High-dimensional Bayesian optimization

A survey and benchmark

Miguel González-Duque - DTU BioSustain - 04/07/2024

About me

miguelgondu.com/about

B.Sc & M.Sc in Mathematics

Universidad Nacional de Colombia

PhD

IT University of Copenhagen, Creative AI Lab

Machine Learning programmer

University of Copenhagen, MLLS centre mlls.dk

Menu for the day

1. A brief intro to optimization
2. GPs and BO
    2.1. Why scaling GPs is difficult
3. Methods for scaling BO and how they perform
4. Testing on small molecules, prelim. Results

About today

High-dimensional Bayesian **optimization**

About today

High-dimensional Bayesian **optimization**

High-dimensional **Bayesian optimization**

About today

High-dimensional Bayesian **optimization**

High-dimensional **Bayesian optimization**

High-dimensional Bayesian optimization

About today

Goals for the day

1. Give you an overview of HDBO

2. Understand the needs of practitioners

High-dimensional Bayesian optimization

About today

A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences

Miguel González-Duque*
University of Copenhagen

Richard Michael
University of Copenhagen

Simon Bartels
Paul Sabatier University Tolouse

Yevgen Zainchkovskyy
Technical University of Denmark

Søren Hauberg
Technical University of Denmark

Wouter Boomsma
University of Copenhagen

https://arxiv.org/abs/2406.04739

About today

Slides are available online!

I will show this link at the end

High-dimensional Bayesian optimization

1. A brief intro to optimization

Optimization problems live in a triangle

Black-box problems as the most generic ones

Several approaches to black-box optimization (not only BO)

When and why should we use BO.

Start the flowchart

2. GPs and BO

GPs as a powerful method for regression w. Uncertainty

The toy example of my blogpost

BO: using GPs and their uncertainty to smartly explore an input space

- Acq. Functions: leveraging uncertainty to balance exploration and exploitation

The three reasons why GPs are believed to scale poorly.

1. Kernels as proxies for distance, and the curse of dim. (Blogpost example)

1. Optimizing the acquisition function is difficult

1. Dataset size

3. Methods for scaling BO and how they perform

Introduce our paper, introduce the taxonomy

Start building up the flowchart

Can you use gradient info?

Can you assume that many of your variables don’t matter?

Can you assume that linearly combining your variables makes sense?

Can you learn latent representations from tons of unsupervised examples?

Does your problem have a lot of structure (besides continuous opt.)?

Can you assume that your function can be decomposed into sums of smaller problems?

(Another angle, trust regions)

4. Testing on small molecules, prelim. Results

Our ongoing benchmark

Preliminary results: distinction between structured spaces and non-linear embeddings

SAASBO is pretty good, why?

Vanilla BO is pretty good, why?

Models don’t seem to scale well

General outlook: what should the focus be on?

How should practitioners approach BO?

What do practitioners need?