Enhancing Active Learning in Emukit: an extended experimental analysis

An Open Source project proposal by George Barbulescu | R244 | 30/11/2022

Emukit: Emulation of physical processes with Emukit

- Modular framework for active learning proposed by Paleyes, Andrei, et al. 2021.
- A broad range of contracts for: Bayesian Optimisation, experimental design, sensitivity analysis, quadrature, and multi-fidelity emulation
- Model-agnostic backend
- Related active learning frameworks:
  - Focused on single tasks (e.g., BayesOpt)
  - Tightly-coupled with a modelling framework
- Case studies (e.g., BO for quantum computer memory)

```python
class IModel:
    def predict(self, X: np.ndarray) -> Tuple[np.ndarray, np.ndarray]:
        """
        Predict mean and variance values for given points

        :param X: array of shape (n_points x n_inputs) of points to run prediction for
        :return: Tuple of mean and variance which are 2d arrays of shape (n_points x n_outputs)
        """
        raise NotImplementedError

    def set_data(self, X: np.ndarray, Y: np.ndarray) -> None:
        """
        Sets training data in model

        :param X: new points
        :param Y: function values at new points X
        """
        raise NotImplementedError
```

Emukit Contract

Reference: github.com/EmuKit/emukit

Proposed Open-Source Contributions

1. On Bayesian Optimisation

- Emukit supports a deluge of acquisition functions

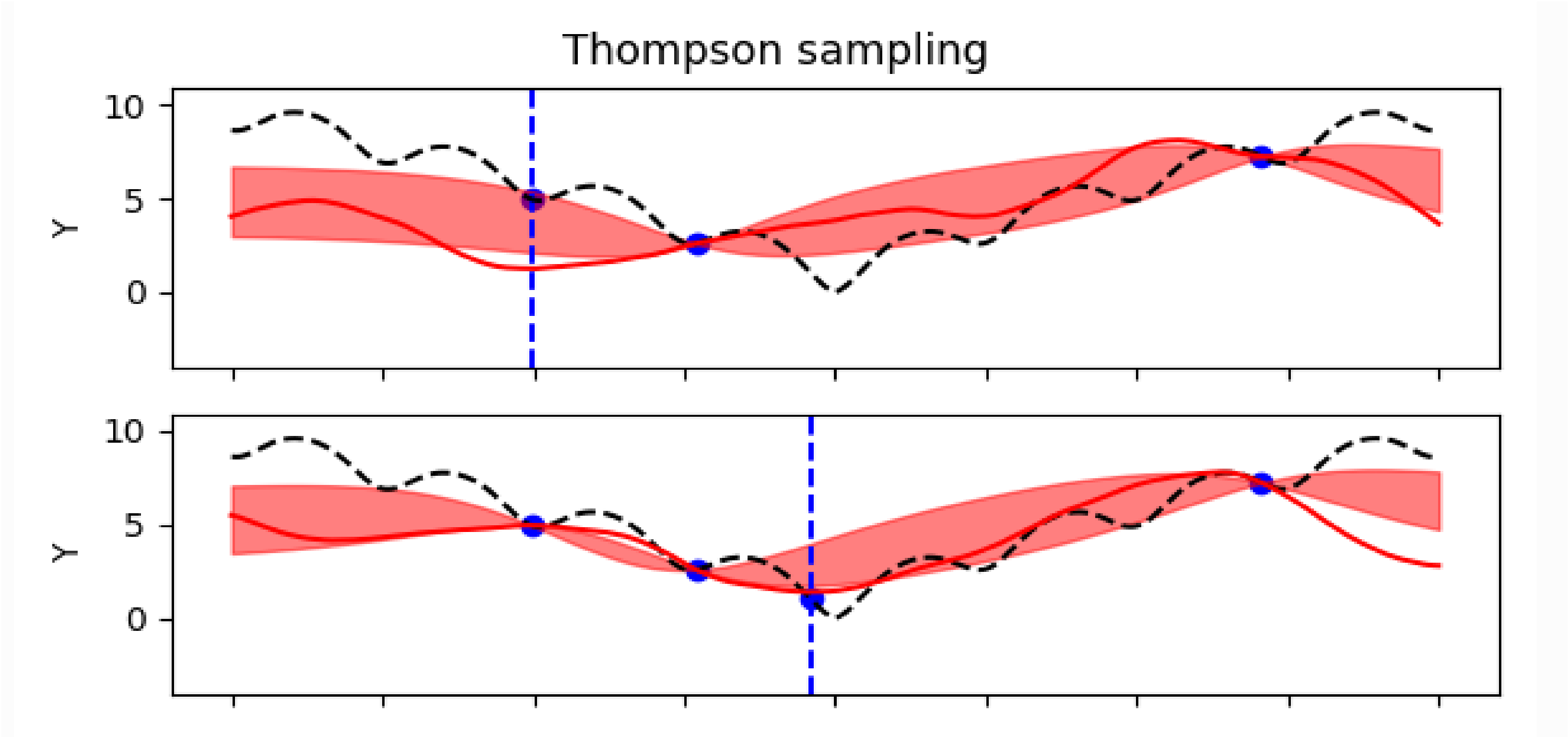
Proposal

- Thompson Sampling acquisition function
- New contract for the surrogate Model
- Thompson Sampling requires a function sample from the prior

Library	Built-in model	Acquisition function
DiceOptim	GP	EI, EQI, KG
laGP	GP	EI
mlrMBO	GP, RF	EI, EQI, UCB, PM
Spearmint	GP	EI, predictive ES
GPyOpt	GP, RF	EI, ES, UCB, PoI
Cornell-MOE	GP	EI, KG, predictive ES
GPflowOpt	GP	EI, UCB, PoF, PoI, max-value ES
pyGPGO	GP, GBM, RF, tSP	EI, PoI, UCB, predictive ES
Emukit	GP	EI, ES, UCB, PoF, PoI, max-value ES
Dragonfly	GP	EI, PoI, TS, UCB
Trieste	GP	EI, UCB, PM, PoF, TS, max-value ES
BoTorch	GP	EI, KG, PM, PoI, UCB, max-value ES

Table from Zhang, Mimi, et al. 2021

Thompson Sampling Acquisition



Example taken from https://num.pyro.ai/en/stable/examples/thompson_sampling.html
Full paper citation: Bingham, Eli, et al. 2019

TS-BO: Motivation

- Thompson Sampling Bayesian Optimization (TS-BO) can be extended to perform evaluations in parallel
- A set of "n" sequential evaluations are equivalent to "n" (a)synchronous evaluations across "T" threads/devices.
- Proof: Kandasamy, Kirthevasan, et al. 2018
- Parallel Bayesian Optimisation with Emukit

2. On Experimental Design

Uncertainty Sampling

- Select most ‘uncertain’ data point

Integrated Variance Reduction

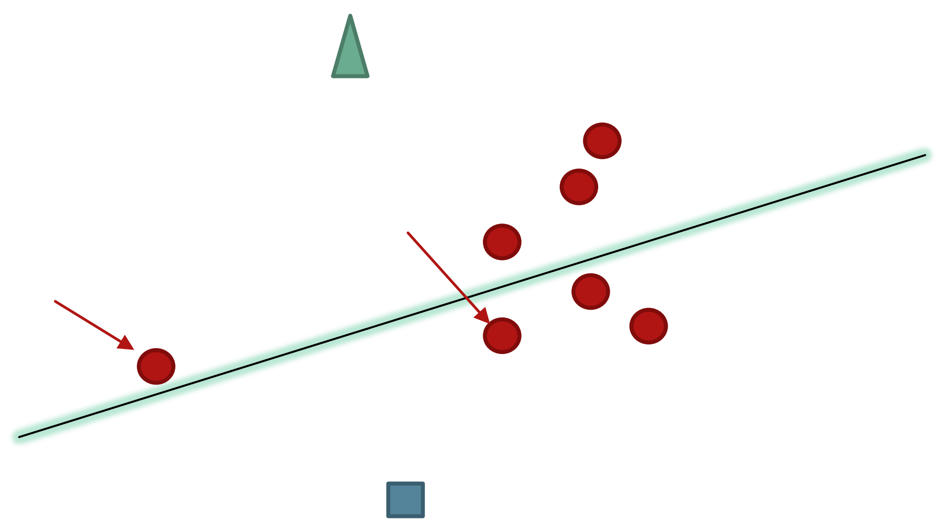
- Expected Error Reduction
- Evaluate ‘x’ if it minimises the **future** variance

Density Weighting

- High-variance samples may be isolated
- Sample from populated regions
- dw(x) = us(x) × *density*(x)

Solutions

- Kernel density
- Approximations (Settles, Burr, 2009)



Example decision boundary for classification task.
Bayesian Experimental Design

Research Contributions

An extended analysis of Emukit's active learning

- An extended analysis with respect to existing methods in Emukit
- Comparison with BoTorch on Bayesian Optimisation

Extensions

- Hyper-parameter tuning case study
- Parallel TS - Bayesian Optimisation

Future vision

- Original proposal: Multi-objective Bayesian Optimisation (MOBO)

Progress

- **Investigate Emukit's open source codebase**
- **Devise a plan for practical contributions**
- **Design a research methodology**
- Engineer the proposed contributions in Emukit
- Evaluate the methods based on the methodology
- Case study extension
- Pull request

Thank you!

Appendix: Emukit Workflow

1. *while **stop condition** is False:*
2. *acquire sample 'x' based on emulator*
3. *run experiment with sample 'x'*
4. *update the emulator with the observed behaviour*

User defines the business problem and injects the model into the Active Learning loop

References

- Paleyes, Andrei, et al. "Emulation of physical processes with emukit." arXiv preprint arXiv:2110.13293 (2021).
- Bingham, Eli, et al. "Pyro: Deep universal probabilistic programming." The Journal of Machine Learning Research 20.1 (2019): 973-978.
- Kandasamy, Kirthevasan, et al. "Parallelised Bayesian optimisation via Thompson sampling." International Conference on Artificial Intelligence and Statistics. PMLR, 2018.
- Settles, Burr. "Active learning literature survey." (2009).

References

- Martinez-Cantin, Ruben. "BayesOpt: a Bayesian optimization library for nonlinear optimization, experimental design and bandits." J. Mach. Learn. Res. 15.1 (2014): 3735-3739.
- Zhang, Mimi, et al. "Bayesian Optimisation for Sequential Experimental Design with Applications in Additive Manufacturing." arXiv preprint arXiv:2107.12809 (2021).