Advanced Machine Learning

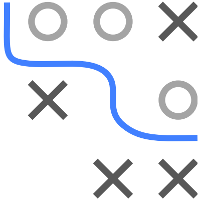
Gaussian Processes: Distribution on Functions

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

Learning goals

- How to model distributions over discrete functions
- The role of the covariance function

DISCRETE FUNCTIONS



For simplicity, let us consider functions with finite domains first.

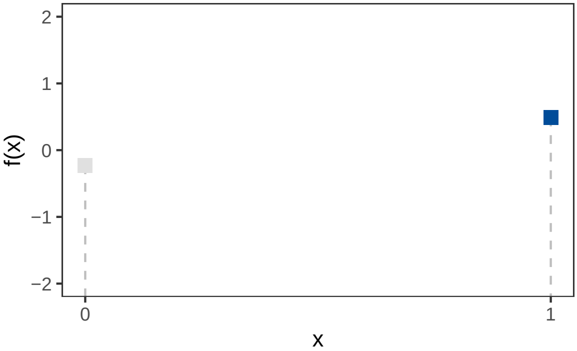
Let $\mathcal{X} = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\}$ be a finite set of elements and $\mathcal{H}$ the set of all functions from $\mathcal{X} \to \mathbb{R}$.
Remark: $\mathcal{X}$ does not mean the training data here but means the “real” domain of the functions.

Since the domain of any $f(.) \in \mathcal{H}$ has only n elements, we can represent the function $f(.)$ compactly as a n-dimensional vector

$$\boldsymbol{f} = \left[f\left(\mathbf{x}^{(1)}\right), \ldots, f\left(\mathbf{x}^{(n)}\right)\right].$$

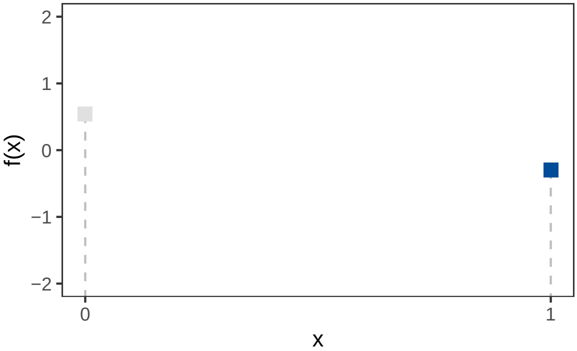
DISCRETE FUNCTIONS

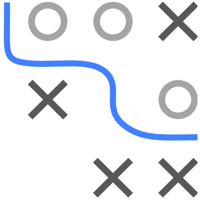
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



DISCRETE FUNCTIONS

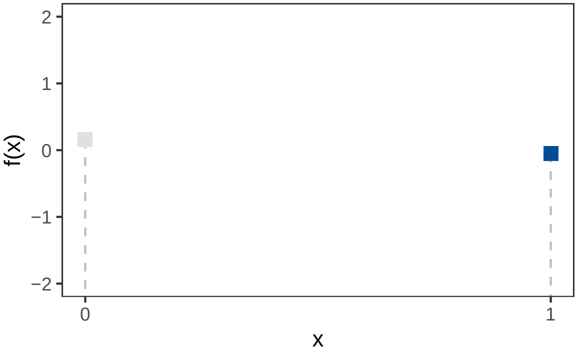
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



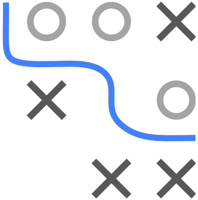


DISCRETE FUNCTIONS

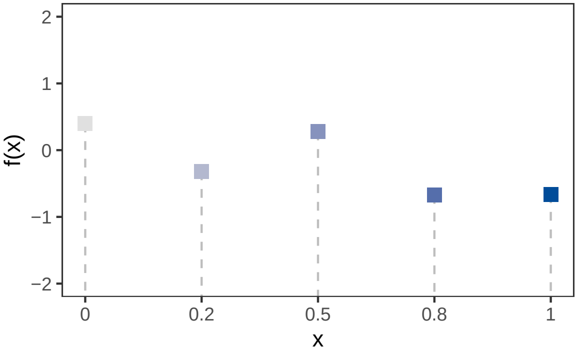
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



DISCRETE FUNCTIONS

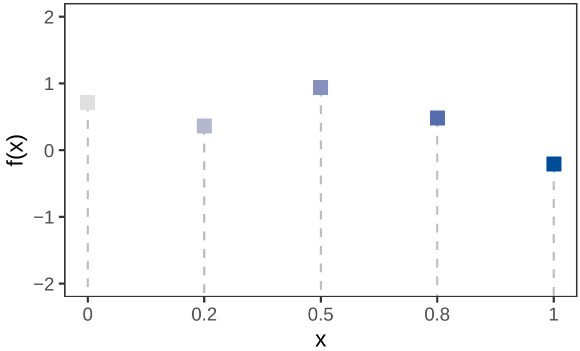


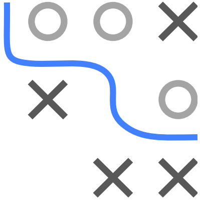
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



DISCRETE FUNCTIONS

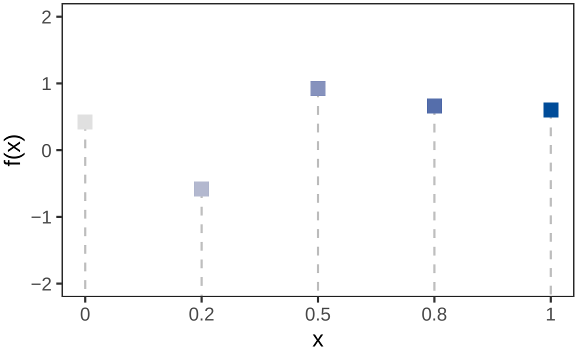
Some examples $f : \mathcal{X} \rightarrow \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:





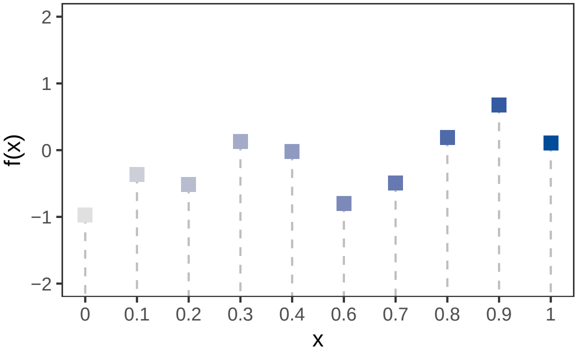
DISCRETE FUNCTIONS

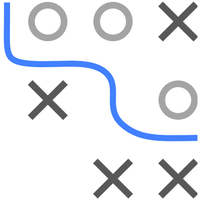
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



DISCRETE FUNCTIONS

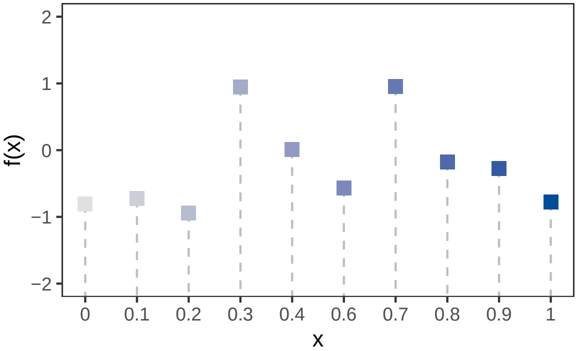
Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:





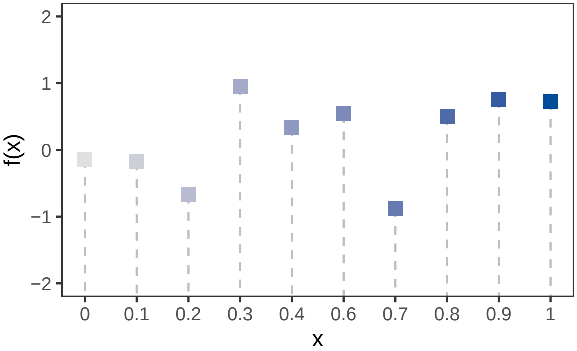
DISCRETE FUNCTIONS

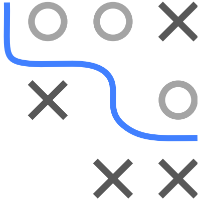
Some examples $f : \mathcal{X} \rightarrow \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:



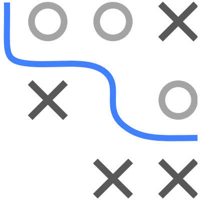
DISCRETE FUNCTIONS

Some examples $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is univariate and finite:





DISTRIBUTIONS ON DISCRETE FUNCTIONS



One natural way to specify a probability function on a discrete function $f \in \mathcal{H}$ is to use the vector representation

$$\boldsymbol{f} = \left[f\left(\mathbf{x}^{(1)}\right), f\left(\mathbf{x}^{(2)}\right), \ldots, f\left(\mathbf{x}^{(n)}\right) \right]$$

of the function.

Let us see $\boldsymbol{f}$ as a n-dimensional random variable. We will further assume the following normal distribution:

$$\boldsymbol{f} \sim \mathcal{N}\left(\boldsymbol{m}, \boldsymbol{K}\right).$$

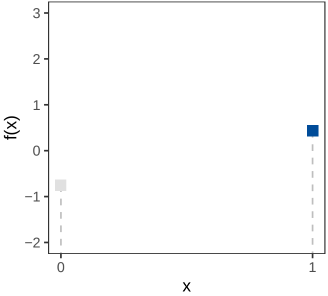
Note: For now, we set $\boldsymbol{m} = \mathbf{0}$ and take the covariance matrix $\boldsymbol{K}$ as given. We will see later how they are chosen / estimated.

DISCRETE FUNCTIONS

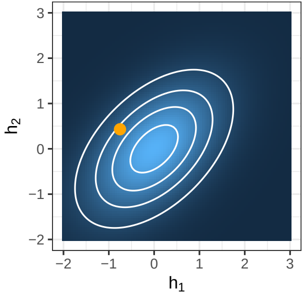
Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a two-dimensional normal variable.

$$\boldsymbol{f} = [f(1), f(2)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$

Sample Function 1, n = 2



Density of a 2−D Gaussian

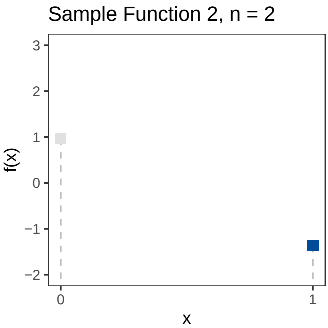


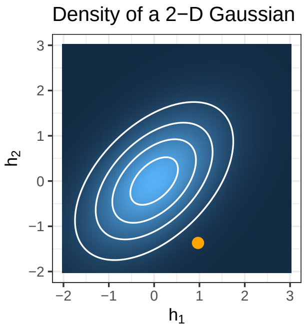
In this example, $m = (0, 0)$ and $K = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

DISCRETE FUNCTIONS

Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a two-dimensional normal variable.

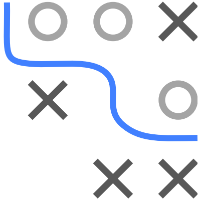
$$\boldsymbol{f} = [f(1), f(2)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$





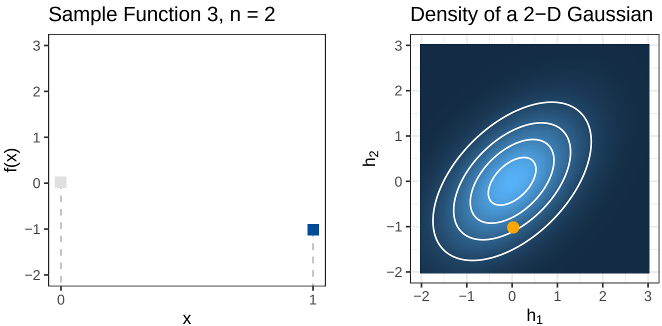
In this example, $m = (0, 0)$ and $K = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

DISCRETE FUNCTIONS



Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a two-dimensional normal variable.

$$\boldsymbol{f} = [f(1), f(2)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$

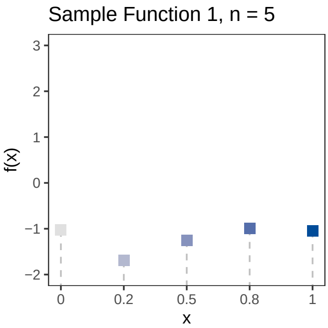


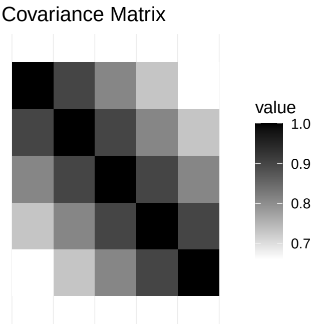
In this example, $m = (0, 0)$ and $K = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

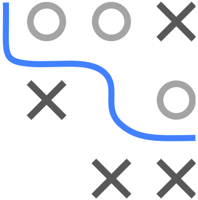
DISCRETE FUNCTIONS

Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a five-dimensional normal variable.

$$\boldsymbol{f} = [f(1), f(2), f(3), f(4), f(5)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$



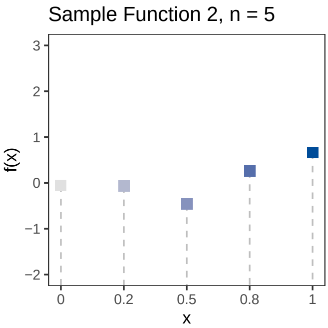


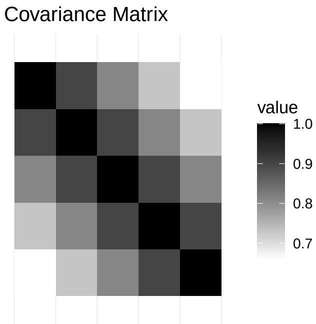


DISCRETE FUNCTIONS

Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a five-dimensional normal variable.

$$\boldsymbol{f} = [f(1), f(2), f(3), f(4), f(5)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$

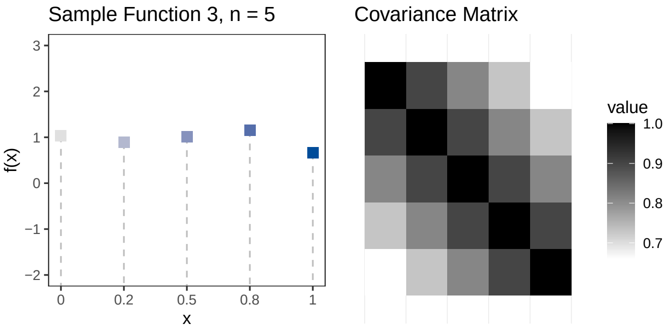




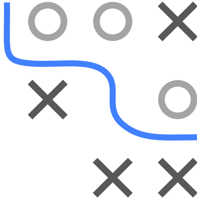
DISCRETE FUNCTIONS

Let $f : \mathcal{X} \to \mathbb{R}$. Sample functions by sampling from a five-dimensional normal variable.

$$\boldsymbol{f} = [f(1), f(2), f(3), f(4), f(5)] \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$



ROLE OF THE COVARIANCE FUNCTION



- "Meaningful" functions (on a numeric space $\mathcal{X}$) may be characterized by a spatial property:

  > If two points $\mathbf{x}^{(i)}, \mathbf{x}^{(j)}$ are close in $\mathcal{X}$-space, their function values $f(\mathbf{x}^{(i)}), f(\mathbf{x}^{(j)})$ should be close in $\mathcal{Y}$-space.

  In other words: If they are close in $\mathcal{X}$-space, their functions values should be **correlated**!

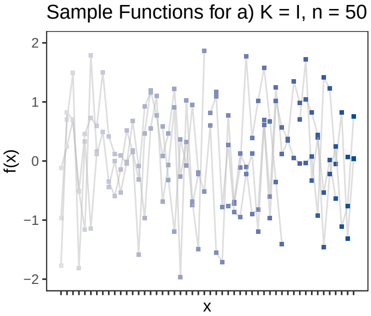
- We can enforce that by choosing a covariance function with

$$\boldsymbol{K}_{ij} \text{ high, if } \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \text{ close.}$$

ROLE OF THE COVARIANCE FUNCTION / 2

Covariance controls the “shape” of the drawn function. Consider cases of varying correlation structure

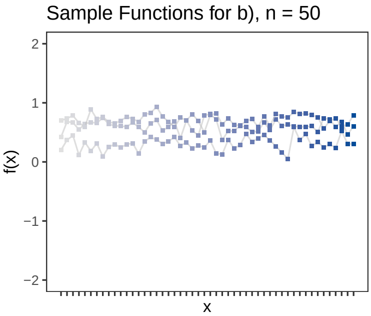
a) uncorrelated: $\boldsymbol{K} = \boldsymbol{I}$.



Points are uncorrelated. We sample white noise.

ROLE OF THE COVARIANCE FUNCTION

b) Correlation almost 1: $\boldsymbol{K} = \begin{pmatrix} 1 & 0.99 & \dots & 0.99 \\ 0.99 & 1 & \dots & 0.99 \\ 0.99 & 0.99 & \ddots & 0.99 \\ 0.99 & \dots & 0.99 & 1 \end{pmatrix}.$

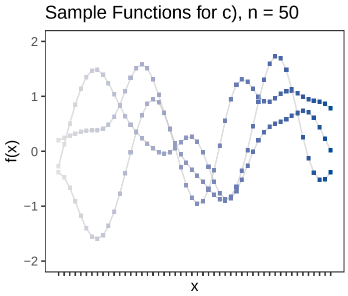


Points are highly correlated. Functions become very smooth and flat.

ROLE OF THE COVARIANCE FUNCTION / 2

- We can compute the entries of the covariance matrix by a function that is based on the distance between $\mathbf{x}^{(i)}, \mathbf{x}^{(j)}$, for example:

   c) Spatial correlation: $K_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp\left(-\frac{1}{2}\left|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right|^2\right)$



Function exhibit interesting, variable shape.

NB: $k(\cdot, \cdot)$ is called **covar. function** or **kernel**, we will study it more later.