Bayesian Optimisation

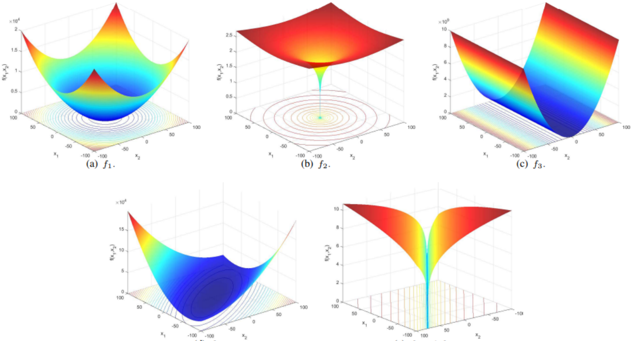
Tinkle Chugh

Department of Computer Science
University of Exeter, UK

In this Lecture

- Acquisition functions in single-objective Bayesian Optimisation
- Exploitation vs Exploration Trade-off

Recap of Single-objective Optimisation



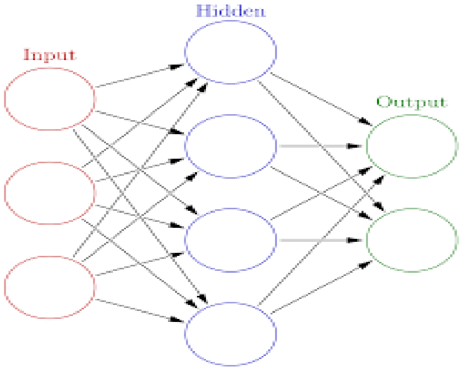
(a) f_1. (b) f_2. (c) f_3.

(d) f_4. (e) f_5 and f_6.

- Analytical formulation is not available
- Evaluation of the function is expensive

Examples of BO

Hyperparameter tuning in Neural Networks[1]

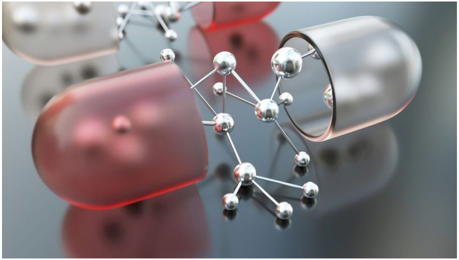


- Number of nodes, hidden layers
- Weights and bias

[1] Wu et al. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization, Journal of Electronic Science and Technology, 17, 2019, 26-40.

Examples of BO

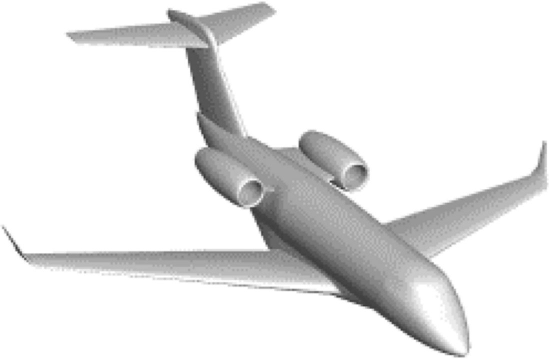
Drug Discovery[2]



[2] Colliandre L, Muller C. Bayesian Optimization in Drug Discovery. Methods Mol Biol. 2024;2716:101-136.

Examples of BO

Engineering Design[3]



[3] Forrester, A. I. J., Sobester, A., & Keane, A. J. (2008). Engineering Design via Surrogate Modelling: A Practical Guide. Wiley

Examples of BO

- Material Science[4]
- Financial Modelling and Portfolio Optimisation[5]
- Reinforcement learning[6]

[4] Xue, D., Balachandran, P., Hogden, J. et al. Accelerated search for materials with targeted properties by adaptive design. Nat Commun 7, 11241 (2016).

[5] Gupta, A. A Bayesian Approach to Financial Model Calibration, Uncertainty Measures and Optimal Hedging. Oxford University, UK, 2010.

[6] Paul et al. , JMLR, Robust Reinforcement Learning with Bayesian Optimisation and Quadrature, JMLR, 21, 1-31, 2020.

Bayesian optimisation

Problem

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$$

$$\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$$

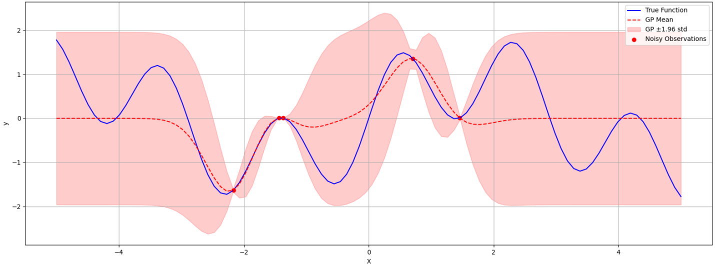
Bayesian optimisation

- Build a probabilistic surrogate model of $f(\mathbf{x})$
- Maximise an *acquisition* function $\alpha(\mathbf{x})$

$$\mathbf{x}' \leftarrow \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmax}} \; \alpha(\mathbf{x})$$

  Acquisition function balances exploitation with exploration
- Expensively evaluate $f(\mathbf{x}')$

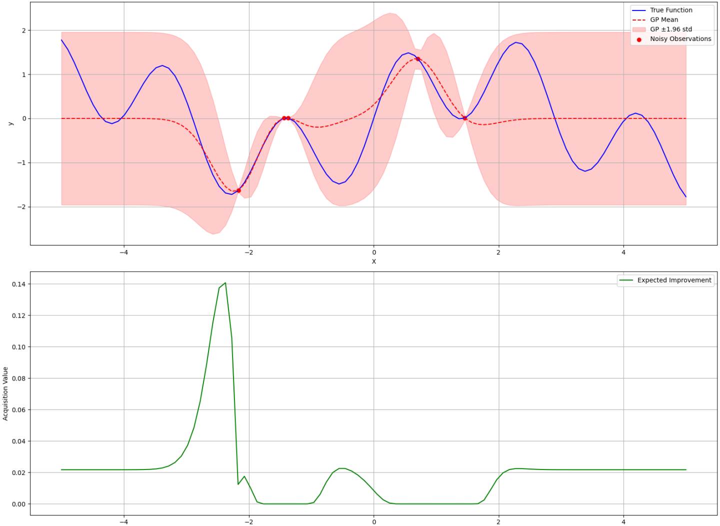
Build surrogate model



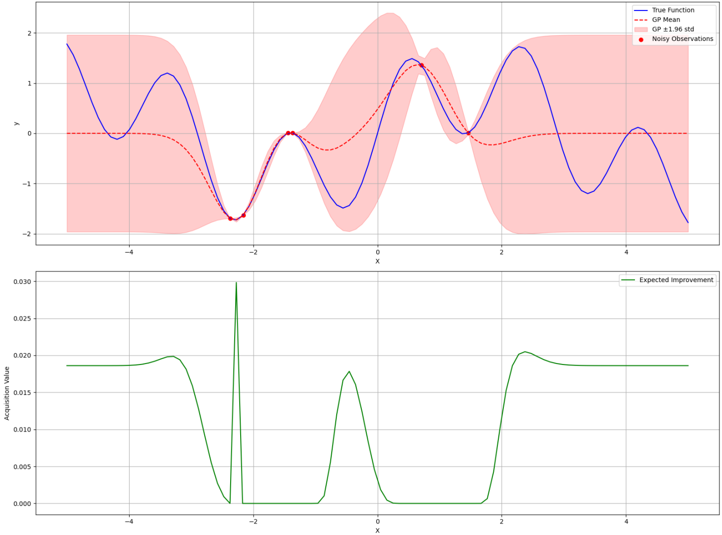
Exploitation and Exploration

$$\mathbf{x}' \leftarrow \text{argmax}_{\mathbf{x}\in\mathcal{X}} \ \alpha(\mathbf{x}, f^\star, \theta)$$

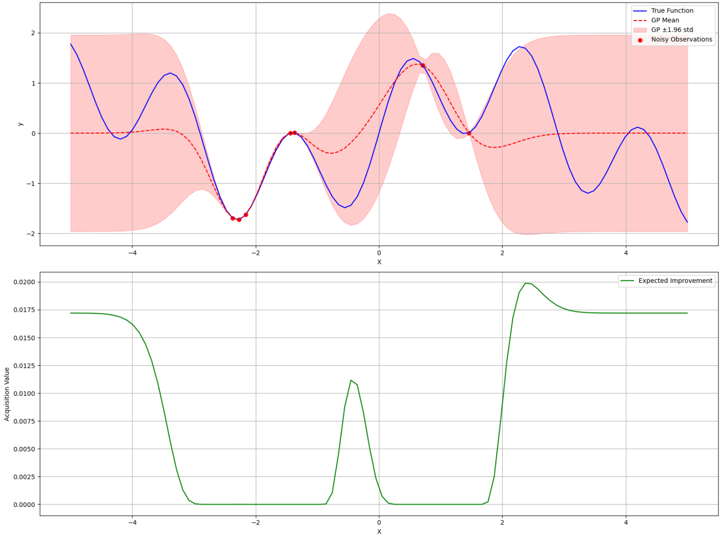
Balance exploration with exploitation



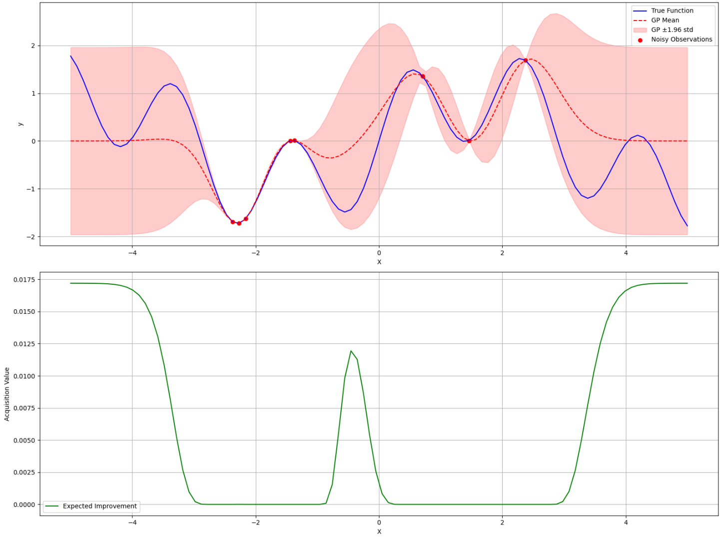
Exploitation and Exploration



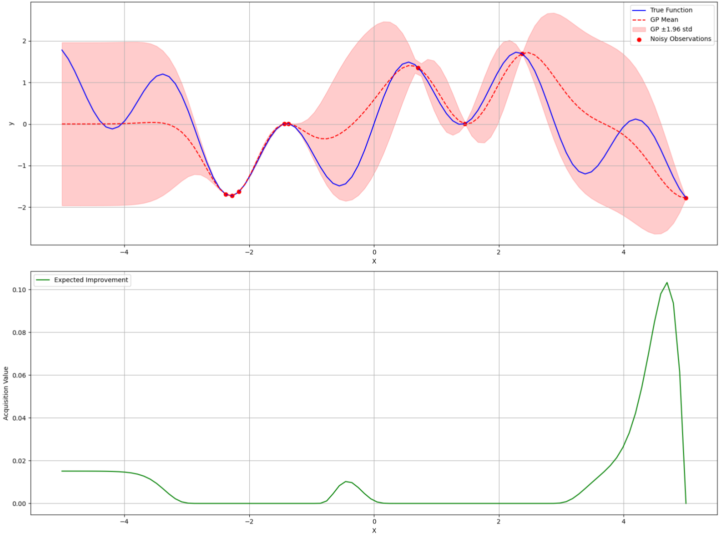
Exploitation and Exploration



Exploitation and Exploration



Exploitation and Exploration



Acquisition functions

- Expected Improvement
- Probability of Improvement
- Lower Confidence Bound

There are a few other acquisition functions in the literature

Expected Improvement

Maximising the expected gain over the best observed value.

The Expected Improvement at a point $\mathbf{x}$ is given by:

$$\mathsf{EI}(\mathbf{x}) = \mathbb{E}[\max(f^\star - f(\mathbf{x}), 0)]$$

where:

- $f^\star$ is the best observed value of the function so far.
- $f(\mathbf{x})$ is the predicted value at point $\mathbf{x}$.

Expected Improvement

The value $f(x)$ is modeled as a normal distribution with mean $\mu(x)$ and standard deviation $\sigma(x)$. Therefore, the Expected Improvement can be expressed as:

$$\mathrm{EI}(x) = \int_{-\infty}^{\infty} \max(f^{\star} - f(\mathbf{x}), 0)\, p(f(\mathbf{x}))\, \mathrm{d}f(\mathbf{x})$$

where $p(f(\mathbf{x}))$ is the probability density function of $f(\mathbf{x})$, which is normally distributed:

$$p(f(\mathbf{x})) = \frac{1}{\sqrt{2\pi\sigma(\mathbf{x})^2}} \exp\left(-\frac{(f(\mathbf{x}) - \mu(\mathbf{x}))^2}{2\sigma(\mathbf{x})^2}\right)$$

Expected Improvement

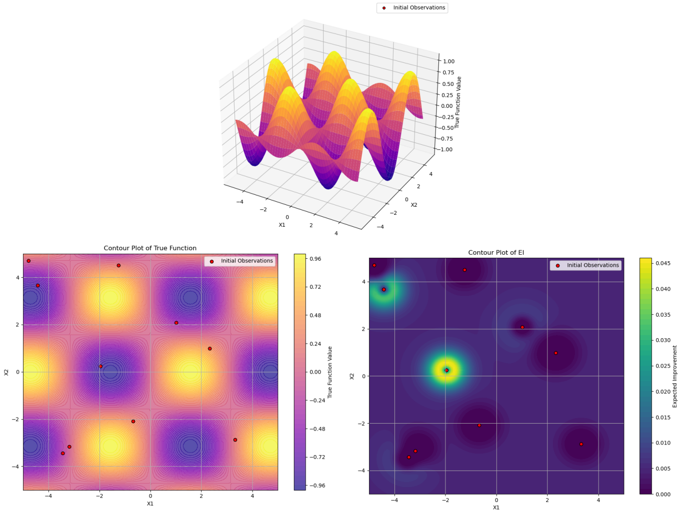
Ths integral can be rewritten using the normal CDF Φ and PDF ϕ:

$$\mathsf{EI}(\mathbf{x}) = (f^\star - \mu(\mathbf{x}))\Phi(Z) + \sigma(\mathbf{x})\phi(Z)$$

where:

- $Z = \frac{f^\star - \mu(\mathbf{x})}{\sigma(\mathbf{x})}$
- $\Phi(Z)$ is the cumulative distribution function of the standard normal distribution.
- $\phi(Z)$ is the probability density function of the standard normal distribution.

Expected Improvement



Probability of Improvement

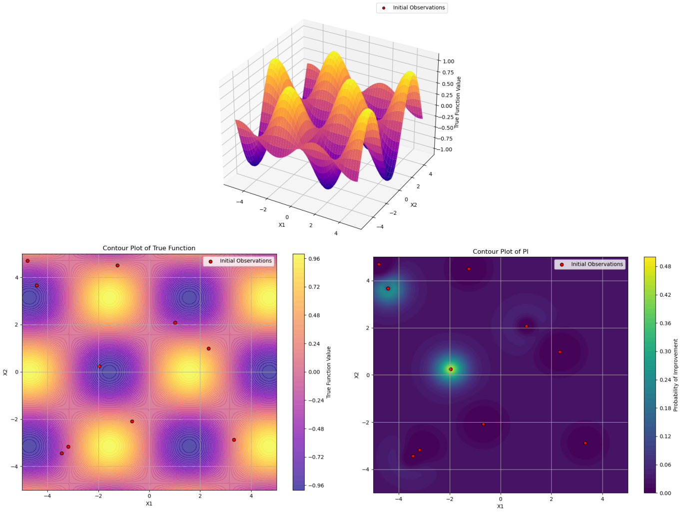
Maximising the probability of doing better than the best value.

$$\begin{aligned}\mathrm{PI}(\mathbf{x}) &= \int_{-\infty}^{f^\star} \phi\left(\frac{f(\mathbf{x}) - \mu(\mathbf{x})}{\sigma(x)}\right)\, df(\mathbf{x}) \\ &= \Phi\left(\frac{f^\star - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right)\end{aligned}$$

where:

- ϕ is the probability density function (PDF) of the standard normal distribution.
- Φ is the cumulative distribution function (CDF) of the standard normal distribution.

Probability of Improvement



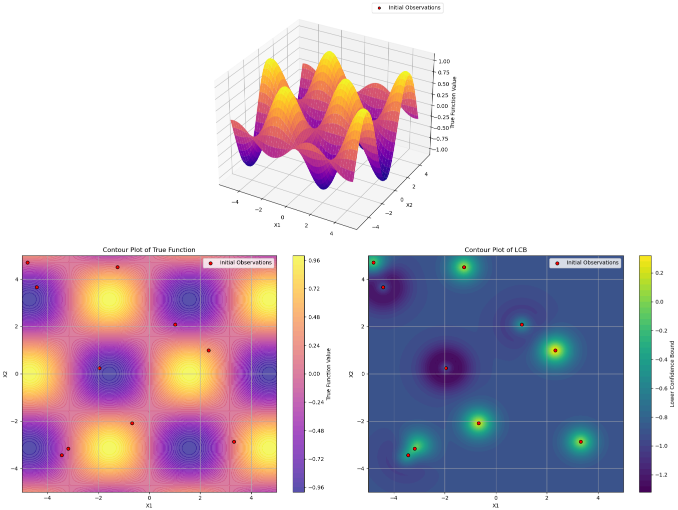
Lower Confidence Bound

Balancing exploration and exploitation by considering the lower bound of the confidence interval of the predicted mean.

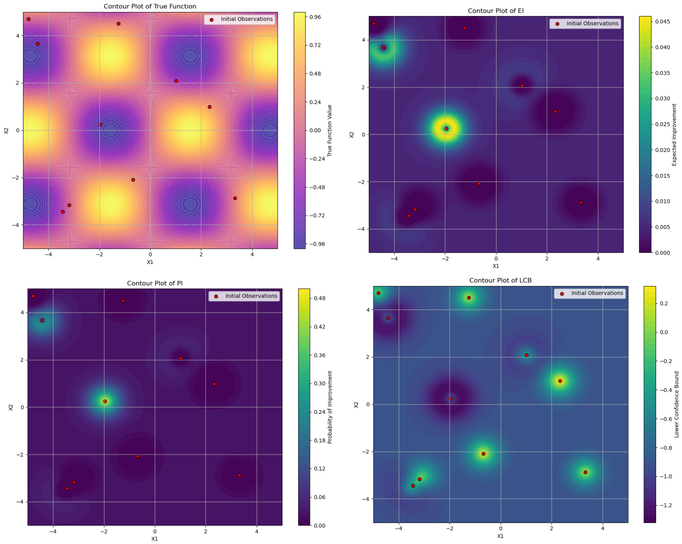
$$\mathrm{LCB}(\mathbf{x}) = \mu(\mathbf{x}) - \kappa \cdot \sigma(\mathbf{x}) \tag{1}$$

In BO, can maximise $-\mu(\mathbf{x}) + \kappa \cdot \sigma(\mathbf{x})$

Lower Confidence Bound



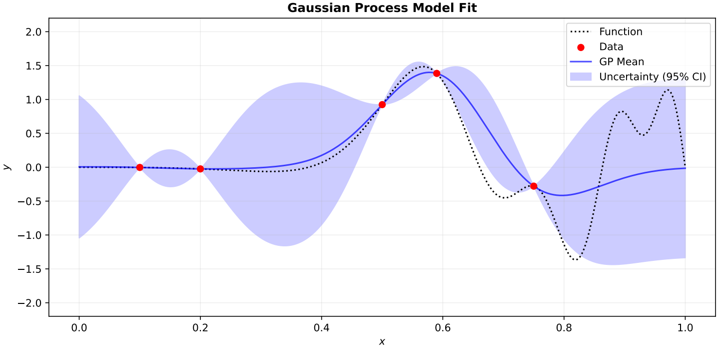
Acquisition Functions



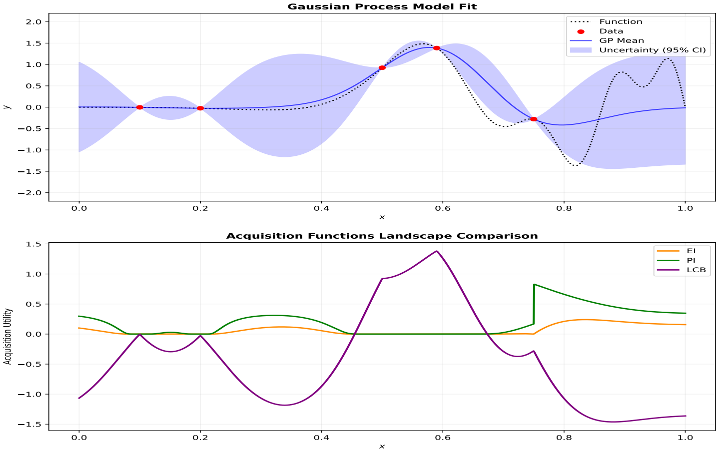
A Toy Example

$$\begin{aligned} \text{minimise} \quad & f(x) = \sin(3\pi x^3) - 0.5\sin(8\pi x^3) \\ \text{subject to} \quad & 0 \leq x \leq 1 \end{aligned}$$

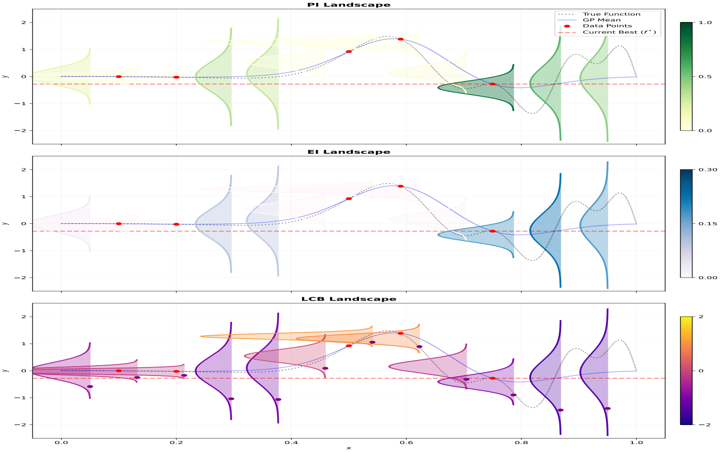
A Toy Example



A Toy Example



A Toy Example

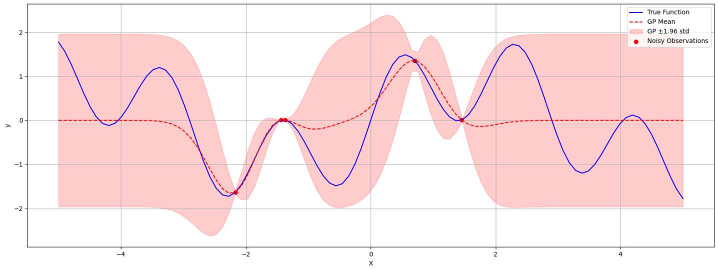


Multi-objectivisation of Single-objective BO

Two Goals

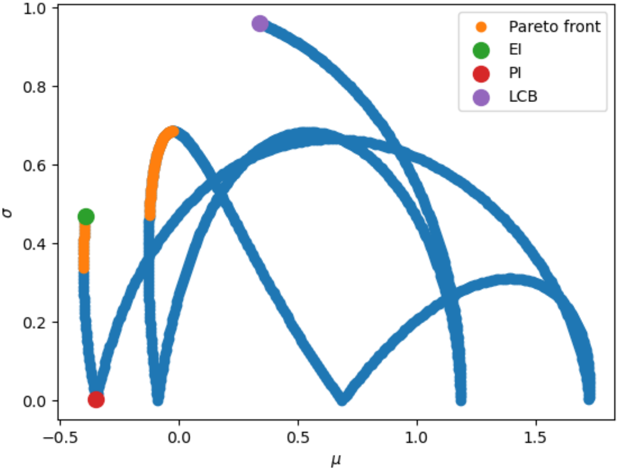
1. Exploitation - Minimising Posterior mean, μ
2. Exploration - Maximising Posterior uncertainty, σ

Multi-objectivisation of Single-objective BO



Multi-objectivisation of Single-objective BO

Trade-off between exploitation and exploration



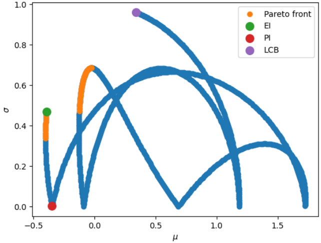
ϵ-greedy Acquisition Function (George et al. Greed is Good: Exploration and Exploitation Trade-offs in Bayesian Optimisation. ACM Transactions on Evolutionary Learning and Optimization (2021))

With probability ϵ:

explore by selecting a random solution from the exploration-exploitation Pareto set

else:

choose the most exploitative solution



ϵ-greedy acquisition

if $\texttt{rand}() < \epsilon$:

$\tilde{\mathcal{P}} \leftarrow \texttt{MOOptimise}_{\mathbf{x}\in\mathcal{X}}(\mu(\mathbf{x}), \sigma(\mathbf{x}))$

$\mathbf{x}' \leftarrow \texttt{randomChoice}(\tilde{\mathcal{P}})$

else:

$\mathbf{x}' \leftarrow \text{argmin}_{\mathbf{x}\in\mathcal{X}}\, \mu(\mathbf{x})$ *Exploit*

Other Important Elements in BO

- Kernel selection
- Intialising the points
- Optimising the acquisition function

Kernel selection

- Automatic kernel selection[7]
- Composite kernel[8]

[7] Abdessalem et al. Automatic Kernel Selection for Gaussian Processes Regression with Approximate Bayesian Computation and Sequential Monte Carlo. Frontiers in Built Environment, 2017

[8] Chugh et al. Trading-off Data Fit and Complexity in Training Gaussian Processes with Multiple Kernels, LOD 2019.

Initialising the model

- Random
- Uniform
- Design of experiment - Latin hypercube, Sobol

Optimising Acquisition function

- Gradient Based - Conjugate gradient, BFGS
- Evolutionary - Differential Evolution, CMA-ES, NSGA-II

Resources

- Summer school on GPs[9]
- Book on BO[10]
- BO workshops and sessions [11]

[9] https://gpss.cc/

[10] https://bayesoptbook.com/

[11] https://uq.math.cnrs.fr/june24, https://saeopt.bitbucket.io/