

# ベイズ最適化

# Bayesian Optimization

# BO

明治大学 理工学部 応用化学科  
データ化学工学研究室 金子 弘昌

# ベイズ最適化 (BO) とは？

- ✓ ガウス過程による回帰により説明変数と目的変数との間で回帰モデルを構築する
  - ガウス過程による回帰の詳細はこちら  
<https://datachemeng.com/gaussianprocessregression/>
- ✓ 推定値とその分散を利用して、目的変数の値がより大きくなる (小さくなる) 可能性の高いサンプル候補を見つける
- ✓ 実験計画法やハイパーパラメータの最適化に活用できる
  - 実験計画法の詳細はこちら  
<https://datachemeng.com/designofexperimentscodes/>

# ベイズ最適化のための準備

✓ ガウス過程についてはこちら

<https://datachemeng.com/gaussianprocessregression/>

✓ 実験計画法についてはこちら

<https://datachemeng.com/designofexperimentscodes/>

# ベイズ最適化をするときの前提

- ✓ 説明変数と目的変数のあるサンプルがいくつか存在する
  - たとえば、実験条件 (説明変数) をいくつか変えて実験し、収率 (目的変数) を測定した結果がある
- ✓ 目的変数の値がなるべく**大きく**なるような説明変数の値の候補をしりたい
  - たとえば、収率がなるべく高くなるような実験条件をしりたい
  - 目的変数の値をなるべく小さくしたいときは、目的変数に-1をかければよい

# ガウス過程による回帰

- ✓説明変数と目的変数のあるサンプルを用いて、ガウス過程による回帰により回帰モデルを構築する

# 回帰モデルを用いた探索

- ✓ 回帰モデルがあれば、
  - 仮想的なサンプルとして説明変数の値をいろいろ変えてモデルに入力
  - 目的変数の推定値が最大になる説明変数の値を選択

                  ということができる

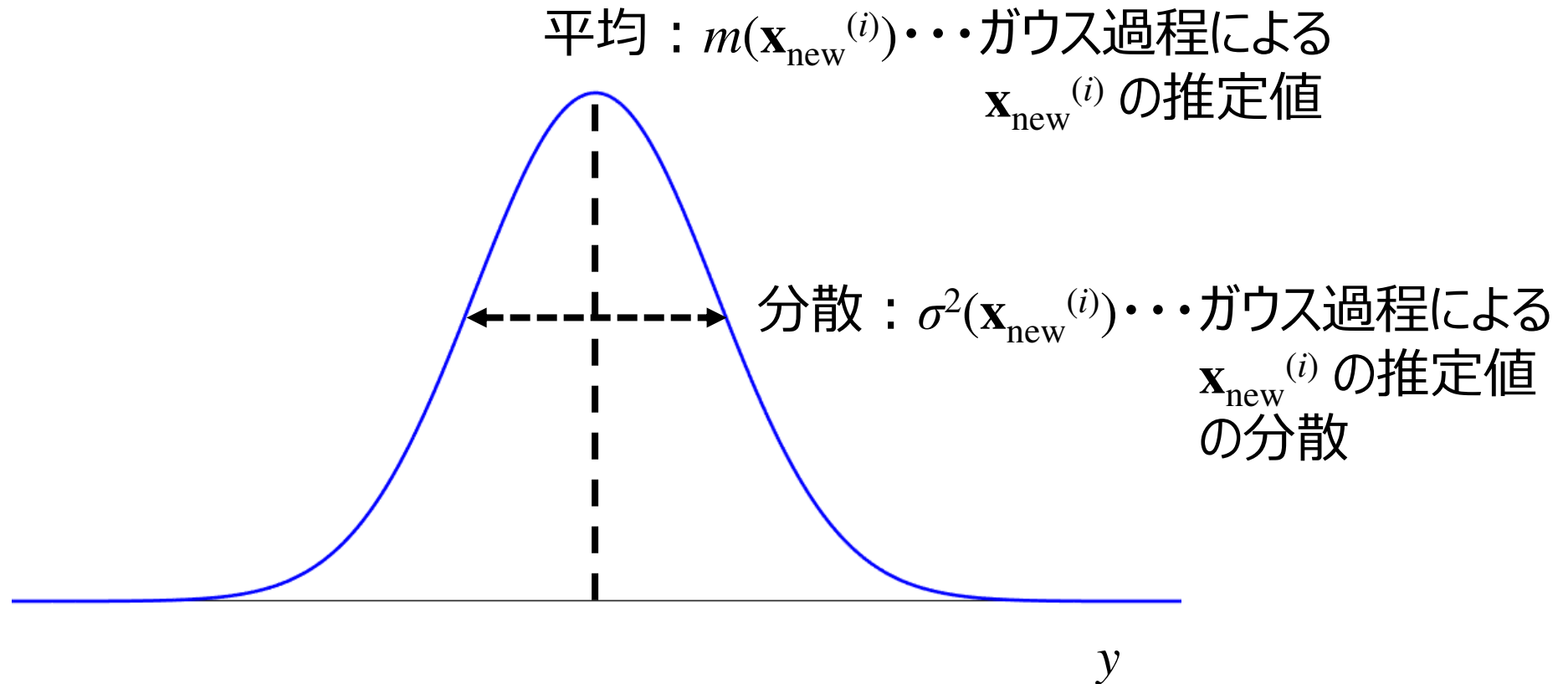
- ✓ ただ、この目的変数の推定値を最大化する方法では、既存のサンプルに近いところにあるものが、新しいサンプル候補として選ばれやすい
- ✓ 新しいサンプル候補の選択 → 実際に実験して目的変数(収率)を測定、を繰り返したとき、同じようなサンプル候補ばかりになってしまう
  - 面白みがない

# 獲得関数

- ✓ ベイズ最適化では、獲得関数が最大になる説明変数の値を選択
  
  - ✓ 獲得関数の計算には、目的変数の推定値だけでなく推定値のばらつきも利用
  
  - ✓ 獲得関数 (acquisition function)
    - Probability of Improvement (PI)
    - Expected Improvement (EI)
    - Mutual Information (MI)
- など

# Probability of Improvement (PI)

- ✓ 既存のサンプルにおける目的変数の最大値より大きくなる確率
- ✓ あるサンプル候補  $\mathbf{x}_{\text{new}}^{(i)}$  のPIの値を  $\text{PI}(\mathbf{x}_{\text{new}}^{(i)})$  とする
  - 正規分布を考える

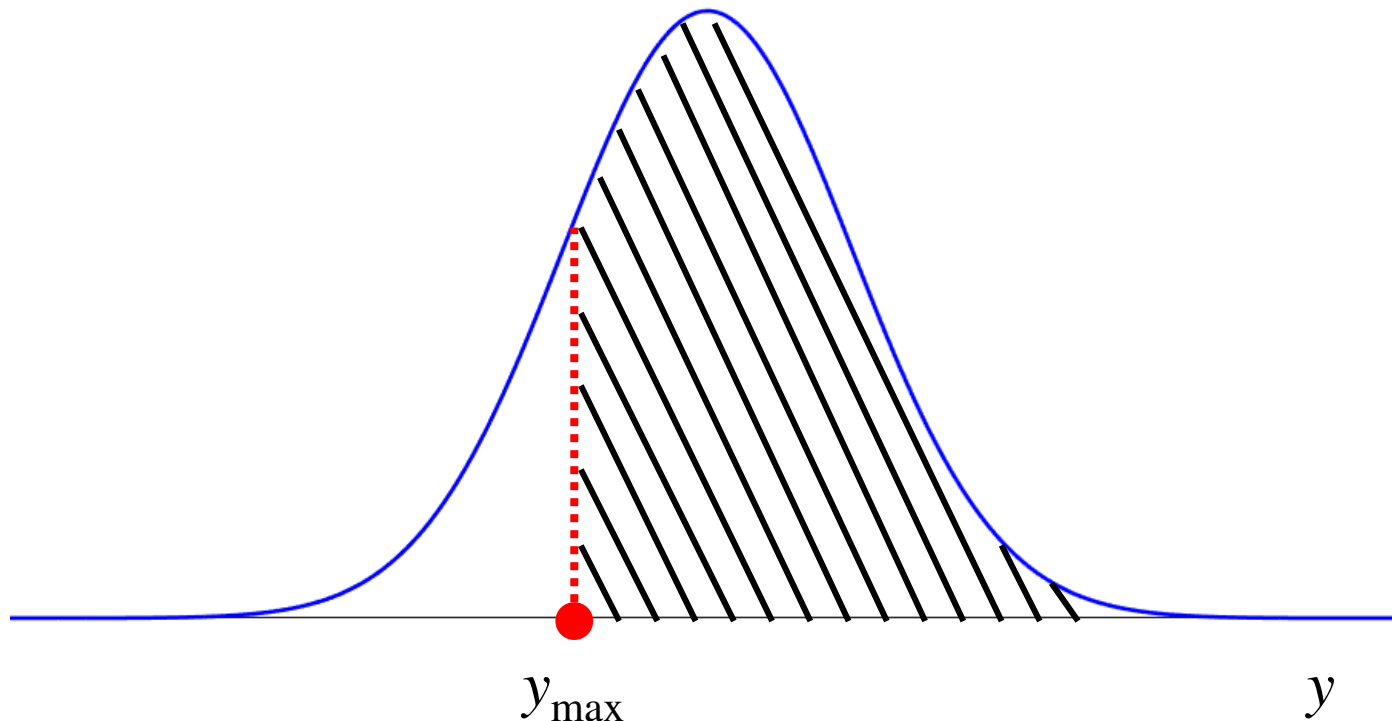




# PIの図解

✓  $PI(\mathbf{x}_{\text{new}}^{(i)})$  は、既存のサンプルにおける目的変数の最大値  $y_{\text{max}}$  から無限大まで正規分布を積分した値 (下の斜線の面積)

- 実際には、 $y_{\text{max}}$  ではなく、 $y_{\text{max}} + \varepsilon$ 
  - $m(\mathbf{x}_{\text{new}}^{(i)}) = y_{\text{max}}$  のとき  $PI(\mathbf{x}_{\text{new}}^{(i)}) = 0.5$  で最大になりうるため
  - たとえば、 $\varepsilon = 0.01 \times (\text{既存のサンプルの目的変数の標準偏差})$



$$\text{PI}(\mathbf{x}_{\text{new}}^{(i)}) = \int_{y_{\text{max}} + \varepsilon}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2(\mathbf{x}_{\text{new}}^{(i)})}} \exp\left\{-\frac{1}{2\sigma^2(\mathbf{x}_{\text{new}}^{(i)})} \left(x - m(\mathbf{x}_{\text{new}}^{(i)})\right)^2\right\} dx$$

# Expected Improvement (EI)

✓ 既存のサンプルにおける目的変数の最大値の更新幅の期待値

$$\begin{aligned} \text{EI}(\mathbf{x}_{\text{new}}^{(i)}) &= \left( m(\mathbf{x}_{\text{new}}^{(i)}) - y_{\text{max}} - \varepsilon \right) \text{PI}(\mathbf{x}_{\text{new}}^{(i)}) \\ &\quad + \sigma(\mathbf{x}_{\text{new}}^{(i)}) \frac{1}{\sqrt{2\pi\sigma^2(\mathbf{x}_{\text{new}}^{(i)})}} \exp \left\{ -\frac{1}{2\sigma^2(\mathbf{x}_{\text{new}}^{(i)})} \left( y_{\text{max}} + \varepsilon - m(\mathbf{x}_{\text{new}}^{(i)}) \right)^2 \right\} \end{aligned}$$

# Mutual Information (MI)

- ✓ 推定値 + “ばらつき”、が最大になるように選択し、“ばらつき”を実験回数ごとに更新していく
- ✓  $t$  回目の実験パラメータの候補を選択するとき

$$\text{MI}(\mathbf{x}_{\text{new}}^{(i)}) = m(\mathbf{x}_{\text{new}}^{(i)}) + \phi_t(\mathbf{x}_{\text{new}}^{(i)})$$

$$\phi_t(\mathbf{x}_{\text{new}}^{(i)}) = \sqrt{\alpha} \left( \sqrt{\sigma^2(\mathbf{x}_{\text{new}}^{(i)}) + \gamma_{t-1}} - \sqrt{\gamma_{t-1}} \right)$$

$$\gamma_t = \gamma_{t-1} + \sigma^2(\mathbf{x}_{\text{new}}^{(i)})$$

ただし、 $\gamma_0 = 0$ ,  $\alpha = \log \frac{2}{\delta}$       文献では、 $\delta = 10^{-6}$

E. Contal, V. Perchet, N. Vayatis, “Gaussian Process Optimization with Mutual Information”,  
arXiv:1311.4825v3

<https://arxiv.org/abs/1311.4825v3>

# 適応的な実験計画法

- ① PI もしくは EI もしくは MI の値が最大となるサンプル候補で、実験する
- ② 実験したら、そのサンプルを既存のサンプルに追加し、再度  
ガウス過程による回帰を行う

✓ ①② を繰り返すことで、目的変数の値を大きくすることを目指す

# どうやって実際にベイズ最適化するか？

✓ベイズ最適化をするための MATLAB や Python のプログラムを作りました！

[https://github.com/hkaneko1985/design\\_of\\_experiments](https://github.com/hkaneko1985/design_of_experiments)

# サンプルの候補が多いときは？

- ✓ 基本的には、グリッドサーチで新しいサンプル候補を探索する
  - 説明変数の数が3、各説明変数で値の候補が 10 あるとすると、サンプル候補の数は、 $10^3 = 1,000$
  
- ✓ サンプル候補の数が多いときは？
  - 説明変数の数が10、各説明変数で値の候補が 10 あるとすると、サンプル候補の数は、 $10^{10}$
  
  - ① ランダムにサンプリングする
    - たとえば、乱数で1,000,000サンプル候補を発生させる
  - ② 遺伝的アルゴリズムなどの最適化手法を使う