Stat 238, Fall 2025
Project Minis

Alexander Strang
Due: by **5:00 pm** Friday, March 28th, 2025

Mini Project 5: Bayesian Optimization

Policies

- This is a group project. You may work with up to 2 partners.
- All submissions must be properly type-set and sourced. This means providing formal citations and a complete bibliography when sources are used. Please export any notebooks as a single pdf and merge all components into one file. Only one group member should submit to gradescope. They must tag the other members when they submit.
- Project minis should be uploaded to Gradescope under the appropriate mini-assignment.

Prompts

This is a coding project. You will be asked to test and extend the Bayesian optimization algorithms you developed in lab 5 on a broader suite of test functions, to select one based on those tests, and apply it to a real optimization problem.

(a) (8 points) Complete the Q 2.7 - 2.10 in Lab 5 for the Ackley and Michalwiscz functions.

(b) (4 points) Select a test function from an applied optimization problem. This test function should be selected to model a real scenario where blackbox optimization (in particular, Bayesian Optimization using a GP surrogate) is relevant. These are optimization problems that share the following features:

- The objective is very slow or expensive to evaluate (as for large computer simulations, physical experiments, or for the hyperparameters of models that are expensive to train and test).
- Objective evaluations are noisy. This is an optional, but common, feature.
- Low-dimensional space of design parameters. Bayesian optimization is usually performed in moderately low-dimensional spaces. This is an optional, but common, feature. It enables global function approximation. If the design space is too large, Bayesian optimization will spend most of its effort exploring.
- Smooth but complicated objective. In particular the objective is usually not convex, and is often highly multimodal. This motivates approaches based on global function

approximation. If the objective is simple, then standard iterative optimizers based on local approximations are sufficient. This is an optional, but common, feature.

- Constraints. If the objective is simple (e.g. convex), but the design parameters are constrained, then, restricted to valid inputs, the optimization problem may still be difficult to solve with standard methods. Bayesian optimization offers two advantages here: global function approximation (avoids local maxima and requires exploration), and, if the objective is defined outside of the region of allowed inputs, then it may be sampled outside the constrained region to improve estimation inside the constrained region. This is a common, but optional feature.

You are welcome, and encouraged, to choose an optimization problem relevant to your own research or field. I've included some resources below to help you find examples.

There are a variety of libraries and packages that implement objectives from real applications. I've listed some sources below.

- **CompModels.** This is an R package. You can find the package here. It is documented here. It includes approximately 10 blackbox functions, largely drawn from engineering problems. See the documentation for two examples (e.g. the pressure vessel example).
- **BayesianOptimization.** This is a package built on scikitlearn and implemented in python. It includes examples that focus on hyperparameter optimization for classifier models. It is available here.
- **SAMBO:** This is a more implementation focused optimization package based on surrogate modeling. You could look here for a reference blackbox optimizers. SAMBO is available here.

For example applied papers and examples therein see:

- Sauer et al. *Vecchia-Approximated Deep Gaussian Processes for Computer Experiments.* Journal of Computational and Graphical Statistics. 2023.
- Gramarcy, RB. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences.* Chapman and Hall. 2020.
- Binois et al. *A Survey on High-dimensional Gaussian Process Modeling with Application to Bayesian Optimization.* 2022.
- Rajaram et al. *Empirical Assessment of Deep Gaussian Process Surrogate Models for Engineering Problems.* Aerospace Research Central. 2020.

(c) (4 points) Propose at least two Gaussian Process (GP) priors for your surrogates that you believe to be reasonable choices, and that you expect to produce different optimization behavior.

Choosing a GP means adopting a mean function and a covariance function.

1. Explain your choices for each (why are they reasonable). You may consider any of the three modeling aims discussed in class as arguments in favor of a prior (veracity, simplicity, tractability). Note: most surrogate models are chosen to be weakly informative based on loose smoothness guarantees.

2. Explain the difference between the models, how this effects sample draws from each model, and how you expect this to effect the behavior of the optimizer.

If you are lost here, Chapter 4 from Rasmussen and Williams (see the reference texts list on Ed) is a helpful resource.

(d) (6 points) Select one acquisition function/Bayesian optimization procedure, and apply it to your chosen example for both choices of your prior. You do not need to implement the optimization procedure yourself, but, must fully explain any tool you use. It must accept a GP prior specification, objective function, and standard acquisition function (e.g. an upper credible bound).

To compare the behavior of the optimizer:

- Make a plot showing the sequence of sampled objective function values and their running minimum.
- Make a plot showing the distance between the n^{th} and $n-1^{st}$ sample point from $n = 2$ up until convergence. This will help you evaluate how the procedure focuses in on a proposed minimizer.
- Make a plot showing the negative log posterior probability of each observed sample, $-\log(p(f(x_j)|f(x_{j-1}), f(x_{j-2}), \dots, f(x_1))$. This is a measure of how surprising each new observation is. The decay in the surprise measures concentration of the surrogate distribution to the true objective along the sequence of sampled points.
- A scatter plot showing the sequence of sample inputs at different stages of the procedure. I suggest coloring each point by the stage at which it was selected, and, if you have more than 2 design parameters, running PCA on the full set of sample points, then plotting the sequence of sample points in the space spanned by the first two principal components.