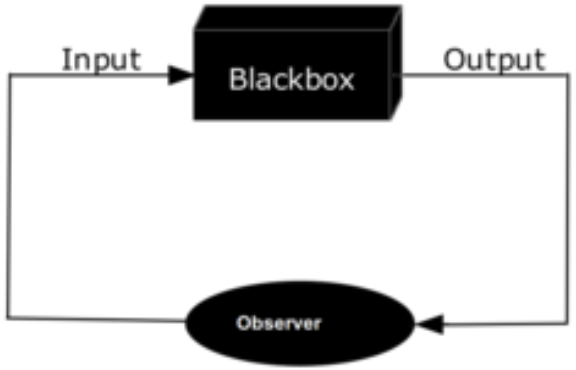
공학분야 연구/개발에서 흔히 나타나는 탐색 과정

1. **입력 결정:** 조성, 온도, 공정 시간, 형태, 강도… → **연속적, 다변수: '차원의 저주'**
2. **물성 측정:** 점도, 강성, 저항, 반응률, 속도 … → **물성/성능 '최적화'**
3. **평가 및 실험 재설계 → 요구 물성/성능 만족 여부 판단, 실험 '반복'**

- Grid search
- Model-based
- Heuristic / physics / insight

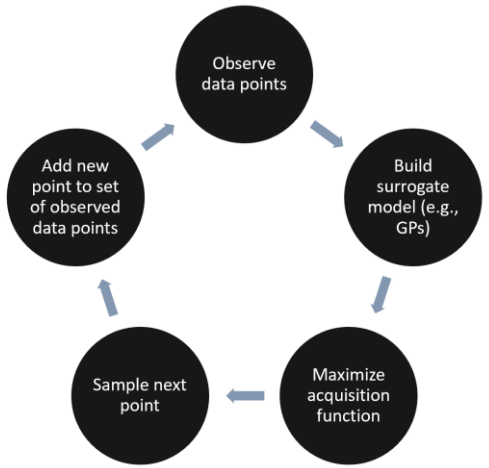
(Input) Parameter Searching



Q. 물리적/경험적 통찰이 없을 때, 측정된 데이터로부터 다음 실험을 어떻게 설계하는가?

Ans: 데이터 기반 모델링!

- Grid search보다 효율적
- 문제의 종류에 비의존적 (일관성)
- 데이터의 경향성 파악 가능



Bayesian optimization

Procedure

1. Given observed data (score vs. parameter)
2. Conduct GPR (mean & variance)
3. Calculate acquisition function (expected improvement, EI)
   : exploring & exploiting issue
4. Sample maximum EI point
5. Measure score for suggested point
6. Repeat 1-5

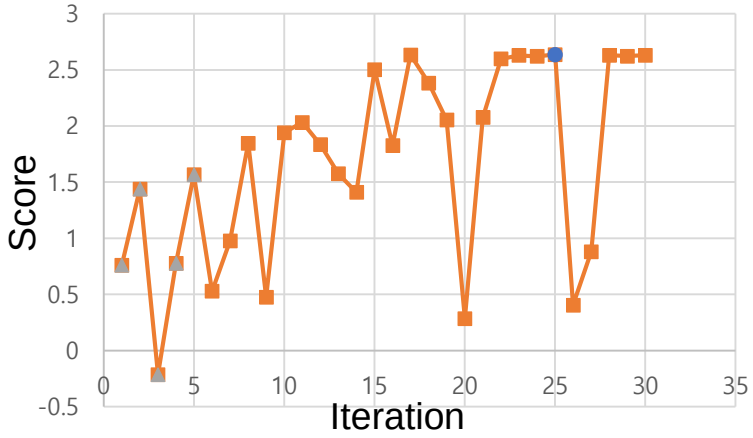
Expected improvement

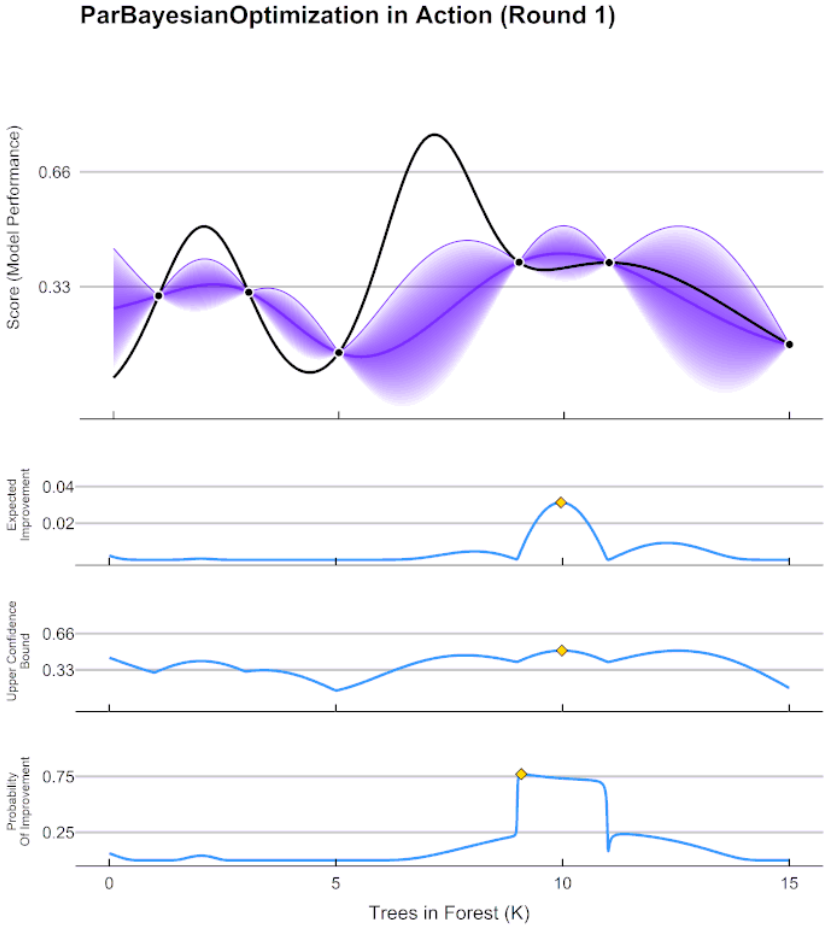
$$EI(x) = (\mu(x) - f(x_{\text{best}}))\Phi(Z) + \sigma\varphi(Z)$$

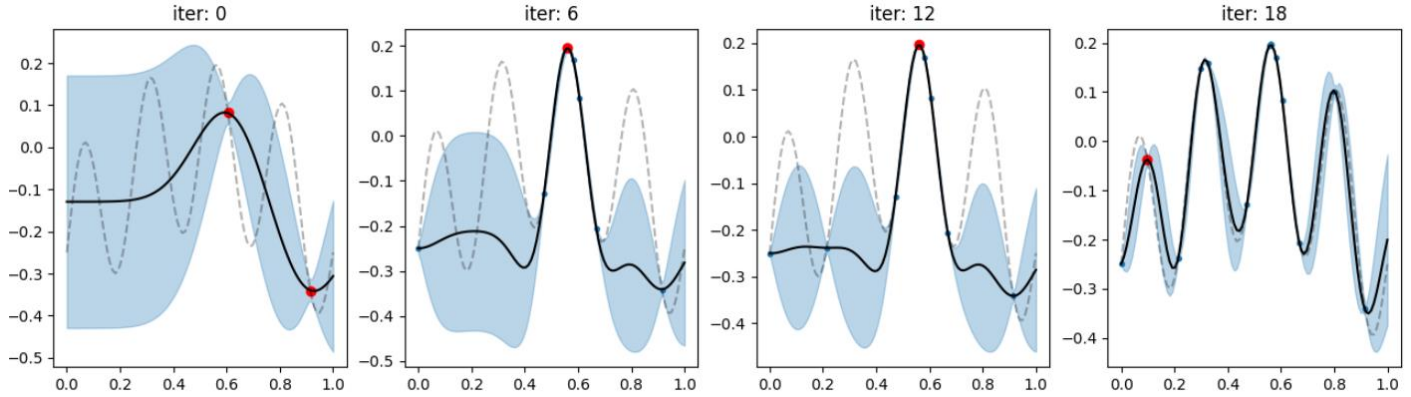
$$Z(x) = \frac{\mu(x) - f(x_{\text{best}})}{\sigma(x)}$$

$\Phi(Z)$: Cumulative normal distribution

$\varphi(Z)$: Normal distribution







Multi-objective problem?

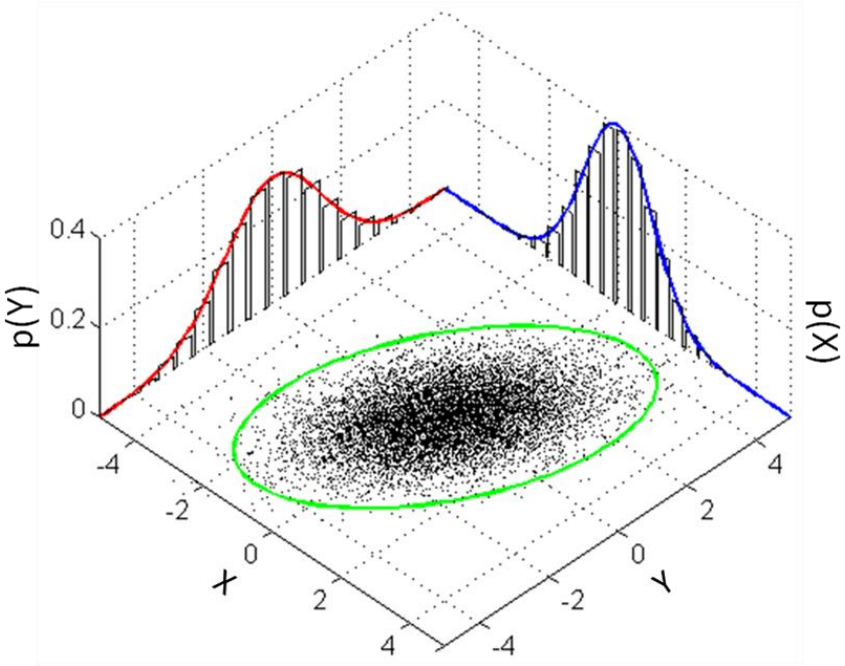
- Weighted sum of each score (Field knowledge)

- Multi-objective optimization (Pareto front)

Gaussian process regression

Multivariate normal distribution

$$N(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{8\pi^3 \det(\boldsymbol{\Sigma})} \exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})\cdot\boldsymbol{\Sigma}^{-1}\cdot(\mathbf{X}-\boldsymbol{\mu})\right)$$



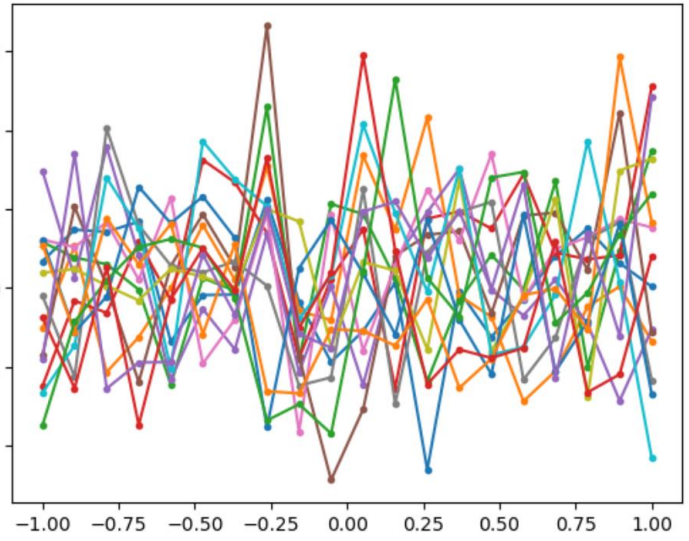
Function sampling

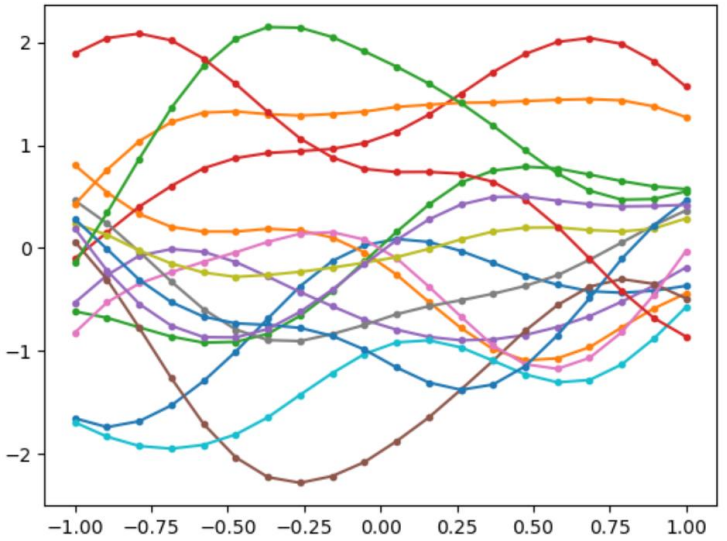
- Sampling from infinite dimensional distribution
- Covariance based on distance → smooth function
- Observed data → conditional sampling

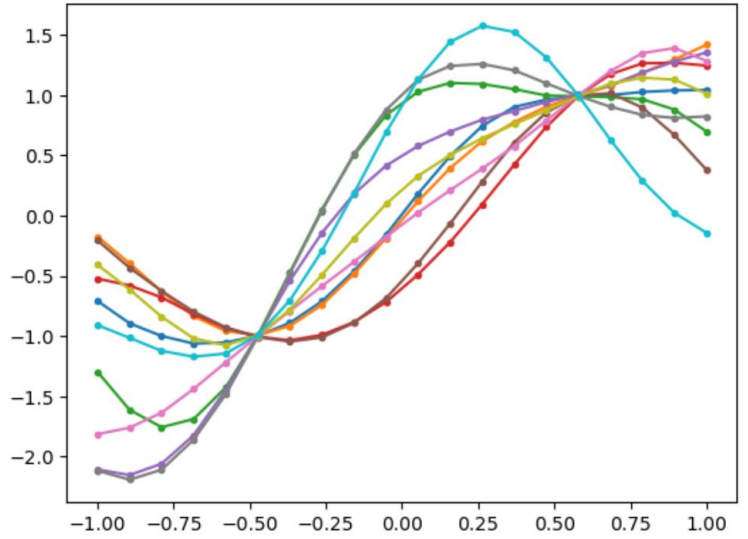
→ Prediction w/ mean & variance

$\mathbf{X} = y_{1\sim20}, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathbf{I}$ $\quad \boldsymbol{\mu} = \mathbf{0}, \Sigma_{ij} = A\exp(-d(x_i, x_j)/l)$ $\quad y_6 = -1, y_{16} = 1$

15 samples of **X**

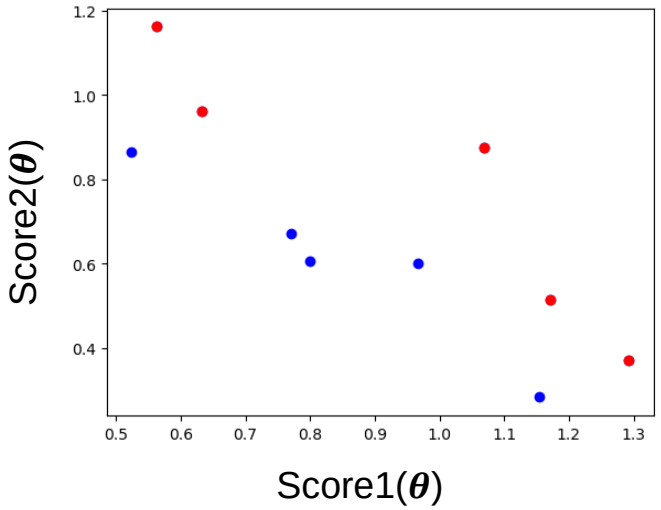


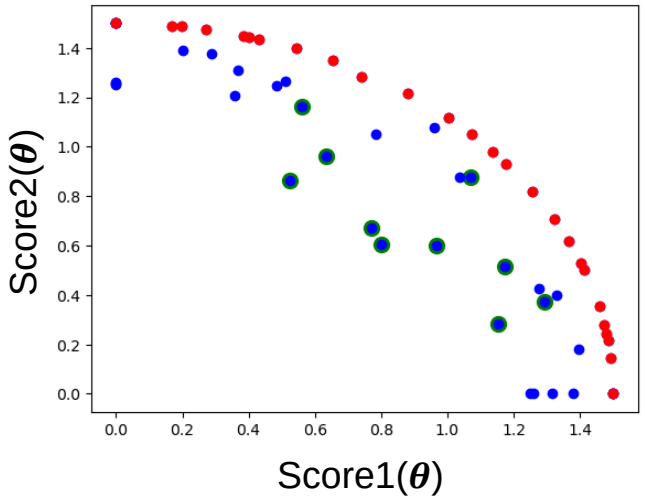




Multi objective Bayesian optimization

Pareto front





Procedure

1. Given observed data (score vs. parameter)
2. Conduct GPR for every score
3. Calculate acquisition function :**EHVI**
4. Sample maximum EHVI point
5. Measure scores for suggested point
6. Repeat 1-5

Expected Hyper Volume Improvement (EHVI)