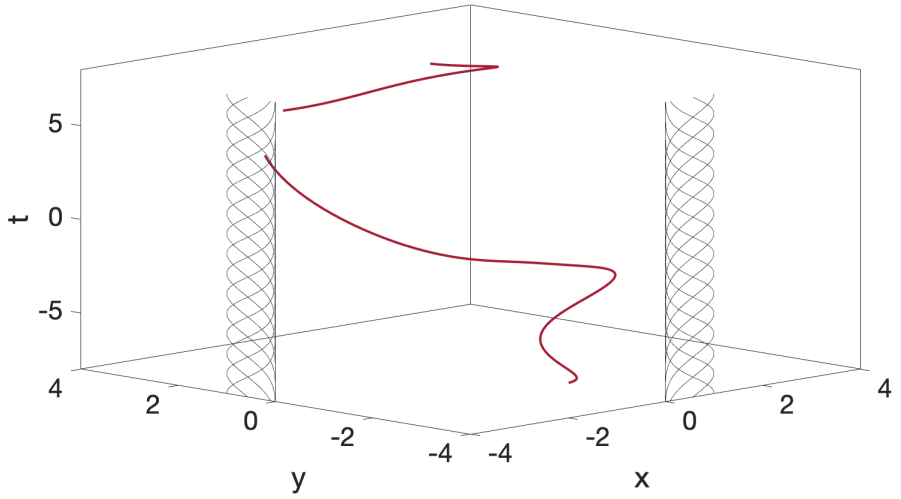
Multiple Outputs

The output domain can be anything isomorphic to $\mathbb{R}^N$



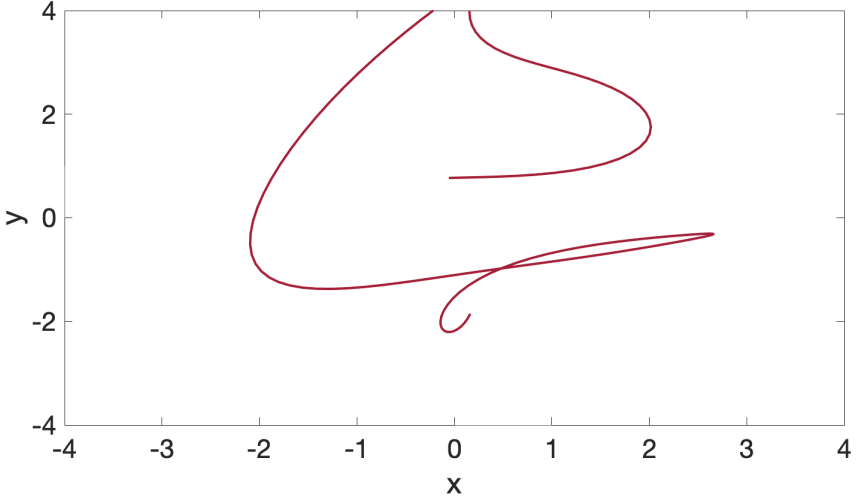
Multiple Outputs

The output domain can be anything isomorphic to $\mathbb{R}^N$



Multiple Outputs

The output domain can be anything isomorphic to $\mathbb{R}^N$



Summary:

- Gaussian distributions can be used to learn functions
- Analytical inference is possible using general linear models

$$f(x) = \phi(x)^\intercal \boldsymbol{w} = \phi_x^\intercal \boldsymbol{w}$$

- Then the posterior on both w and f is Gaussian
- The choice of features $\phi : \mathbb{X} \to \mathbb{R}$ is essentially unconstrained