

Bayesian optimization

Unknown function $f(x)$

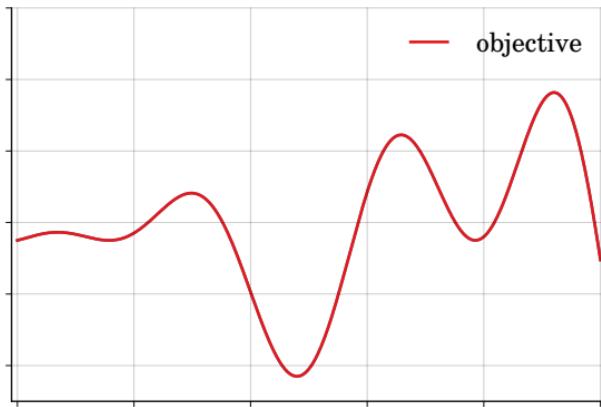
Goal : find $x^* = \operatorname{argmax} f(x)$

Setting : evaluating $y = f(x)$ is slow

- optimizing DL hyperparams
- drilling for oil

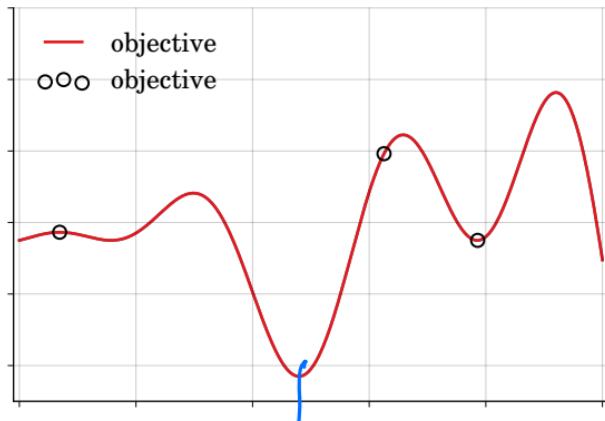
Idea : probabilistic model $p(y|x)$
as a surrogate for f .

Bayesian optimization cartoon



- ① Get initial sample.

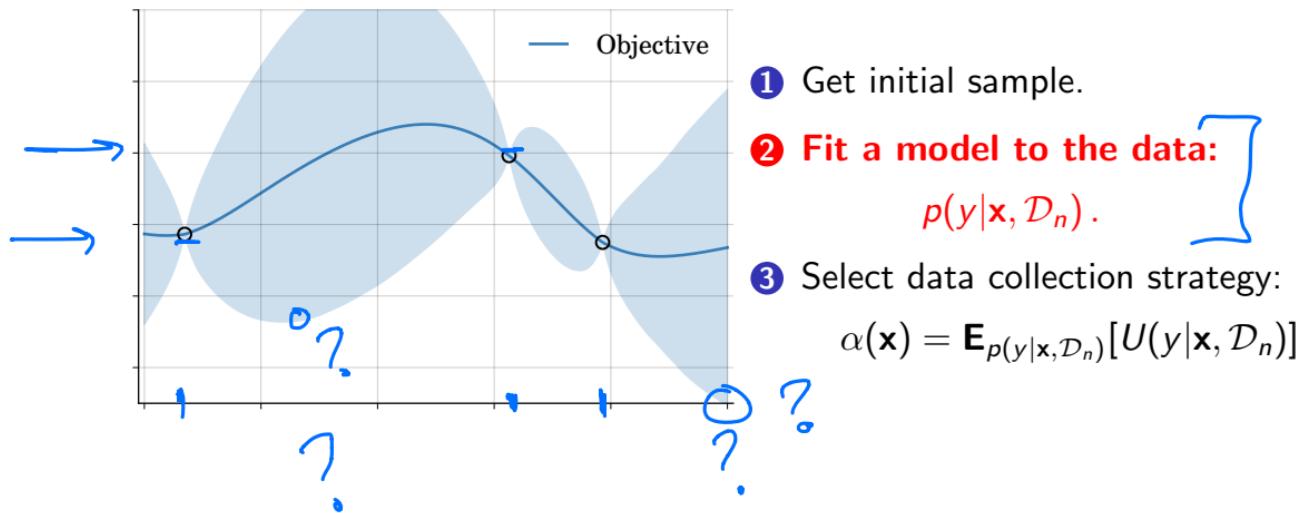
Bayesian optimization cartoon



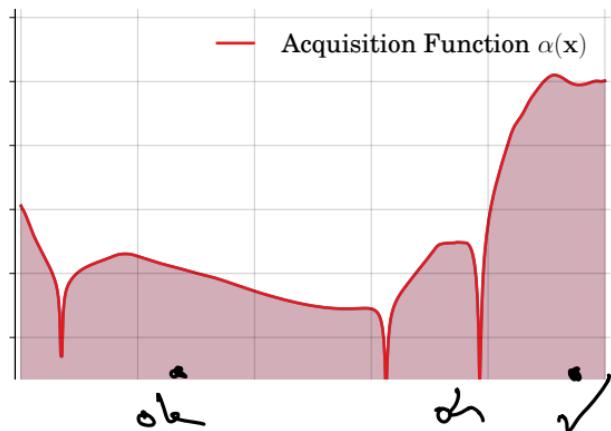
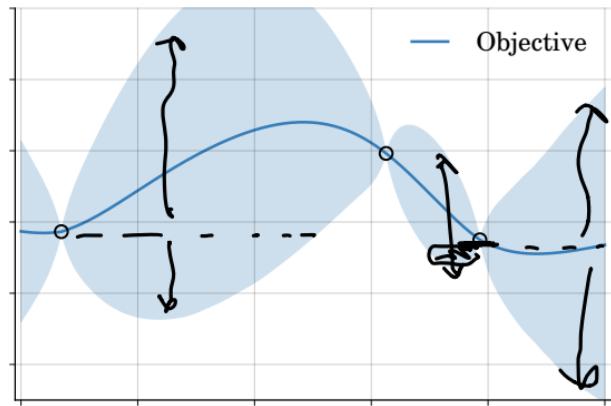
① Get initial sample.

$$\mathcal{D}_o = \{x, y\}$$

Bayesian optimization cartoon



Bayesian optimization cartoon



- ① Get initial sample.
- ② Fit a model to the data:

$$p(y|x, \mathcal{D}_n).$$

- ③ Select data collection strategy:

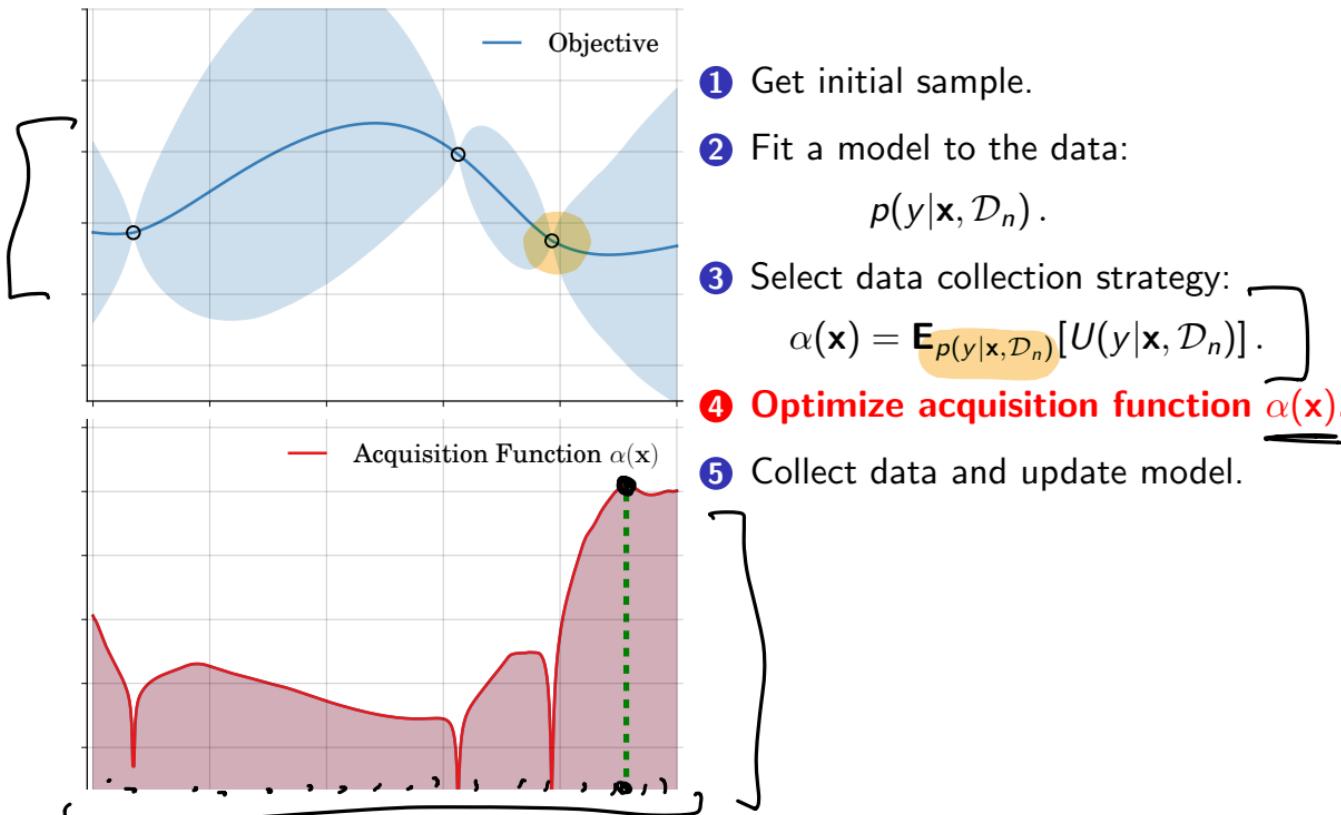
$$\alpha(x) = \mathbf{E}_{p(y|x, \mathcal{D}_n)}[\overline{U(y|x, \mathcal{D}_n)}].$$

- ④ Optimize acquisition function $\alpha(x)$.

Expected Improvement (EI)

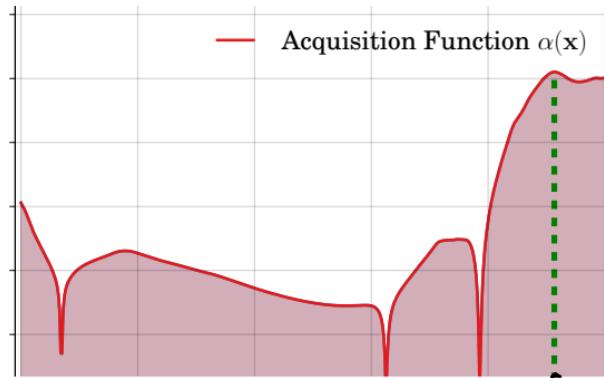
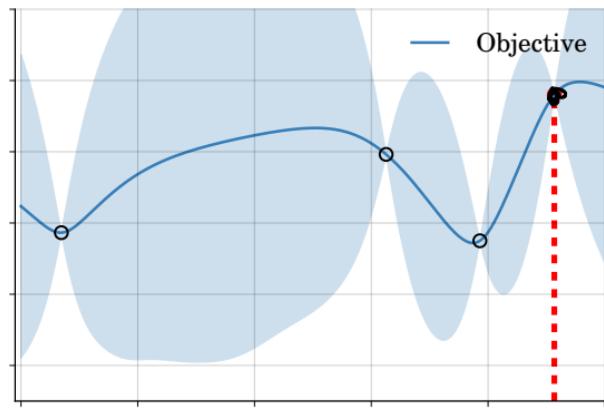
$$\mathbf{E}_{p(f_{\text{fb}})} \left[\max(0, f(x^*) - y^{\text{max}}) \right]$$

Bayesian optimization cartoon



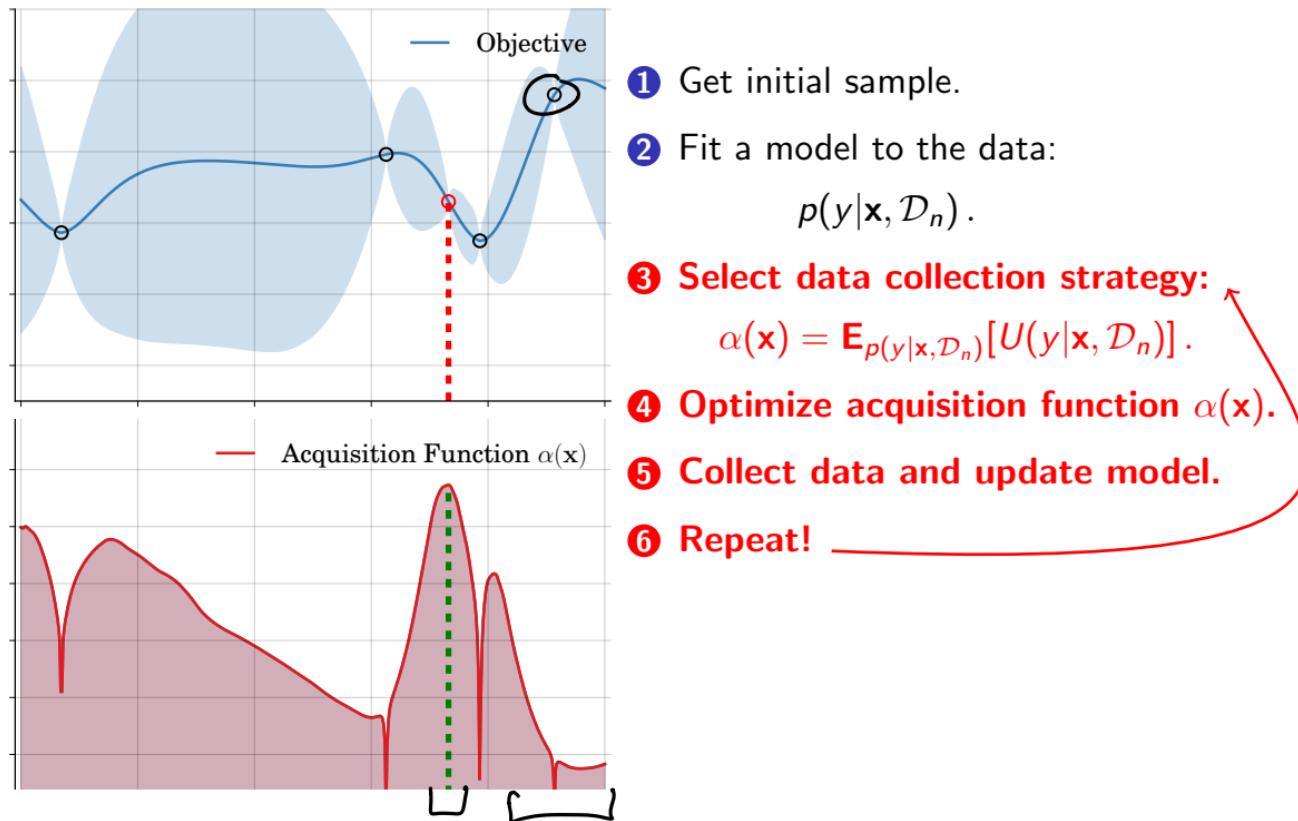
Figures: Miguel Hernández-Lobato

Bayesian optimization cartoon

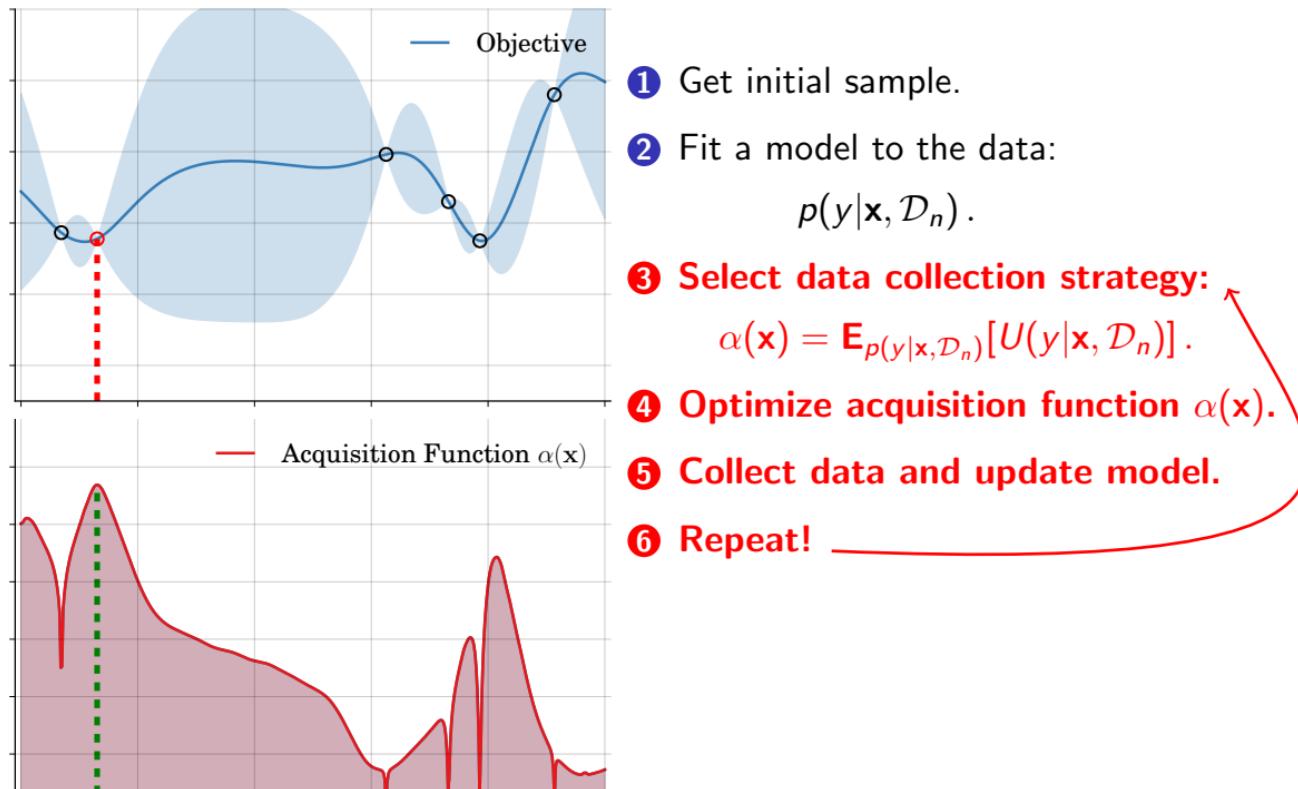


- ① Get initial sample.
- ② Fit a model to the data:
 $p(y|\mathbf{x}, \mathcal{D}_n)$.
- ③ Select data collection strategy:
 $\alpha(\mathbf{x}) = \mathbf{E}_{p(y|\mathbf{x}, \mathcal{D}_n)}[U(y|\mathbf{x}, \mathcal{D}_n)]$.
- ④ Optimize acquisition function $\alpha(\mathbf{x})$.
- ⑤ **Collect data and update model.**
- ⑥ Repeat!

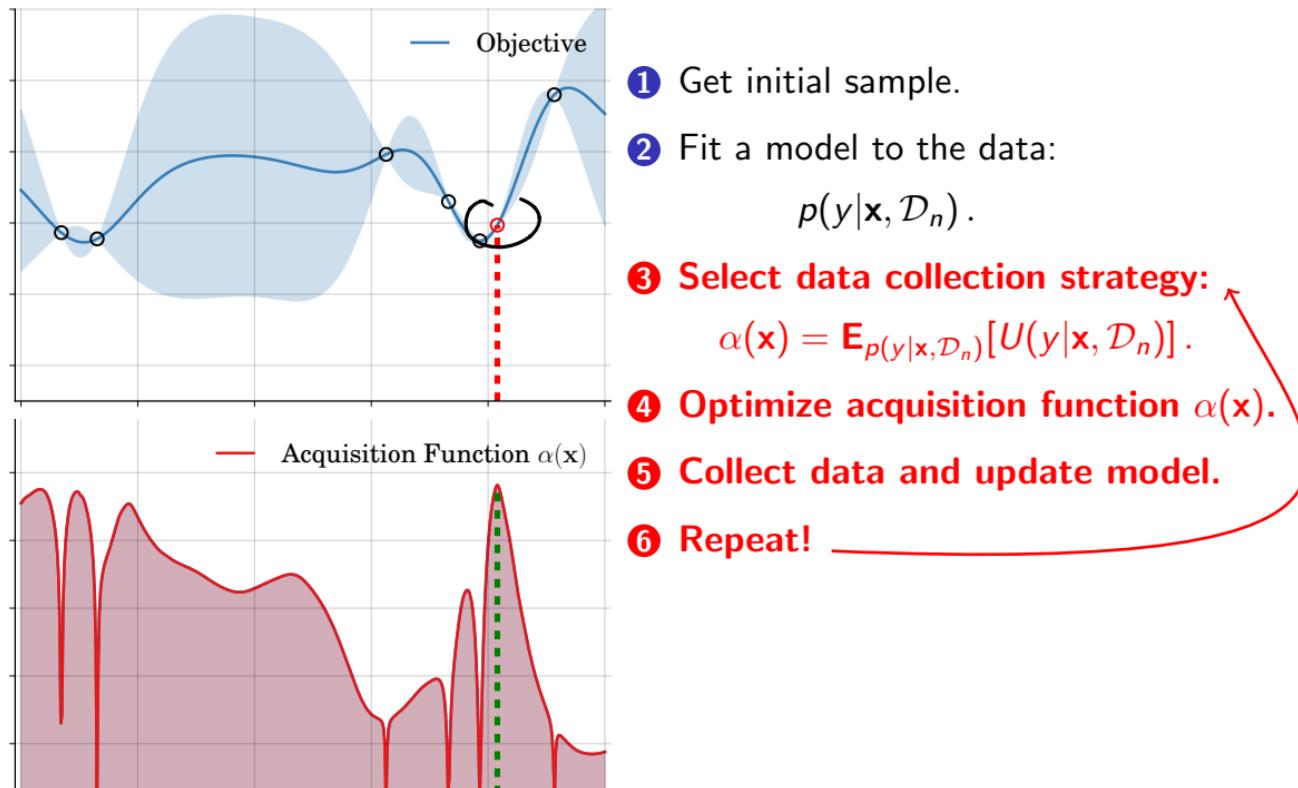
Bayesian optimization cartoon



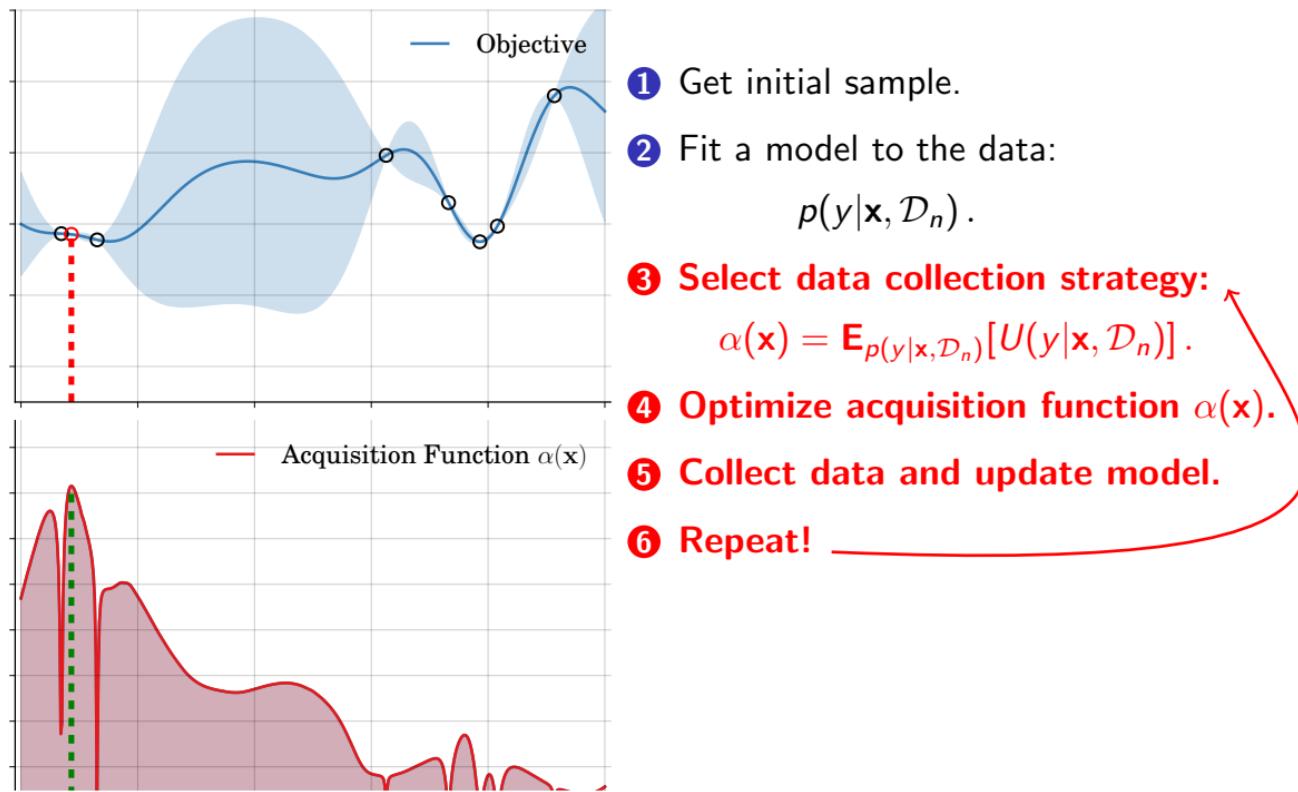
Bayesian optimization cartoon



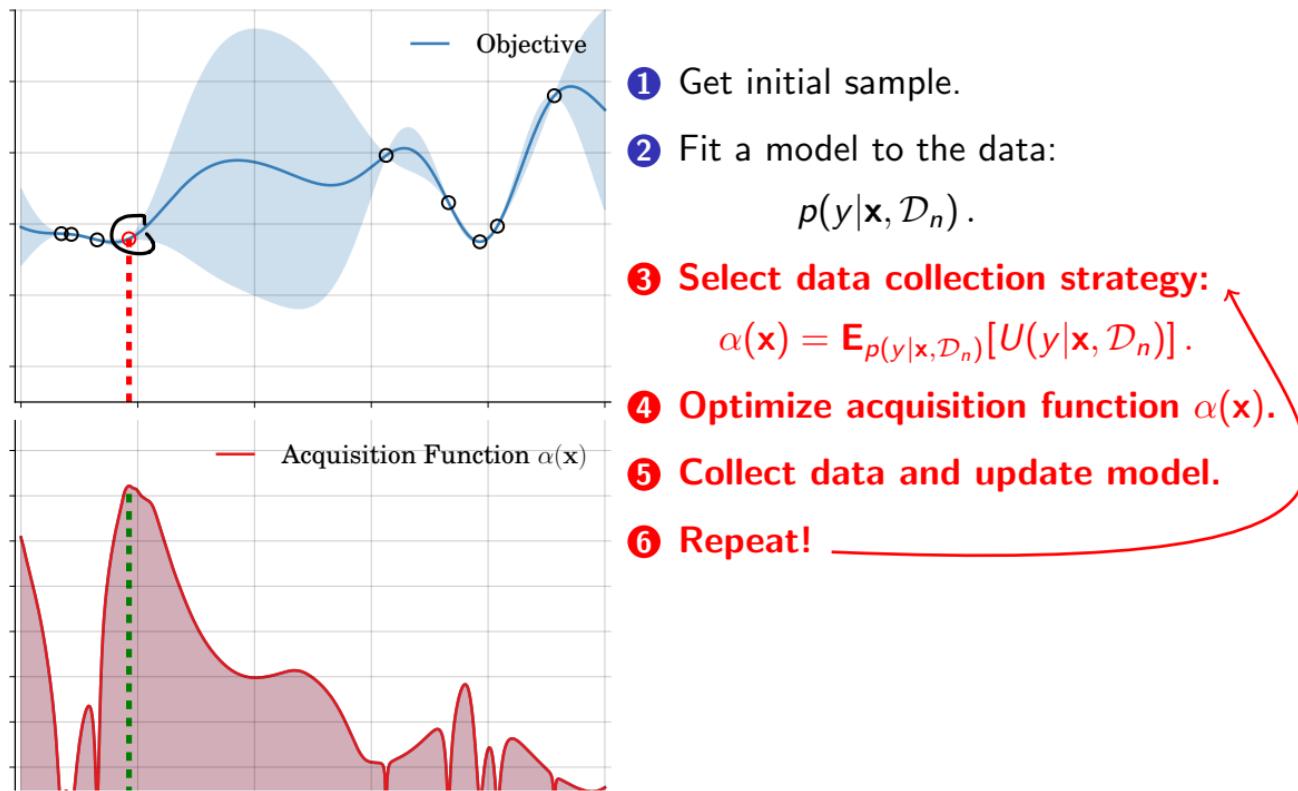
Bayesian optimization cartoon



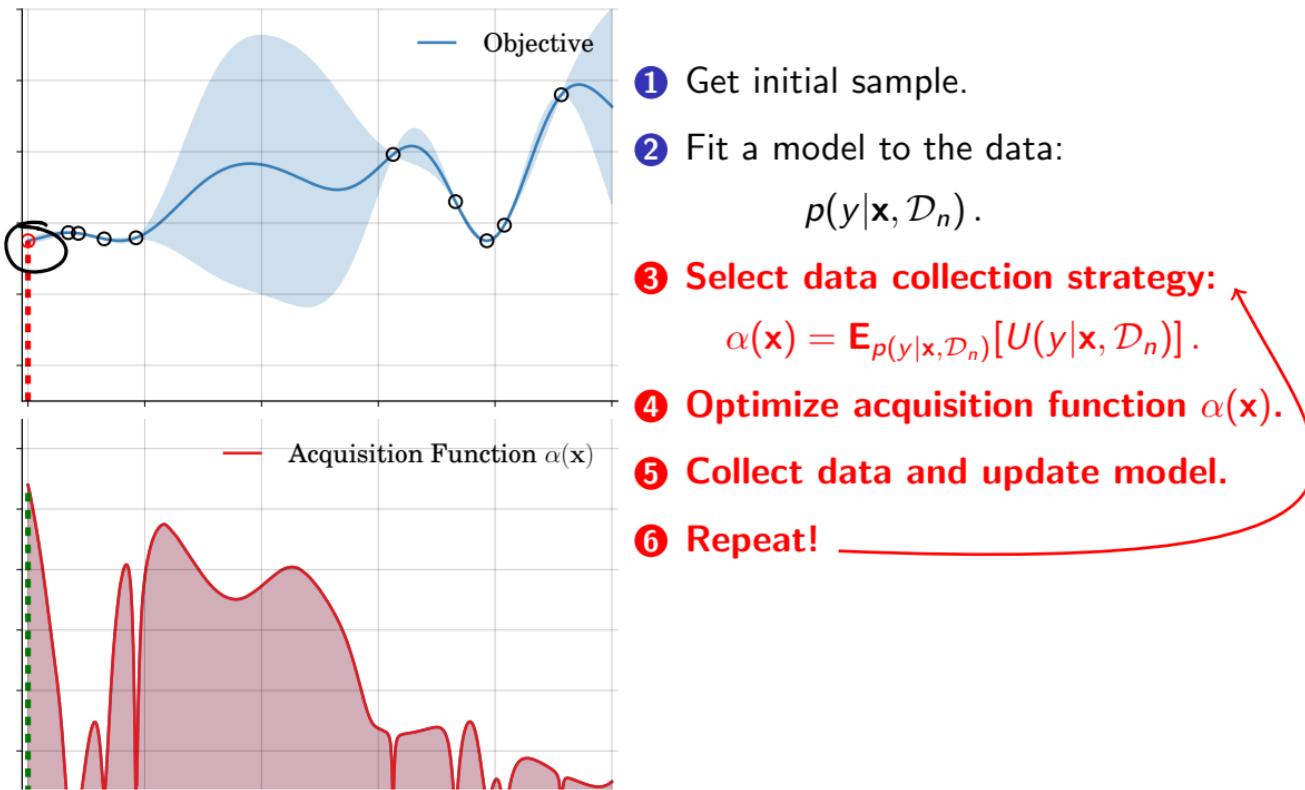
Bayesian optimization cartoon



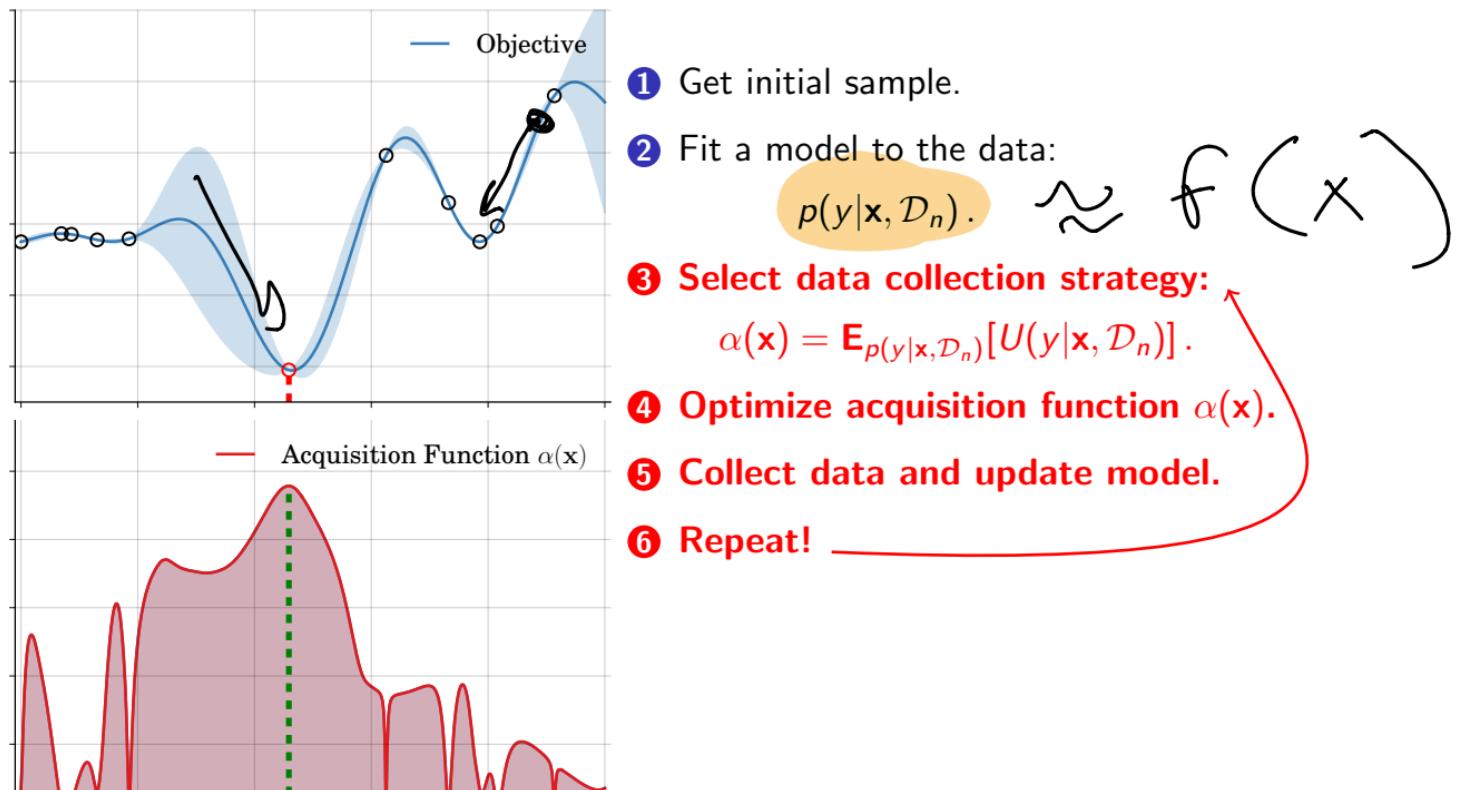
Bayesian optimization cartoon



Bayesian optimization cartoon



Bayesian optimization cartoon



Acquisition functions

$$\alpha_{EI}(x') = \underset{p(\theta|D)}{E} \left[\text{Imp}(x') \right] = \underset{p(\theta|D)}{E} \left[\max \left\{ 0, f(x, \theta) - y^* \right\} \right]$$

↑
max, at x^*

$$\alpha_{ES}(x') = H(x^* | D) - H(x^* | D \cup \{x', y'\})$$

- Predictive entropy search

- Max-value entropy search (y^*)

Deep Learning -

①

Use as surrogate.

$$f_{\text{ij}}(x', \mathcal{D})$$

GP?

BNN?

②

Bayes Opt has trouble in high dimensions of x .

$$p(y|z, \mathcal{D})$$



Soln : VAE :

$$- |z| \ll |x|$$

- z is continuous

do BO,
here,