# Bayesian Optimization in Machine Learning

José Jiménez

Master Thesis
Master's degree in Statistics and Operations Research

supervised by Josep Ginebra[1]
[1] Department of Statistics and Operations Research. ETSEIB.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

UNIVERSITAT DE BARCELONA

# Goals of this thesis

This Master's thesis aims to be a multi-objective optimization task:

- Provide an introduction to both Gaussian Process regression and Bayesian optimization.
- Show that the Bayesian Optimization framework works in several real-world machine learning tasks.
- Write a complete software package (pyGPGO) for users to apply Bayesian Optimization in their research.

# Goals of this thesis

This Master's thesis aims to be a multi-objective optimization task:

- Provide an introduction to both Gaussian Process regression and Bayesian optimization.
- Show that the Bayesian Optimization framework works in several real-world machine learning tasks.
- Write a complete software package (pyGPGO) for users to apply Bayesian Optimization in their research.

# Goals of this thesis

This Master's thesis aims to be a multi-objective optimization task:

- Provide an introduction to both Gaussian Process regression and Bayesian optimization.
- Show that the Bayesian Optimization framework works in several real-world machine learning tasks.
- Write a complete software package (pyGPGO) for users to apply Bayesian Optimization in their research.

# Goals of this thesis

This Master's thesis aims to be a multi-objective optimization task:

- Provide an introduction to both Gaussian Process regression and Bayesian optimization.
- Show that the Bayesian Optimization framework works in several real-world machine learning tasks.
- Write a complete software package (pyGPGO) for users to apply Bayesian Optimization in their research.

# Organization of the work

Organized in 5 self-contained chapters.

- **Chapter 2** focuses on an introduction to regression problems using Gaussian Processes. These are surrogate models we will use for Bayesian Optimization.
- **Chapter 3** covers the main topic in this work, Bayesian Optimization.
- **Chapter 4** presents experiments using the Bayesian Optimization framework. Mostly mid-sized supervised-learning problems.
- **Chapter 5** provides technical explanations for pyGPGO, the software developed alongside this manual.

# Organization of the work

Organized in 5 self-contained chapters.

- **Chapter 2** focuses on an introduction to regression problems using Gaussian Processes. These are surrogate models we will use for Bayesian Optimization.
- **Chapter 3** covers the main topic in this work, Bayesian Optimization.
- **Chapter 4** presents experiments using the Bayesian Optimization framework. Mostly mid-sized supervised-learning problems.
- **Chapter 5** provides technical explanations for pyGPGO, the software developed alongside this manual.

# Organization of the work

Organized in 5 self-contained chapters.

- **Chapter 2** focuses on an introduction to regression problems using Gaussian Processes. These are surrogate models we will use for Bayesian Optimization.
- **Chapter 3** covers the main topic in this work, Bayesian Optimization.
- **Chapter 4** presents experiments using the Bayesian Optimization framework. Mostly mid-sized supervised-learning problems.
- **Chapter 5** provides technical explanations for pyGPGO, the software developed alongside this manual.

# Organization of the work

Organized in 5 self-contained chapters.

- **Chapter 2** focuses on an introduction to regression problems using Gaussian Processes. These are surrogate models we will use for Bayesian Optimization.
- **Chapter 3** covers the main topic in this work, Bayesian Optimization.
- **Chapter 4** presents experiments using the Bayesian Optimization framework. Mostly mid-sized supervised-learning problems.
- **Chapter 5** provides technical explanations for pyGPGO, the software developed alongside this manual.

# Organization of the work

Organized in 5 self-contained chapters.

- **Chapter 2** focuses on an introduction to regression problems using Gaussian Processes. These are surrogate models we will use for Bayesian Optimization.
- **Chapter 3** covers the main topic in this work, Bayesian Optimization.
- **Chapter 4** presents experiments using the Bayesian Optimization framework. Mostly mid-sized supervised-learning problems.
- **Chapter 5** provides technical explanations for pyGPGO, the software developed alongside this manual.

# Gaussian Process Regression: basic definitions

**Definition 1.**

A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution. This process is defined by two functions. Its *mean function*:

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{1}$$

and its *covariance function*:

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))\,(f(\mathbf{x}') - m(\mathbf{x}'))\right] \tag{2}$$

We say that $f$ is a Gaussian Process with mean $m(\mathbf{x})$ and covariance function $k(x, x')$ and write:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{3}$$

# Gaussian Process Regression: basic definitions

**Definition 1.**

A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.This process is defined by two functions. Its *mean function*:

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{1}$$

and its *covariance function*:

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))\,(f(\mathbf{x}') - m(\mathbf{x}'))\right] \tag{2}$$

We say that $f$ is a Gaussian Process with mean $m(\mathbf{x})$ and covariance function $k(x, x')$ and write:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{3}$$

# Gaussian Process Regression: basic definitions

**Definition 1.**

A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.This process is defined by two functions. Its *mean function*:

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{1}$$

and its *covariance function*:

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))\,(f(\mathbf{x}') - m(\mathbf{x}'))\right] \tag{2}$$

We say that $f$ is a Gaussian Process with mean $m(\mathbf{x})$ and covariance function $k(x, x')$ and write:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{3}$$

# Gaussian Process Regression: basic definitions

**Definition 1.**

A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.This process is defined by two functions. Its *mean function*:

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{1}$$

and its *covariance function*:

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))\left(f(\mathbf{x}') - m(\mathbf{x}')\right)\right] \tag{2}$$

We say that $f$ is a Gaussian Process with mean $m(\mathbf{x})$ and covariance function $k(x, x')$ and write:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{3}$$

# Gaussian Process Regression: basic definitions

Define then a covariance function, such as the *squared exponential* kernel:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}|\mathbf{x} - \mathbf{x}'|^2\right) \tag{4}$$

where $|.|$ denotes the standard $L_2$ norm. Drawing samples from a Gaussian Process, assuming zero mean, for given finite inputs $X_*$ simplifies to sampling from:

$$\mathbf{f}_* \sim \mathcal{N}\left(\mathbf{0}, K(X_*, X_*)\right) \tag{5}$$

# Gaussian Process Regression: prediction

Assume training data $\mathcal{D} = \{(\mathbf{x_i}, y_i) \,|\, i = 1, \ldots, n\}$

## Prediction using GP prior

Let $\mathbf{y}$ and $\mathbf{f}_*$ be jointly Gaussian:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{6}$$

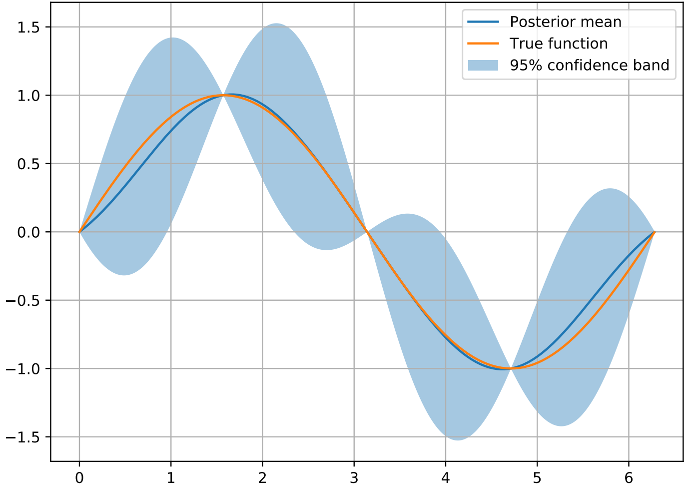
We want to condition $\mathbf{f}_*$ over $\mathbf{y}$.

$$\mathbf{f}_* | \mathbf{y} \sim \mathcal{N}(\overline{\mathbf{f}_*}, Cov(\mathbf{f}_*)) \tag{7}$$

where:

$$\begin{aligned} \overline{\mathbf{f}_*} &= K(X_*,X)\left(K(X,X) + \sigma_n^2 I\right)^{-1} \mathbf{y} \\ Cov(\mathbf{f}_*) &= K(X_*,X_*) - K(X_*,X)\left(K(X,X) + \sigma_n^2 I\right)^{-1} K(X,X_*) \end{aligned} \tag{8}$$

# Gaussian Process Regression: an example



# Gaussian Process Regression: on covariance functions

## Covariance function choices

$$k_{SE}(r) = \exp\left(-\frac{r^2}{2l^2}\right) \tag{9}$$

$$k_{\mathrm{Matern}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}r}{l}\right) \tag{10}$$

$$k_{\mathrm{GE}}(r) = \exp\left(-\left(\frac{r}{l}\right)^{\gamma}\right) \tag{11}$$