BASS como modelo base para optimización Bayesiana

Adrian Tame Jacobo

Temas Selectos de Estadística

3 de junio de 2022

Tabla de contenidos

Introducción

Buscamos resolver el problema

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

donde:

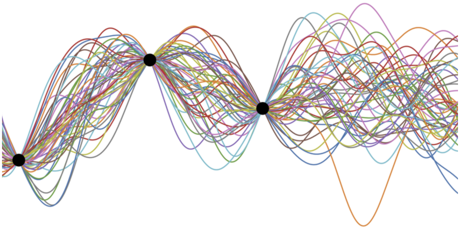
- No tenemos información de las derivadas de f, ni información sobre su linealidad, concavidad, etc. La llamamos una función "caja negra".
- Tenemos que si $x \in \mathbb{R}^d$, entonces $d \lesssim 20$.
- *La función f es continua*.
- f es difícil de evaluar.

Procesos Gausianos

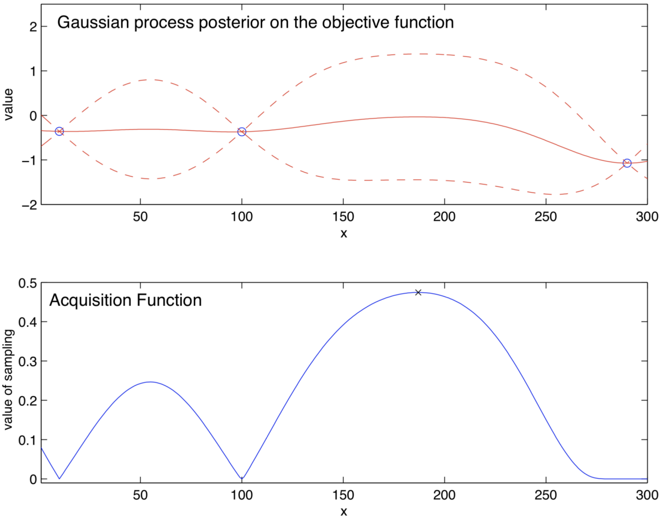
Sean $x_1, x_2, \ldots, x_k \in A$ puntos en los que evaluamos la función f. Sea $[f(x_1), \ldots, f(x_k)]$ un vector de estas evaluaciones.

La distribución resultante o proceso Gausiano asociado sería entonces:

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



Actualización de posterior

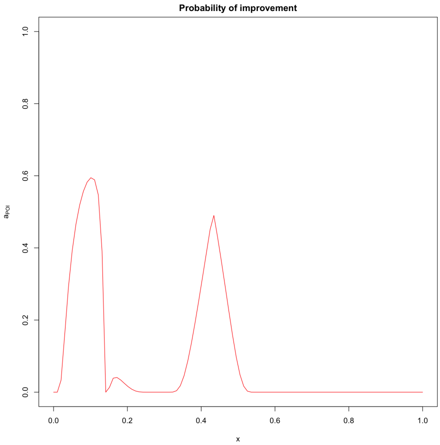


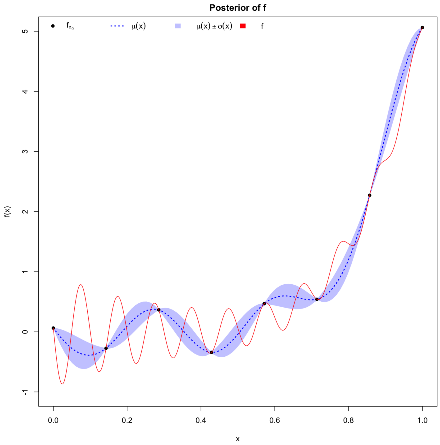
Algoritmo

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

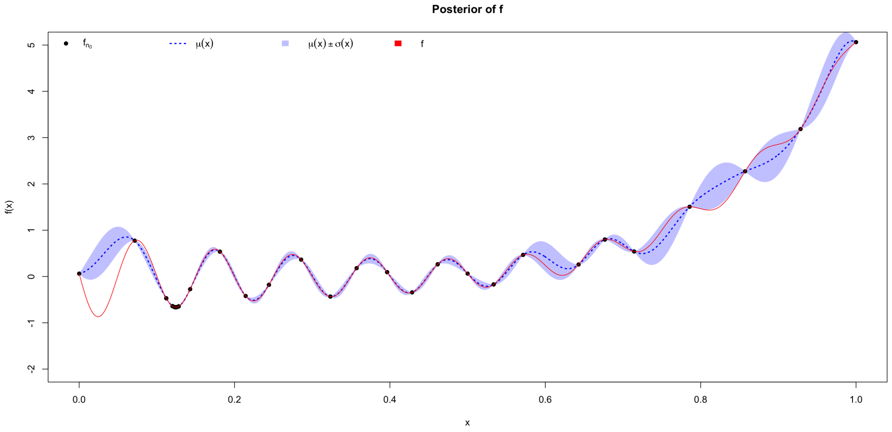


Tabla de contenidos

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:

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

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

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

Después de 45 iteraciones

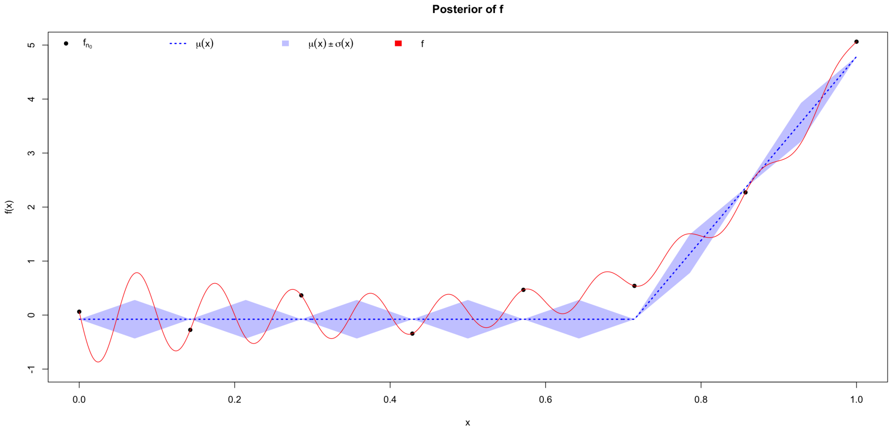
Usamos BASS porque:

- 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.

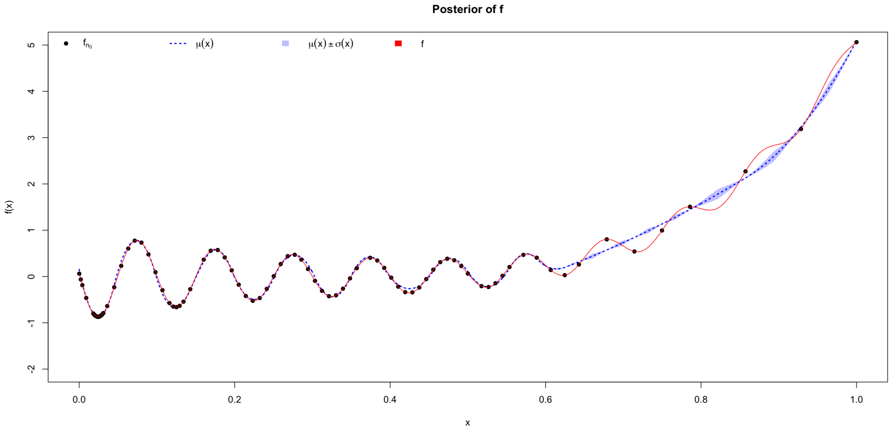
Trade-offs:

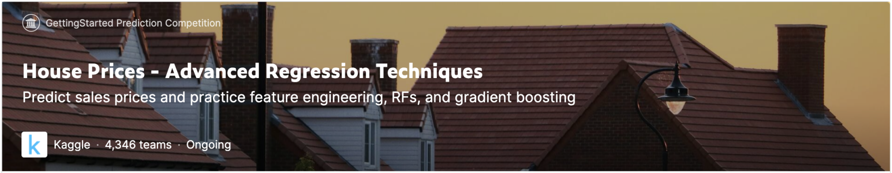
- Procesos Gausianos es óptimo.
- Mucho más lento.

BASS inicial



BASS después de 115 iteraciones





Overview Data Code Discussion Leaderboard Rules

Join Competition ···

Overview

- Description
- Evaluation
- Tutorials
- Frequently Asked Questions

Start here if...

You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition.

Competition Description

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

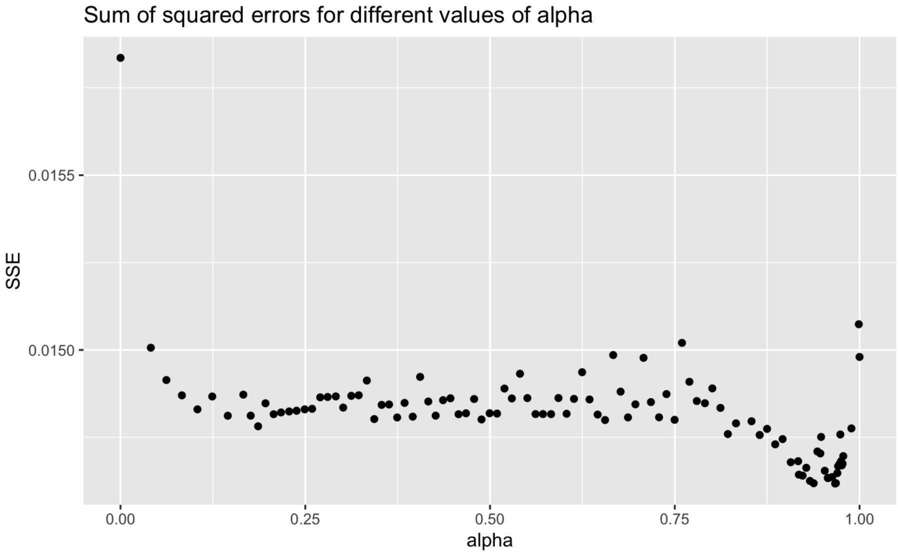
Problema real

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

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

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

Resultados de BASS



Resultados de BASS