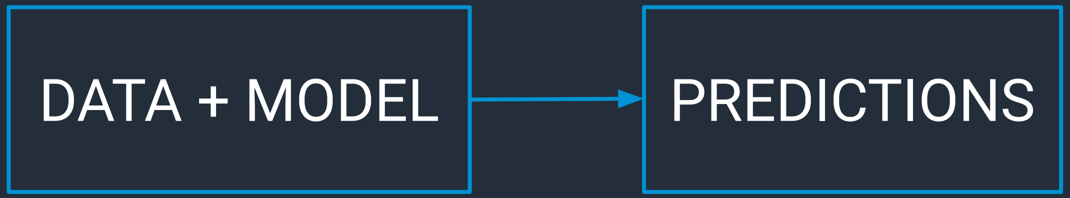
Bayesian Optimisation: Sequential Decision Making Under Uncertainty

•••

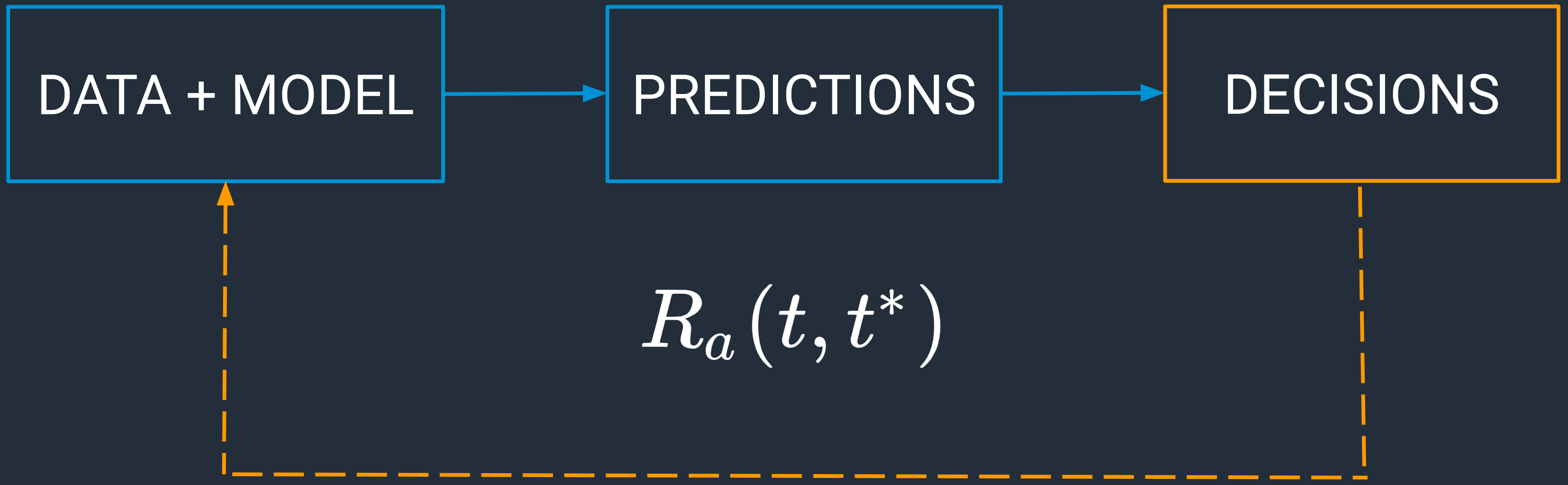
Alessandra Tosi

mind FOUNDRY

Accelerate Science
Winter School
2/3/4 Feb 2021


DATA + MODEL
PREDICTIONS


DATA + MODEL
PREDICTIONS
DECISIONS


DATA + MODEL
PREDICTIONS
DECISIONS
$R_a(t, t^*)$

Outline

- Sequential decision making
- Bayesian optimisation
- Applications

Sequential decision making

Sequential decision making

Multi-armed bandits
Minimize the regret, the expected difference between the reward sum associated with an optimal strategy and the sum of the collected rewards:

$$\rho = T\mu^* - \sum_{t=1}^{T} \hat{r}_t$$

Active learning
A user/oracle is asked to label new data points and reduce the model uncertainty.

Bayesian optimization
Minimize an unknown function

$$x^* = \min_{x \in D} f(x),$$

Reinforcement learning
An agent is interacting with the environment. The goal is to maximize the cumulative expected reward for an action.

Problem definition

The goal of global optimisation

$$x^* = \min_{x \in D} f(x), \qquad f : D \subset \mathbb{R}^n \to \mathbb{R}$$

Problem definition

The goal of global optimisation

$$x^* = \min_{x \in D} f(x), \qquad f : D \subset \mathbb{R}^n \to \mathbb{R}$$

Define a surrogate model for the target function to be optimised.

Frame the optimization problem as a decision problem (decide where to collect the new data point) and define an utility function to do so.

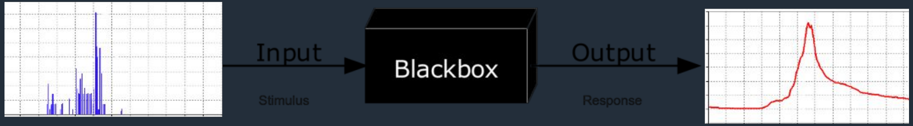

Drinks in Queue


PGN 75

Pain points

Complex target function

- no simple analytical form
- no gradient information
- mixed parameters’ types
- parameters’ constraints

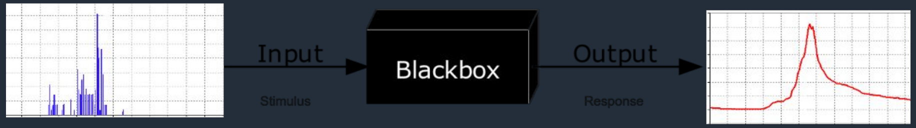


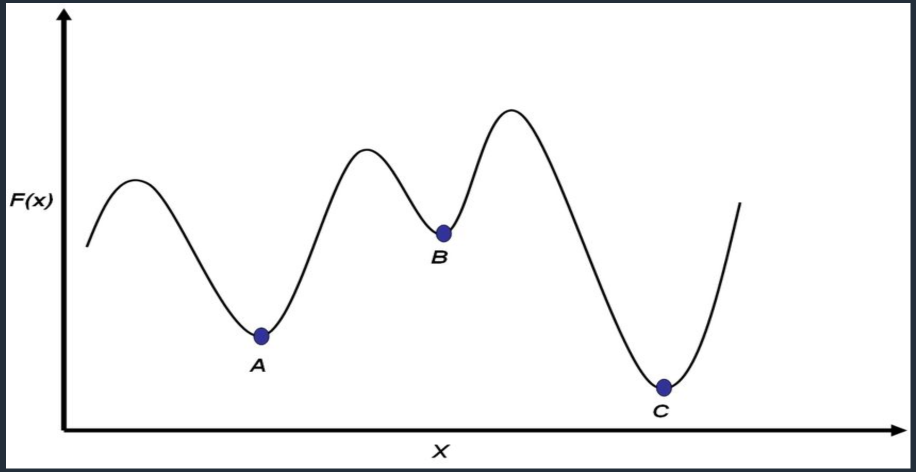
Pain points

Complex target function

- no simple analytical form
- no gradient information
- mixed parameters' types
- parameters' constraints

Multiple local optima





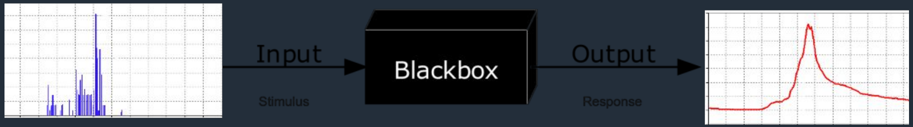
Pain points

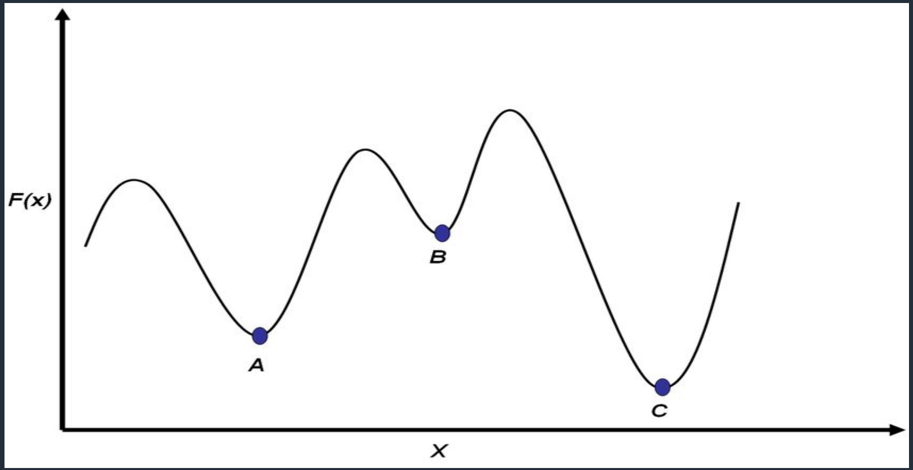
Complex target function

- no simple analytical form
- no gradient information
- mixed parameters' types
- parameters' constraints

Multiple local optima

Expensive function evaluations





Pain points

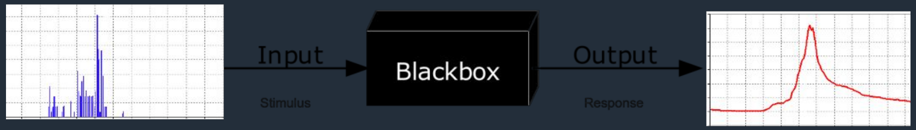
Complex target function

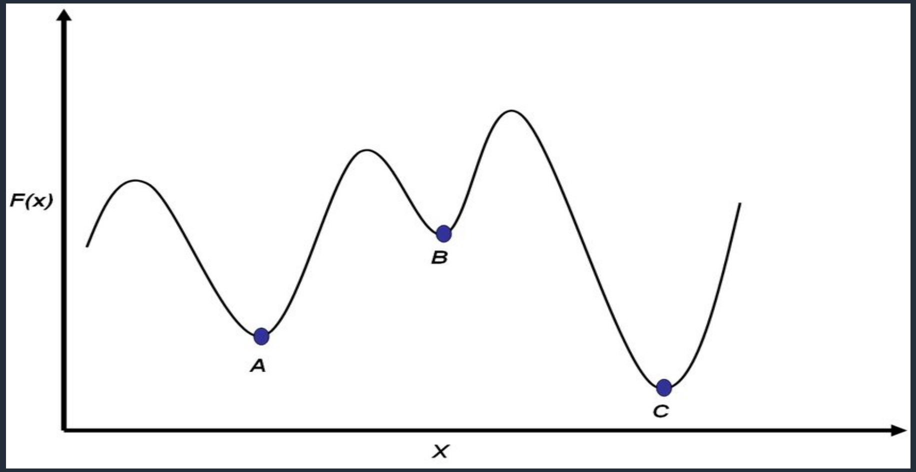
- no simple analytical form
- no gradient information
- mixed parameters’ types
- parameters’ constraints

Multiple local optima

Expensive function evaluations

Noisy function evaluations





Pain points

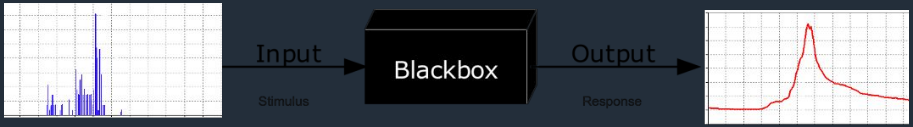
Complex target function

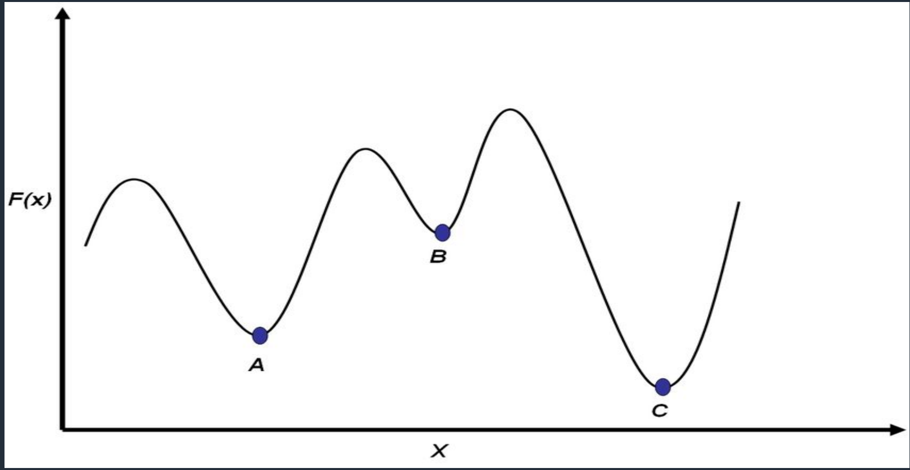
- no simple analytical form
- no gradient information
- mixed parameters' types
- parameters' constraints

Multiple local optima

Expensive function evaluations

Noisy function evaluations





How do we solve it?

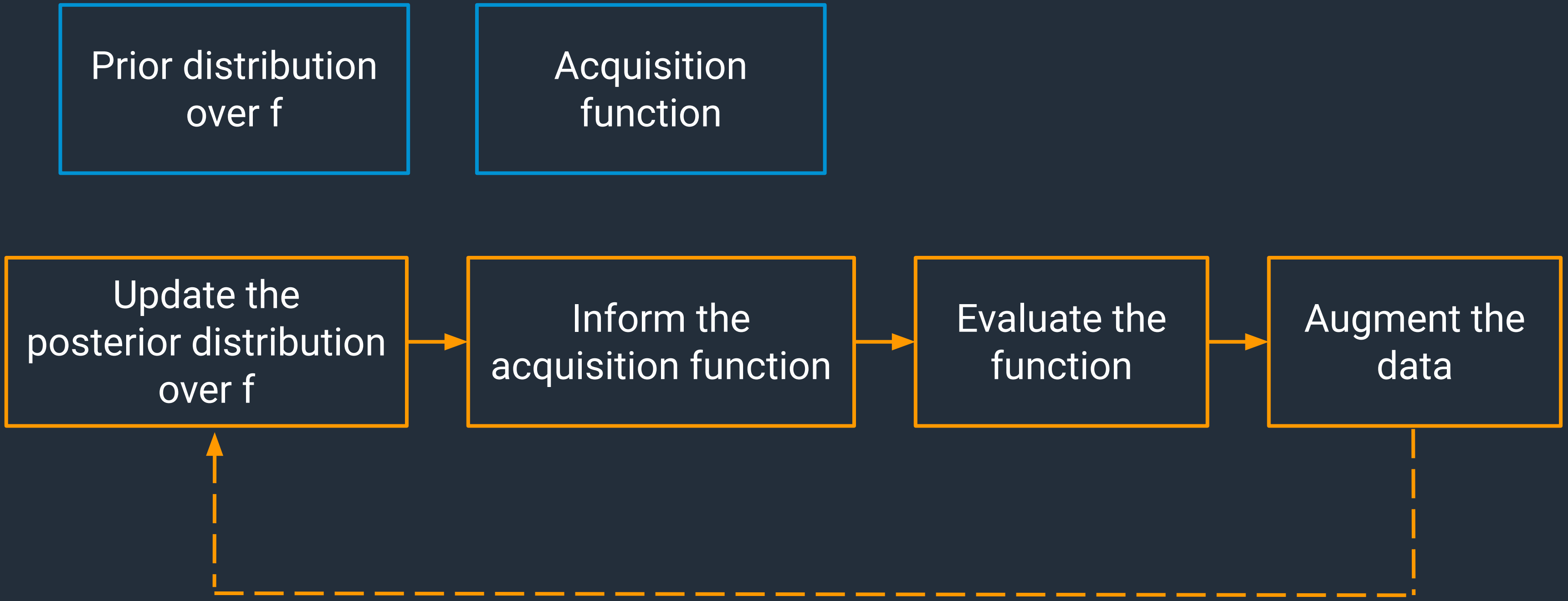
Bayesian optimisation

Prior distribution over f

Acquisition function

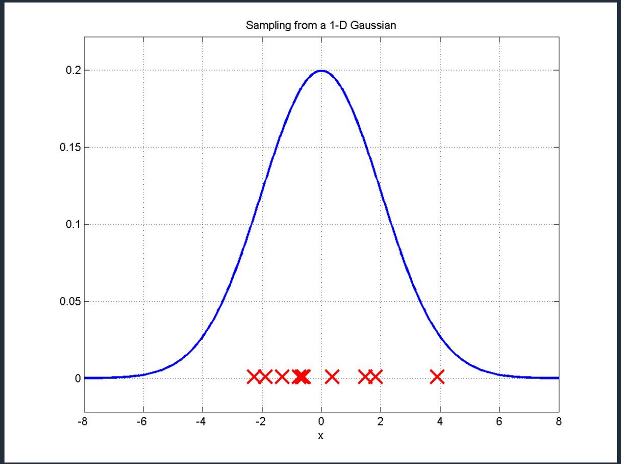
$$x^* = \min_{x \in D} f(x), \qquad f : D \subset \mathbb{R}^n \to \mathbb{R}$$

Bayesian optimisation



Gaussian processes [Rasmussen, C. E. and Williams, C. K. I., Gaussian Processes for Machine Learning. The MIT Press, 2005]

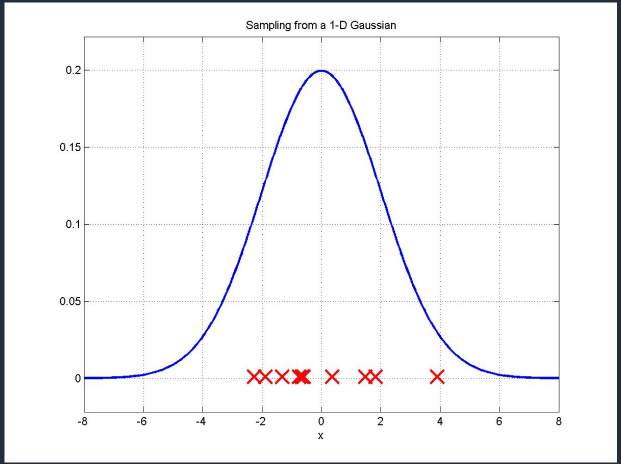
Infinite dimensional Gaussian distributions.

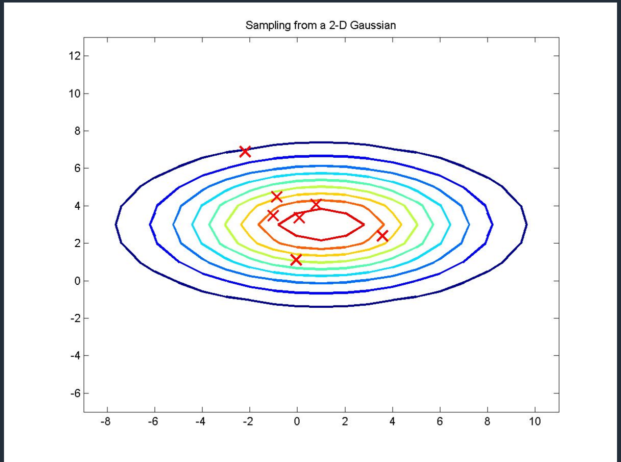


Images from Andreas Damianou - link

Gaussian processes

Infinite dimensional Gaussian distributions.

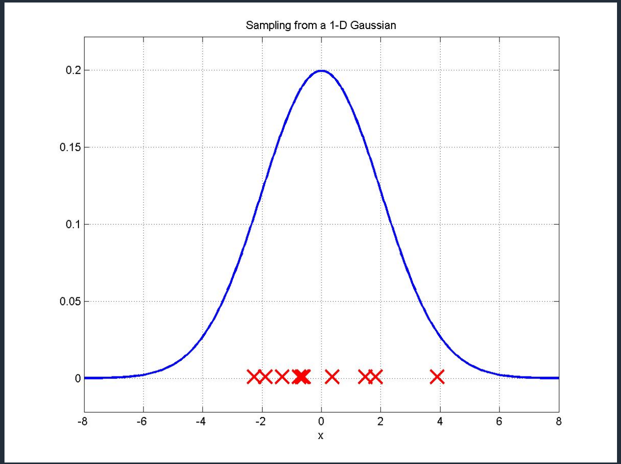


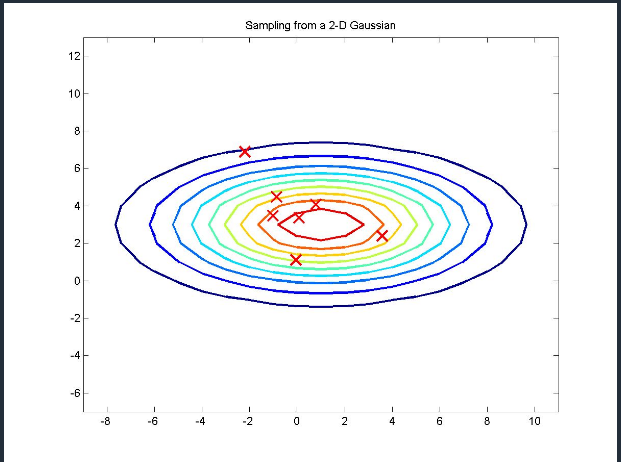


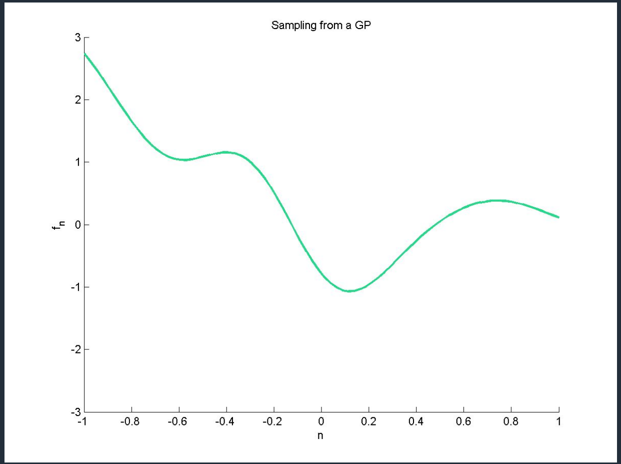
Images from Andreas Damianou - [link]

Gaussian processes

Infinite dimensional Gaussian distributions.



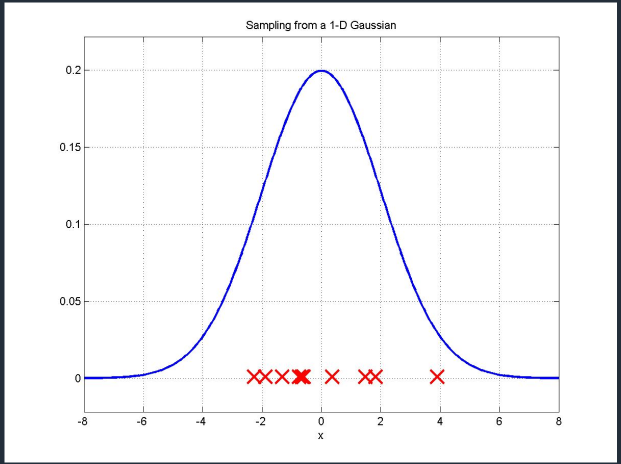


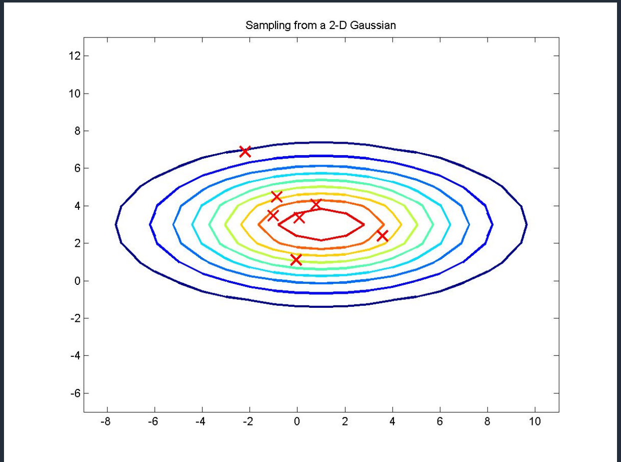


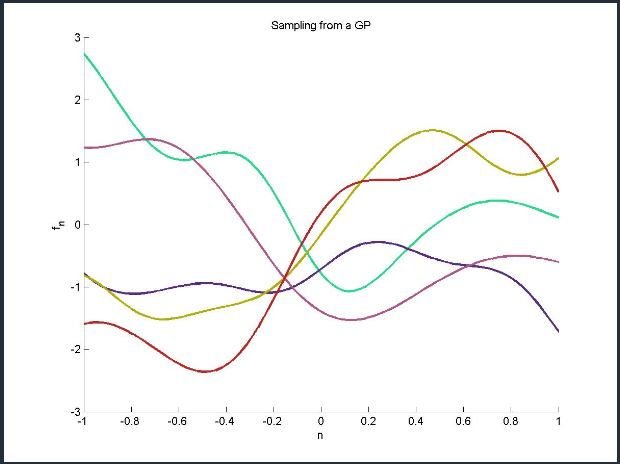
Images from Andreas Damianou - link

Gaussian processes

Infinite dimensional Gaussian distributions.



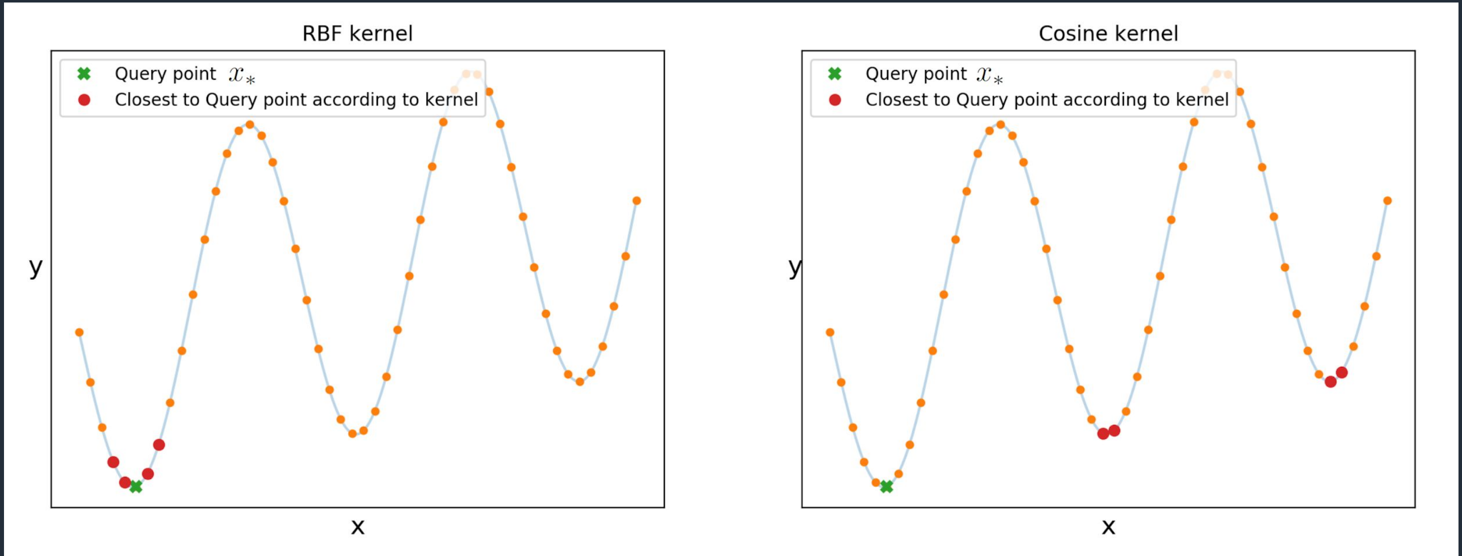




Images from Andreas Damianou - link

Gaussian processes

Kernels? It will be useful for the applications slide.



Images from Andreas Damianou - [link]

Bayesian optimisation

Prior distribution over f

Acquisition function

Acquisition functions

Balance exploration and exploitation.

Acquisition functions

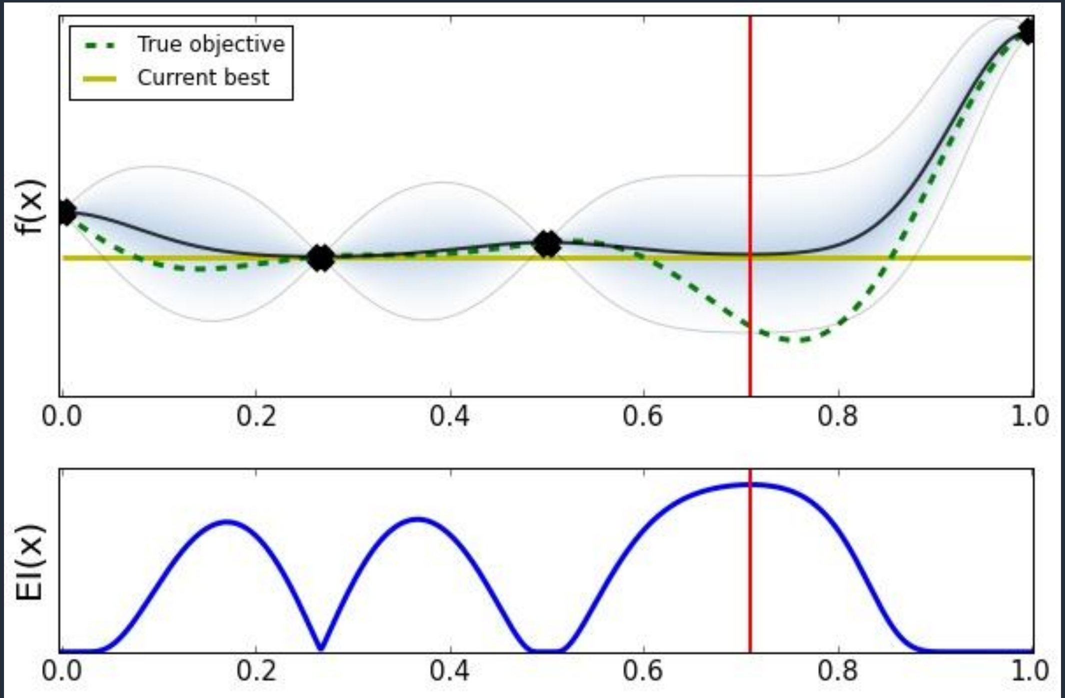


Image from Emukit https://emukit.github.io/bayesian-optimization/

Acquisition functions

Expected improvement [Jones et al., 1998]

$$u(x) = max_x(0, f^*, f(x))$$

$$a_{EI}(x) = \mathbb{E}[u(x)|x]$$

Upper confidence bound [Srinivas et al., 2010]

Probability of improvement

Thompson sampling

Entropy search [Hennig and Schuler, 2013; Hern´andez-Lobato et al., 2014]

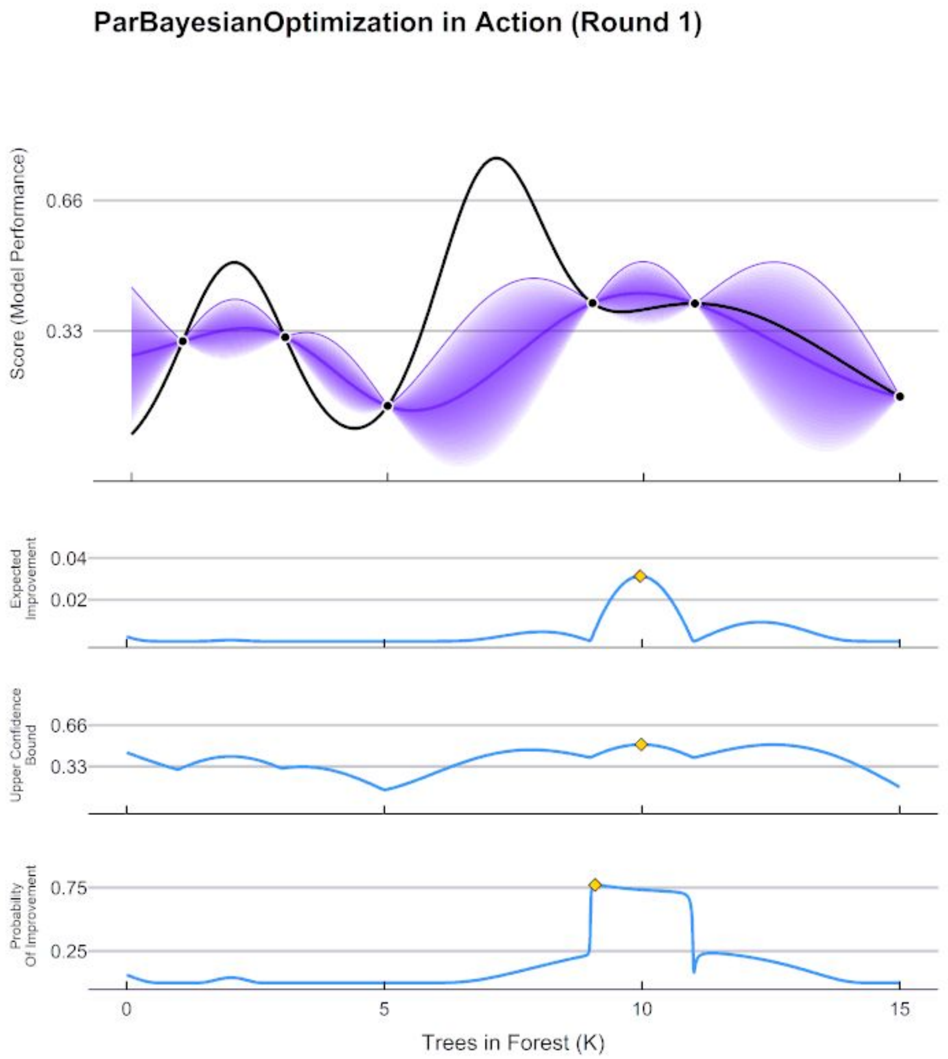


Image from Wikipedia Bayesian Optimization link

Recap

Sequential decision making in machine learning

Bayesian optimisation

- Problem definition
- Gaussian process surrogate model
- Acquisition functions

Coming next

Applications

Applications

Hyperparameters tuning

Drug design

Quantum computer

Robotics

Other

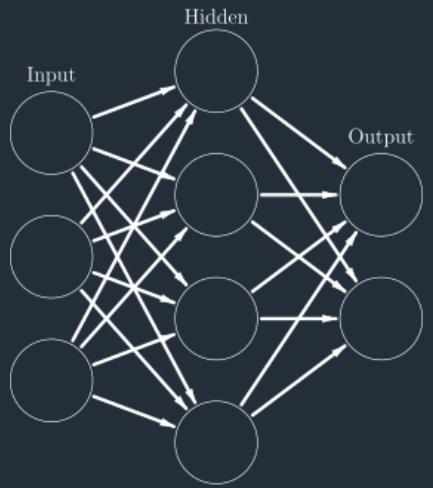
Hyperparameters tuning

Popular application. It works!

Hyperparameters tuning

Popular application. It works!

1. Tune a deep neural network
2. Tune data science pipelines, AutoML
    - Example: Auto Prognosis - code
      [A. M. Alaa and M. van der Schaar, 2018]
3. Tune the parameters of physical simulations



Chemical design

Generating novel molecules with optimised properties.

Recent work from Griffiths and Hern andez-Lobato, 2020. “ Constrained Bayesian optimization for automatic chemical design using variational autoencoders†”, paper

Chemical design

Generating novel molecules with optimised properties.

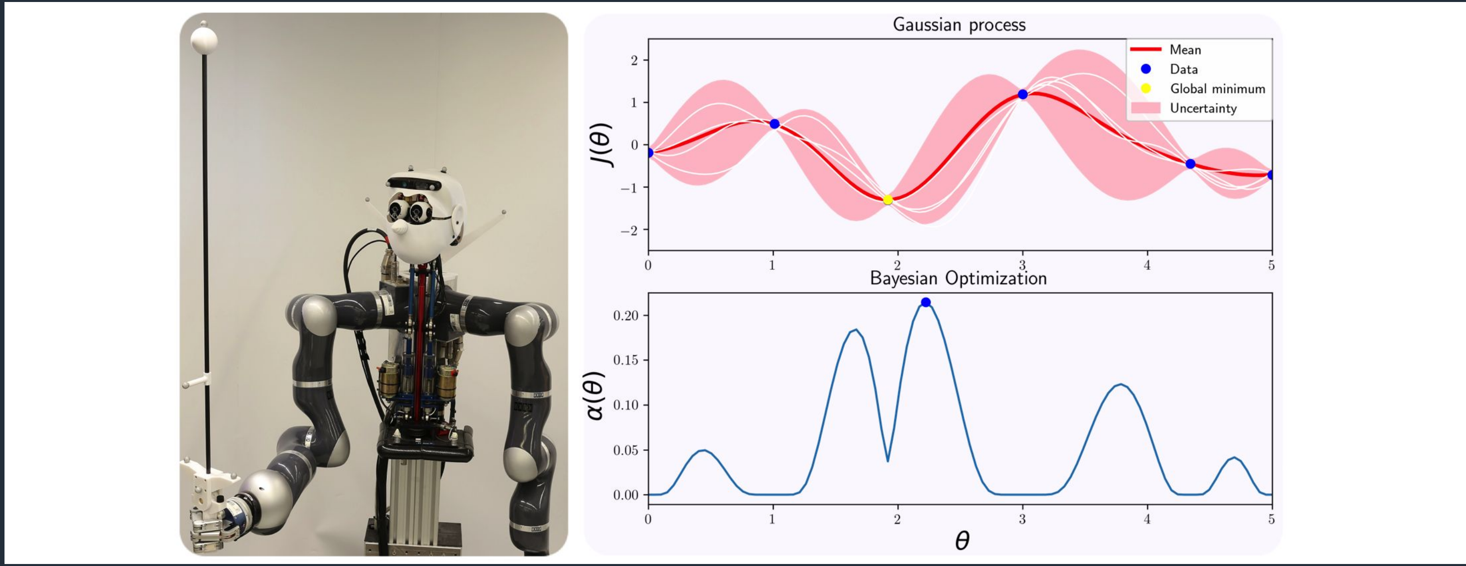
Let’s link this application to the 4 pain points seen before:

1. Complex target function
   - no simple analytical form
   - no gradient information
   - mixed parameters’ types
   - parameters’ constraints
2. Multiple local optima
3. Expensive function evaluations
4. Noisy function evaluations

Recent work from Griffiths and Hern andez-Lobato, 2020. “ Constrained Bayesian optimization for automatic chemical design using variational autoencoders†”, paper

Robotics

Find the optimal value for a robotic controller parameter.



Neumann-Brosig et Al, 2020. “Data-efficient Autotuning with Bayesian Optimization: An Industrial Control Study” NeuMarSchTri18 - link - video

Quantum computing

Training of Quantum Circuits on a Hybrid Quantum Computer

A research lab was applying a hybrid quantum learning scheme on a trapped-ion quantum computer to accomplish a generative modelling task.

Bayesian Optimization was applied by simulating the training procedure for a classical simulator in place of the quantum processor for a given set of parameters.

Key to the success of this project was the incorporation of some domain knowledge into the Bayesian Optimizer - in this case, that some of the optimizable parameters were cyclical in nature.

Zhu et Al, 2019. “Training of quantum circuits on a hybrid quantum computer” - link

Many other applications

A non-exhaustive list

- Experimental design
- Material design
- Mechanical asset design
- Optimising a sensor on a device/robot
- Gene design
- Industrial process optimisation
- Financial portfolio optimisation
- Logistics
- Causal Bayesian optimisation [Aglietti et Al 2020]

Take home message

Integrating the expert knowledge in the iterative decision making process is very important

References - software

- Emukit https://emukit.github.io/about
- GPyOpt https://github.com/SheffieldML/GPyOpt
- GPFlow https://github.com/GPflow/GPflowOpt
- Spearmint https://github.com/HIPS/Spearmint
- Scikit-optimize https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html
- BoTorch https://github.com/pytorch/botorch
- RoBO https://github.com/automl/RoBO
- Hyperopt https://github.com/MBKraus/Hyperopt/blob/master/README.md
- Mind Foundry Optimize https://optaas.mindfoundry.ai/static/swagger/index.html
- Others: please let me know

References

A non-exhaustive list of references

- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P, de Freitas, N., 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proceedings of the IEEE, Vol.104, No.1, January 2016
- Srinivas, N., Krause, A., Kakade, S., Seeger, M., 2009. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In Proceedings of the 27th International Conference on Machine Learning
- Jones, D. R., Schonlau, M., Welch, W. J., 1998. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization, 1998

Thank you