

# BASS como modelo base para optimización Bayesiana

Adrian Tame Jacobo

Temas Selectos de Estadística

3 de junio de 2022

## 1 Optimización Bayesiana

## 2 BASS

Buscamos resolver el problema

$$\max_{x \in A} f(x),$$

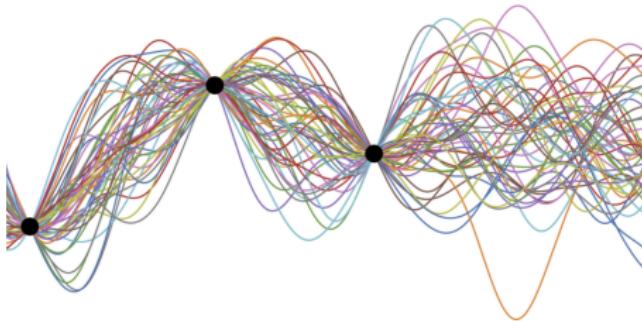
donde:

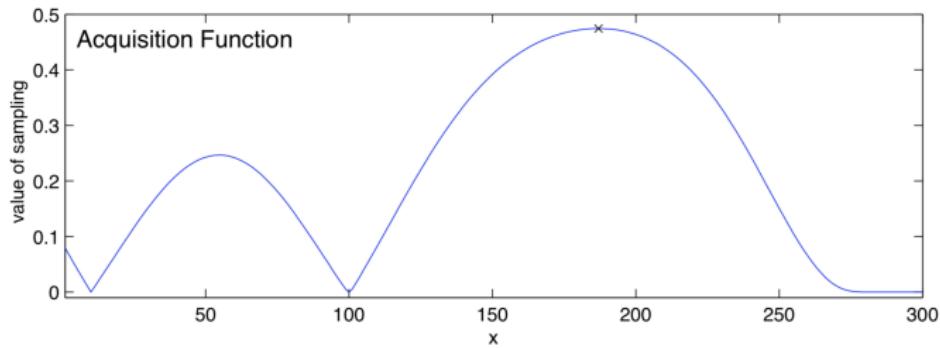
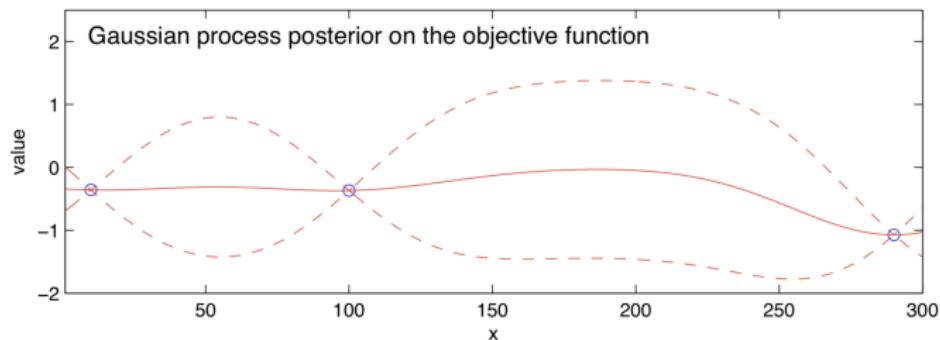
- No tenemos información de las derivadas de  $f$ , ni información sobre su linealidad, concavidad, etc. La llamamos una función “caja negra”.
- Tenemos que si  $x \in \mathbb{R}^d$ , entonces  $d \lesssim 20$ .
- *La función  $f$  es continua.*
- $f$  es difícil de evaluar.

Sean  $x_1, x_2, \dots, x_k \in A$  puntos en los que evaluamos la función  $f$ . Sea  $[f(x_1), \dots, f(x_k)]$  un vector de estas evaluaciones.

La distribución resultante o proceso Gausiano asociado sería entonces:

$$f(x_{1:k}) \sim \mathcal{N}(\mu_0(x_{1:k}), \Sigma_0(x_{1:k}, x_{1:k})).$$





---

**Algorithm 1** Basic pseudo-code for Bayesian optimization

---

Place a Gaussian process prior on  $f$

Observe  $f$  at  $n_0$  points according to an initial space-filling experimental design. Set  $n = n_0$ .

**while**  $n \leq N$  **do**

    Update the posterior probability distribution on  $f$  using all available data

    Let  $x_n$  be a maximizer of the acquisition function over  $x$ , where the acquisition function is computed using the current posterior distribution.

    Observe  $y_n = f(x_n)$ .

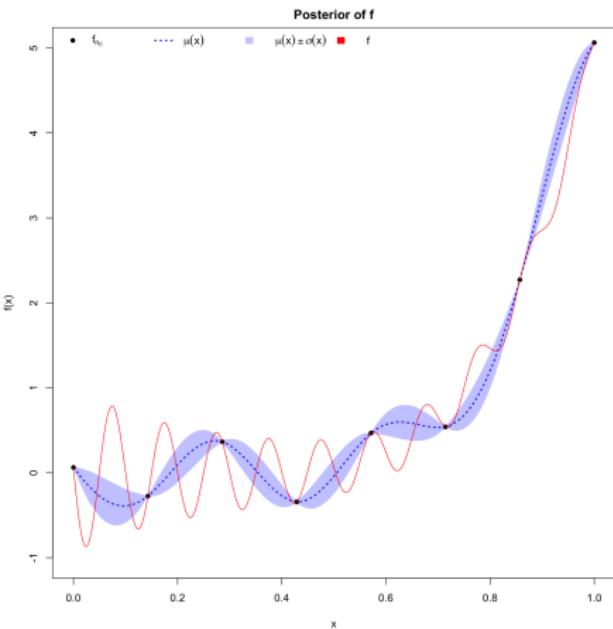
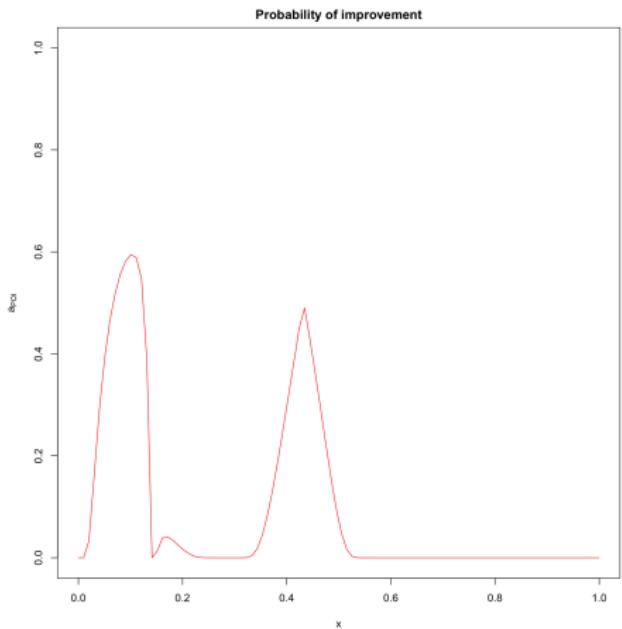
    Increment  $n$

**end while**

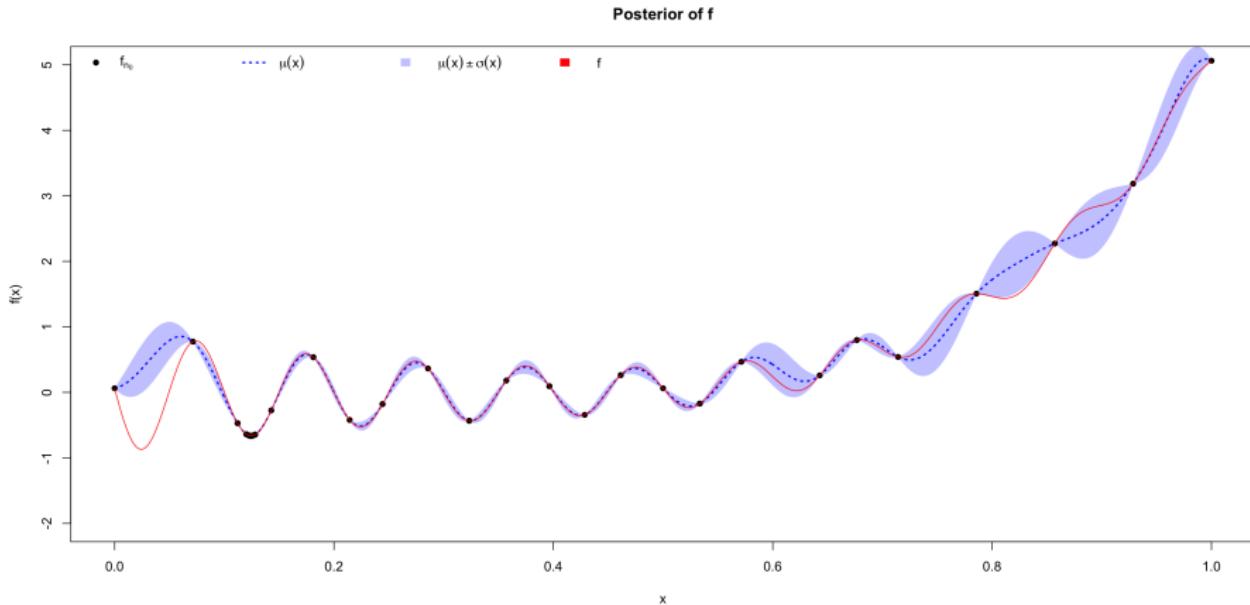
Return a solution: either the point evaluated with the largest  $f(x)$ , or the point with the largest posterior mean.

---

# Posterior y probabilidad de mejora



# Después de 45 iteraciones



1 Optimización Bayesiana

2 BASS

Los modelos BASS son una extensión de los modelos MARS, que hacen regresión lineal a pedazos. BASS se basa en RJMCMC para calcular la mejor distribución.

Modelamos  $y_i$  como sigue:

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$f(\mathbf{x}) = a_0 + \sum_{m=1}^M a_m B_m(\mathbf{x})$$

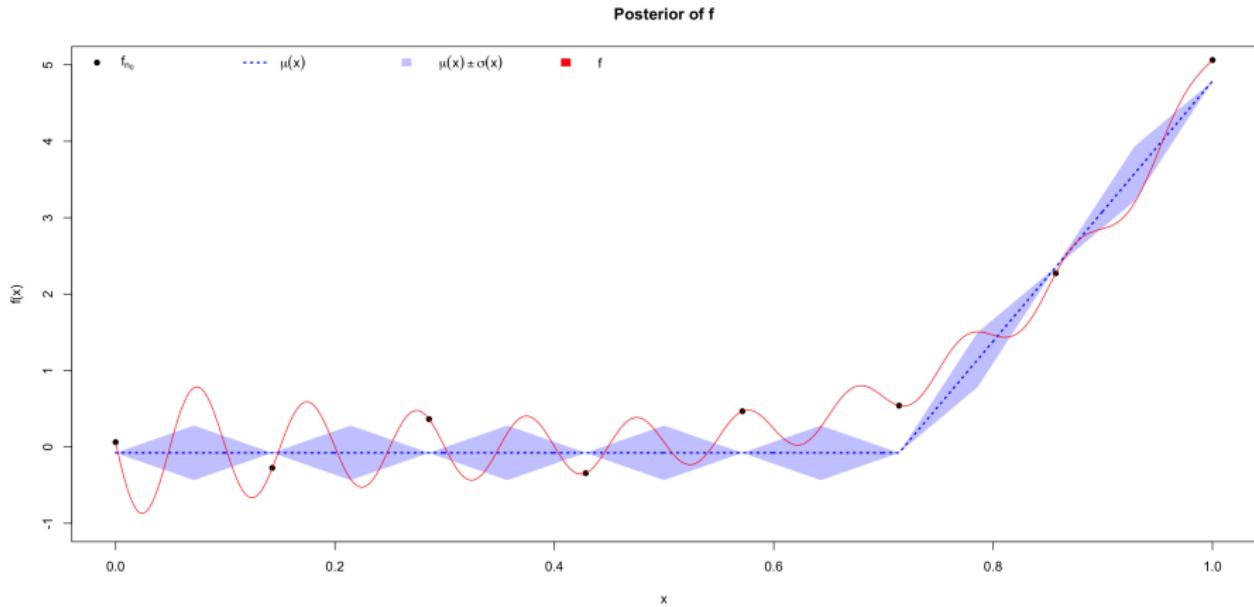
$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} g_{km} [s_{km}(\mathbf{x}_{v_{km}} - t_{km})]_+^\alpha$$

Usamos BASS porque:

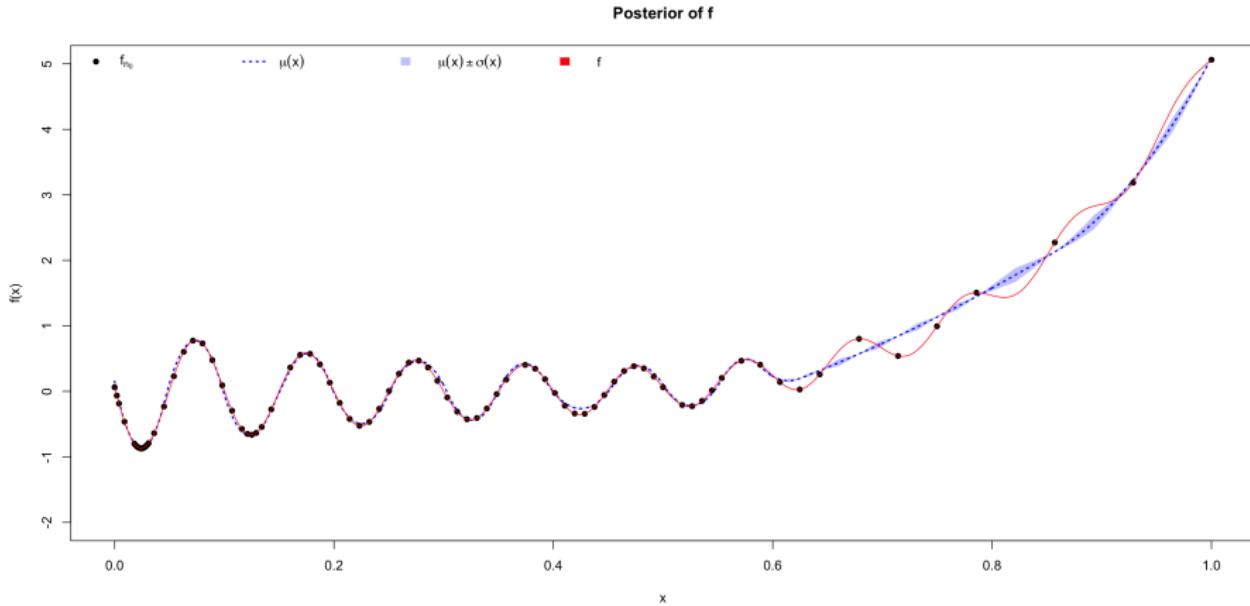
- No es necesario que  $f$  sea continua, puede ser discreta o categórica.
- Puede ser lineal, o de grado  $n$ .
- Mucho más ajustable que procesos Gausianos.

Trade-offs:

- Procesos Gausianos es óptimo.
- Mucho más lento.



# BASS después de 115 iteraciones



The screenshot shows the Kaggle competition page for "House Prices - Advanced Regression Techniques". At the top left is a small icon of a house with a roof and a chimney. Next to it is the text "GettingStarted Prediction Competition". Below this is the main title "House Prices - Advanced Regression Techniques" in large bold letters. Underneath the title is a subtitle "Predict sales prices and practice feature engineering, RFs, and gradient boosting". To the left of the subtitle is a blue "k" icon followed by the text "Kaggle · 4,346 teams · Ongoing". Below the title and subtitle are several navigation links: "Overview" (underlined), "Data", "Code", "Discussion", "Leaderboard", and "Rules". To the right of these links is a dark button with white text that says "Join Competition". At the bottom right of the page is a small ellipsis "...".

The screenshot shows the "Overview" page of the competition. On the left is a sidebar with links: "Description" (which is currently selected and highlighted in blue), "Evaluation", "Tutorials", and "Frequently Asked Questions". The main content area starts with a section titled "Start here if..." which contains text about the competition being suitable for students with basic machine learning experience. Below this is a section titled "Competition Description" which features a colorful illustration of a row of houses with different colors and styles, set against a backdrop of green trees and mountains. A small car is parked in front of one of the houses.

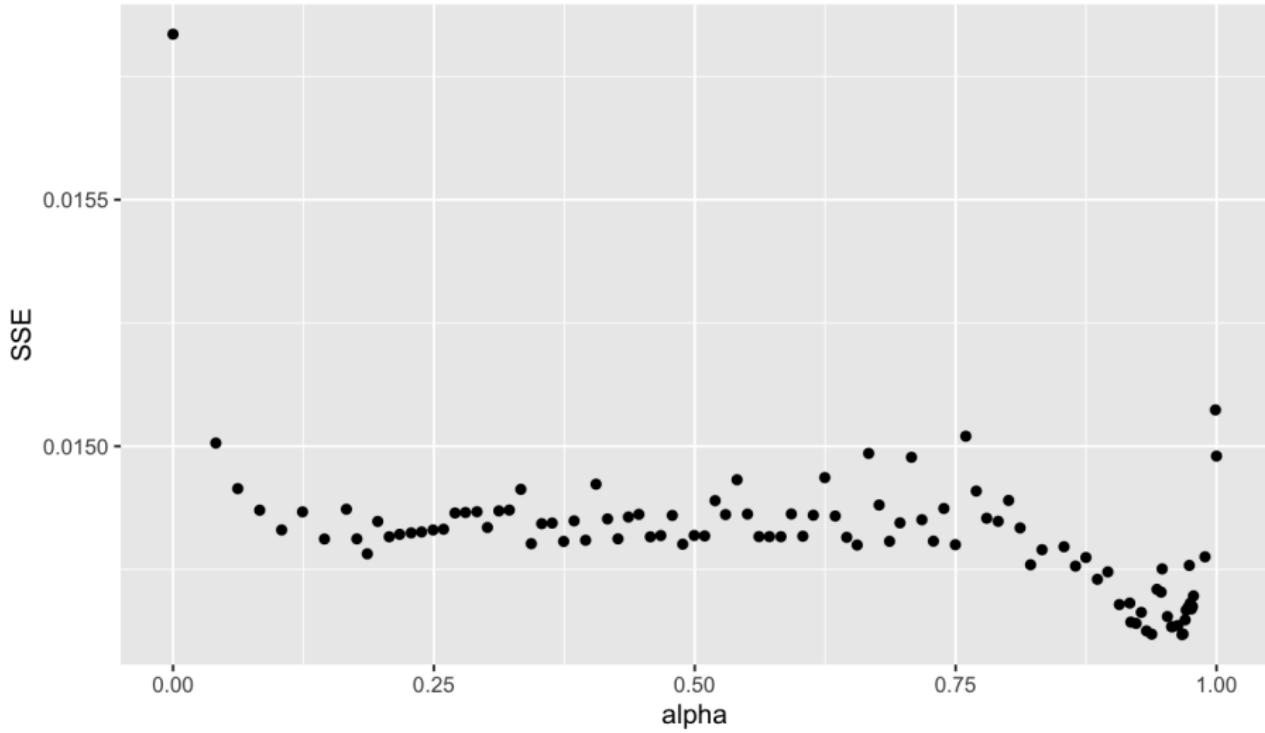
Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the

Queremos ajustar un modelo *elastic net*, de regresión regularizada, cuyos coeficientes se ven como sigue:

$$\hat{\beta} = \arg \min_{\beta} \|y - \mathbf{X}\beta\|^2 + \lambda ((1 - \alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1).$$

Sabemos que  $\alpha \in [0, 1]$ , pero, ¿cuál es la mejor  $\alpha$  para minimizar el error?

Sum of squared errors for different values of alpha



Sum of squared errors for different iterations

