Bayesian optimization

Viacheslav Borovitskiy

St. Petersburg State University St. Petersburg Department of Steklov Mathematical Institute

 $22 \ {\rm October} \ 2020$

Talk structure

1 Introduction

- 2 Gaussian processes
- 3 Bayesian optimization
- 4 Our own research

Talk structure

1 Introduction

- 2 Gaussian processes
- Bayesian optimization
- 4 Our own research

Bayesian optimization (BO)

Optimization algorithm best suited for

- black-box
- expensive evaluations
- low input dimensionality

Basic idea:

- approximate the target function with (probabilistic) model,
- use this model to decide where to evaluate next.

Most popular model for target function - Gaussian process (GP).

BO is useful

Bayesian Optimization in AlphaGo

Yutian Chen, Aja Huang, Ziyu Wang, Ioannis Antonoglou, Julian Schrittwieser, David Silver & Nando de Freitas

DeepMind, London, UK yutianc@google.com

They used GPs to model target function.

Viacheslav Borovitskiy (SPbU, PDMI)

Talk structure

1 Introduction

- 2 Gaussian processes
- Bayesian optimization
- 4 Our own research

Gaussian process regression

 $\mathrm{GP}-\mathrm{distribution}$ over functions.

Bayesian inference for GPs:

prior: hand-picked GP

data: noisy evaluations of the function likelihood: induced by Gaussian noise assumption posterior: another GP

Let us explore this visually ...



Viacheslav Borovitskiy (SPbU, PDMI)

Bayesian optimization

22 October 2020 8 / 19

Viacheslav Borovitskiy (SPbU, PDMI)

Bayesian optimization

22 October 2020 8 / 19



Talk structure

1 Introduction

- 2 Gaussian processes
- 3 Bayesian optimization

4 Our own research

The algorithm

Problem: minimize the target function $\phi : \mathbb{R}^d \to \mathbb{R}$.

At n'th step ϕ has already been evaluated at $x_1, ..., x_n$. How do we choose x_{n+1} ?

Build posterior GP f using data

$$x_1, \dots, x_n, \qquad \qquad \phi(x_1), \dots, \phi(x_n).$$

Choose

$$x_{n+1} = \underset{x \in \mathbb{R}^d}{\operatorname{arg\,max}} \mathbb{P}(f(x) < \underset{i=1..n}{\min} \phi(x_i)). \tag{MPI}$$

 or

$$x_{n+1} = \underset{x \in \mathbb{R}^d}{\operatorname{arg\,max}} \mathbb{E} \max(\min_{i=1..n} \phi(x_n) - f(x), 0). \quad (EI)$$

Automatic exploration/exploitation trade-off.

Let us minimize Forrester function $f(x) = (6x - 2)^2 \sin(12x - 4)$.



Choose some prior as $f_0 \sim GP(?,?)$.

Iteration 1.



Iteration 2.



Iteration 3.



Iteration 4.



Iteration 5.



Iteration 6.



Iteration 7.



Iteration 8.



Iteration 9.



Iteration 10.



Iteration 11.



Iteration 12.



Iteration 13.



Iteration 14.



Iteration 15.



Iteration 16.



Iteration 17.



Iteration 18.



Iteration 19.



Iteration 20.



Let us compare the model after 20 iterations with the target function.



For instance:

- parallel BO
- multi-objective BO
- multi-fidelity BO
- BO with gradients
- BO with constraints

Talk structure

1 Introduction

- 2 Gaussian processes
- Bayesian optimization

4 Our own research

Efficient high-dimensional Thompson sampling



Viacheslav Borovitskiy (SPbU, PDMI)

Matérn Gaussian processes on Riemannian manifolds

Viacheslav Borovitskiy^{*1,4} Alexander Terenin^{*2} Peter Mostowsky^{*1} Marc Peter Deisenroth³ To be presented on NeurIPS 2020.



(a) Ground truth

Viacheslav Borovitskiy (SPbU, PDMI)



(b) Posterior mean



(c) Standard deviation

Models in manifold setting

Matérn Gaussian Processes on Graphs

In review for AISTATS 2021



(a) Mean



(b) Standard deviation

Viacheslav Borovitskiy (SPbU, PDMI)

Thank you for your attention!

via cheslav. borovitskiy @gmail.com



St Petersburg University

Mathematics & Computer Science department

Some figures were taken from: http://inverseprobability.com/talks/.

Viacheslav Borovitskiy (SPbU, PDMI)