Bayesian View on Bandit Problem (MDP formulation)

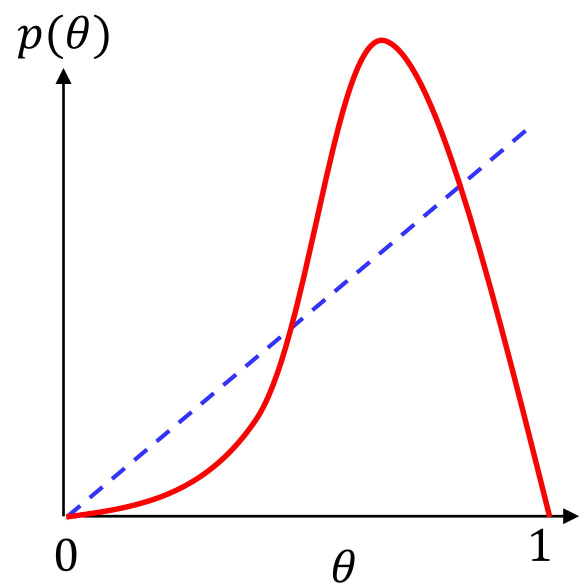
The posterior of the success probability given w_t winnings and l_t loss :

$$\theta_t | w_t, l_t \sim \mathrm{Beta}(\theta | \alpha = 1 + w_t, \beta = 1 + l_t)$$

The mean probability of success :

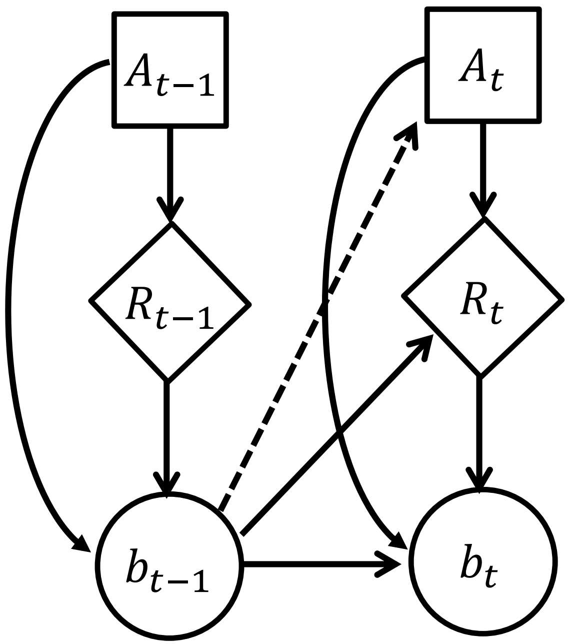
$$\rho_i = \int_0^1 \theta \, \mathrm{Beta}(\theta | \alpha = 1 + w_i, \beta = 1 + l_i) d\theta = \frac{w_i + 1}{w_i + l_i + 2}$$

	data	MLE	Bayesian
	w	$\theta = \frac{1}{1}$	$\theta \vert \mathrm{data} \sim \mathrm{Beta}(2,1)$
	w, w, w, l, w	$\theta = \frac{4}{5}$	$\theta \vert \mathrm{data} \sim \mathrm{Beta}(5,2)$



Bayesian View on Bandit Problem (MDP formulation)

MDP over belief state and finding optimal policy using Dynamic Programming



- $h_t = [(a_1, r_1), (a_1, r_1), \ldots, (a_t, r_t)]$
- $\theta = (\theta_i, \ldots, \theta_n)$: Unknown machine parameters
- Belief state $b_t(\theta) = P(\theta | h_t)$: probability dist. on para.
- Updating **belief state** $b_t(\theta)$ for Binary bandit with prior $\text{Beta}(\theta_i | \alpha, \beta)$: Deterministic

$$b_t(\theta_i) = b_{t-1}[a_t, r_t] = \text{Beta}\big(\theta_i | \alpha + w_{t,i}, \beta + l_{t,i}\big)$$

$w_{t,i}$: Accumulated wins with arm i up to time t

$l_{t,i}$: Accumulated loses with arm i up to time t

Dynamic programming on the value function

$$V_{t-1}(b_{t-1}) = \max_{\pi} E\left[\sum_{t=t}^{T} r_t\right]$$

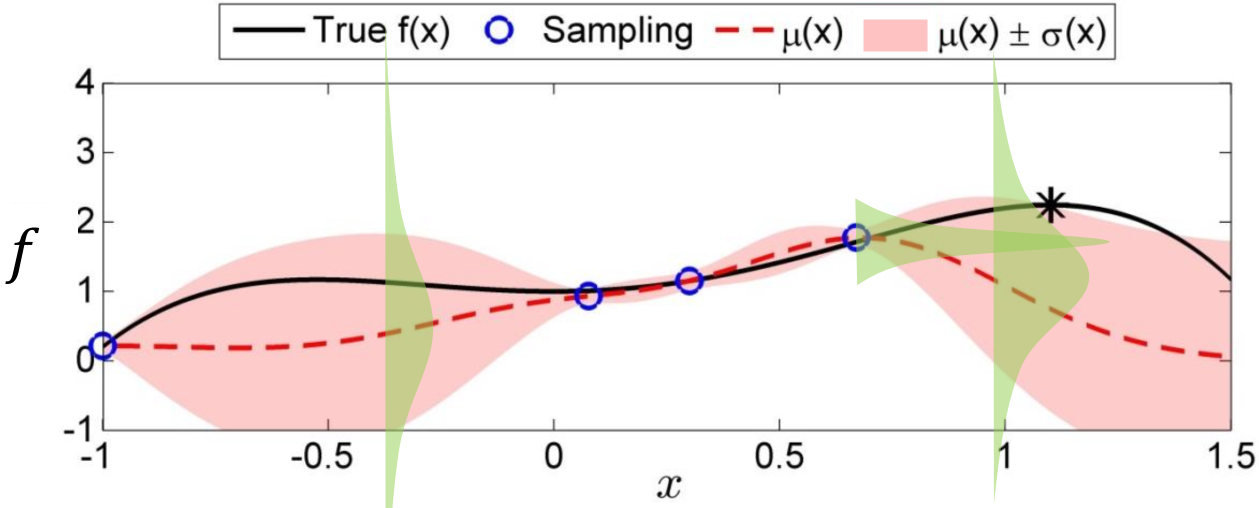
$$= \max_{a_t} \sum_{r_t} P(r_t | a_t, b_{t-1})[r_t + V_t(b_{t-1}[a_t, r_t]\,)]$$

$$P(r|a,b) = \int_{\theta_a} b(\theta_a) P(r|\theta_a)\, d\theta_a$$

∞ – armed Bandit Problem

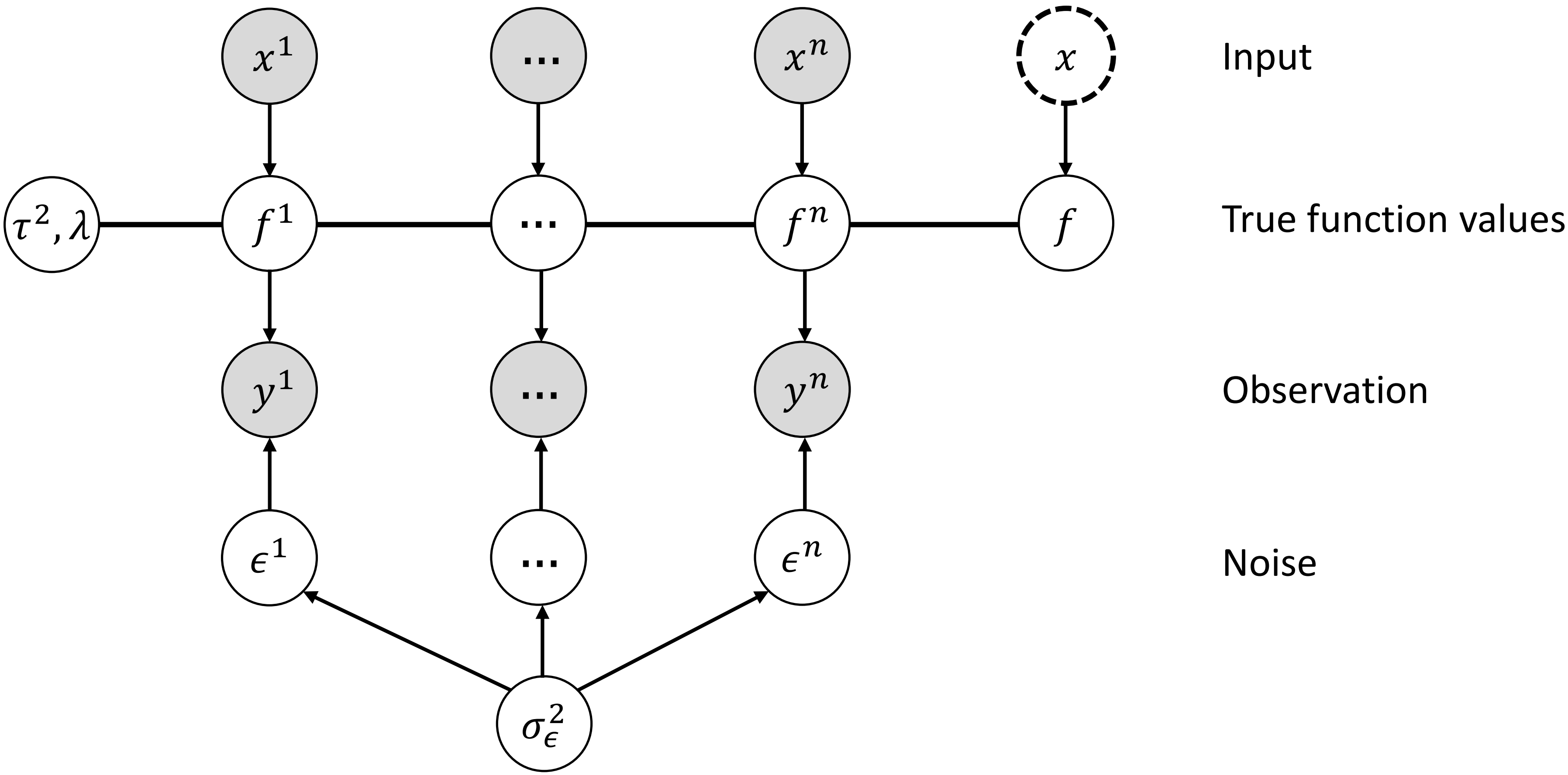
How to solve?

- **Learn target function (exploration)**
- **Improve target value (exploitation)**



Which machine should be selected?

Gaussian Process

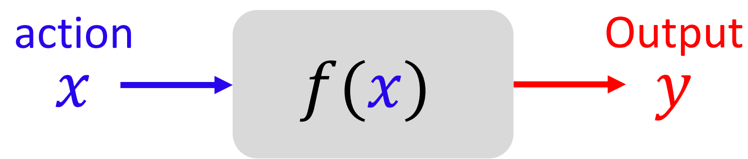


Bayesian Optimization

- **Bayesian Optimization (BO)** is a method to maximize (or minimize) a target value using measurement data from the **unknown target system.**

- BO is composed of three iterative steps:

  (1) Learning : construct a probabilistic model function for a target system
  (2) Optimization : select the next trial input to improve a target
  (3) Observation : measure the output of a target system



- Policy π maps all the history to new action:

$$\pi\colon [(x^1, y^1), (x^2, y^2), \dots, (x^{n-1}, y^{n-1})] \rightarrow x^n$$

- Find the optimal policy π^* that maximizes $E[\sum_{t=1}^{T} y^t]$ or $E[y^T]$

$$x^* = \underset{x}{\operatorname{argmax}}\, f(x)$$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

1. Given the data at nth iteration

$y^i = f^i + \epsilon^i$

Inputs	Latent function values	Observations
$\boldsymbol{x}^{1:n} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^i, \dots, \boldsymbol{x}^n\}$	$\boldsymbol{f}^{1:n} = \{f^1, \dots, f^i, \dots, f^n\}$	$\boldsymbol{y}^{1:n} = \{y^1, \dots, y^i, \dots, y^n\}$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

1. Given the data at nth iteration

Inputs	Latent function values	Observations
		$y^i = f^i + \epsilon^i$
$\boldsymbol{x}^{1:n} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^i, \dots, \boldsymbol{x}^n\}$	$\boldsymbol{f}^{1:n} = \{f^1, \dots, f^i, \dots, f^n\}$	$\boldsymbol{y}^{1:n} = \{y^1, \dots, y^i, \dots, y^n\}$

2. Prior on the function values $\boldsymbol{f}^{1:n}$ is represented as Gaussian Process (GP)

$$p(\boldsymbol{f}^{1:n}) = GP\big(m(\cdot), k(\cdot,\cdot)\big) : \quad \begin{bmatrix} f^1 \\ \vdots \\ f^n \end{bmatrix} \sim N\left(\begin{pmatrix} m(\boldsymbol{x}^1) \\ \vdots \\ m(\boldsymbol{x}^n) \end{pmatrix}, \begin{bmatrix} k(\boldsymbol{x}^1, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^1, \boldsymbol{x}^n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^n, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^n, \boldsymbol{x}^n) \end{bmatrix} \right)$$

$m(\cdot)$: mean function

$k(\cdot,\cdot)$: kernel function

$$m(\boldsymbol{x}) = \boldsymbol{0} \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \gamma \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}')^T \mathrm{diag}(\boldsymbol{\lambda})^{-2} (\boldsymbol{x} - \boldsymbol{x}') \right)$$

$$\mathrm{cov}\big(f^i, f^j\big) = \mathrm{E}\Big[\big(f^i - m(\boldsymbol{x}^i)\big)\big(f^j - m(\boldsymbol{x}^j)\big)\Big] \approx k\big(\boldsymbol{x}^i, \boldsymbol{x}^j\big)$$

Exponential Square kernel function: $k\big(\boldsymbol{x}^i, \boldsymbol{x}^j\big) = \gamma \exp\left(-\frac{1}{2} \big(\boldsymbol{x}^i - \boldsymbol{x}^j\big)^T \mathrm{diag}(\boldsymbol{\lambda})^{-2} \big(\boldsymbol{x}^i - \boldsymbol{x}^j\big) \right)$

Kernel function?

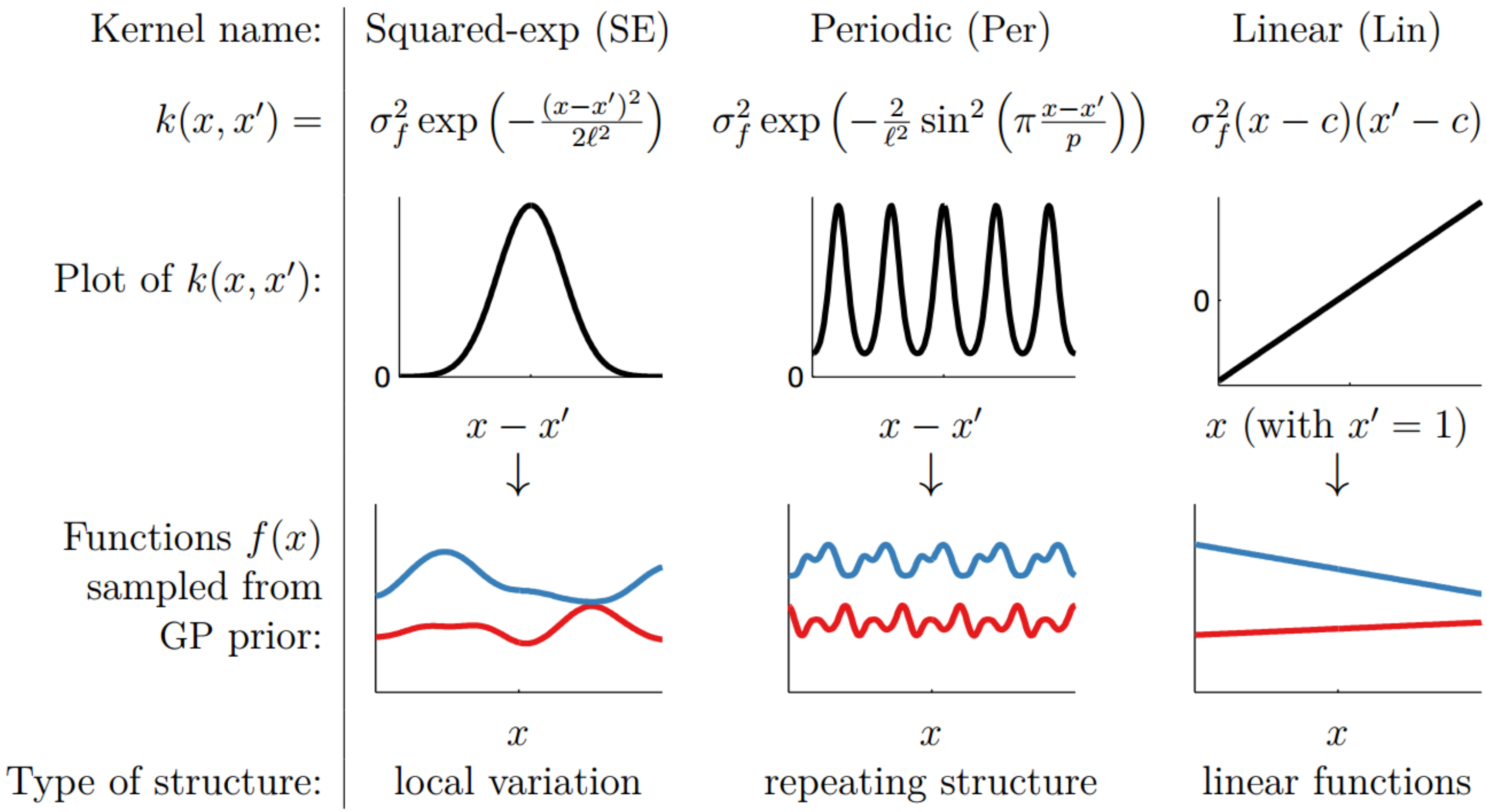


Figure form http://www.cs.toronto.edu/~duvenaud/

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

1. Given the data at nth iteration

Inputs	Latent function values	Observations
		$y^i = f^i + \epsilon^i$
$\boldsymbol{x}^{1:n} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^i, \dots, \boldsymbol{x}^n\}$	$\boldsymbol{f}^{1:n} = \{f^1, \dots, f^i, \dots, f^n\}$	$\boldsymbol{y}^{1:n} = \{y^1, \dots, y^i, \dots, y^n\}$

2. Prior on the function values $\boldsymbol{f}^{1:n}$ is represented as Gaussian Process (GP)

$$p(\boldsymbol{f}^{1:n}) = GP\big(m(\cdot), k(\cdot,\cdot)\big) : \quad \begin{bmatrix} f^1 \\ \vdots \\ f^n \end{bmatrix} \sim N\left(\begin{matrix} m(\boldsymbol{x}^1) \\ \vdots \\ m(\boldsymbol{x}^n) \end{matrix}, \begin{bmatrix} k(\boldsymbol{x}^1, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^1, \boldsymbol{x}^n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^n, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^n, \boldsymbol{x}^n) \end{bmatrix} \right)$$

$m(\cdot)$: mean function

$k(\cdot,\cdot)$: kernel function

$$m(\boldsymbol{x}) = \boldsymbol{0} \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \gamma \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}')^T \operatorname{diag}(\boldsymbol{\lambda})^{-2} (\boldsymbol{x} - \boldsymbol{x}') \right)$$

3. Likelihood is constructed base the assumption on the noise, i.e., i.i.d. Gaussian noise

$$p(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}) = N(\boldsymbol{f}^{1:n}, \sigma_\epsilon^2 \mathbf{I})$$

The hyper-parameters $\boldsymbol{\theta} = (\sigma_\epsilon, \sigma_s, \boldsymbol{\lambda})$ for the noise model and the kernel function are determined as ones maximizing the marginal log-likelihood of the training data $\boldsymbol{D}^n = \{(\boldsymbol{x}^i, y^i) | i = 1, \dots, n\}$ as

$$\begin{aligned} \boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log p\,(\boldsymbol{y}^{1:n} | \boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left(-\frac{1}{2} (\boldsymbol{y}^{1:n})^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \boldsymbol{y}^{1:n} - \frac{1}{2} \log |(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})| - \frac{n}{2} \log 2\pi \right) \end{aligned}$$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

1. Given the data at nth iteration

Inputs	Latent function values	Observations
		$y^i = f^i + \epsilon^i$
$\boldsymbol{x}^{1:n} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^i, \dots, \boldsymbol{x}^n\}$	$\boldsymbol{f}^{1:n} = \{f^1, \dots, f^i, \dots, f^n\}$	$\boldsymbol{y}^{1:n} = \{y^1, \dots, y^i, \dots, y^n\}$

2. Prior on the function values $\boldsymbol{f}^{1:n}$ is represented as Gaussian Process (GP)

$$p(\boldsymbol{f}^{1:n}) = GP\big(m(\cdot), k(\cdot,\cdot)\big) : \quad \begin{bmatrix} f^1 \\ \vdots \\ f^n \end{bmatrix} \sim N\left(\begin{matrix} m(\boldsymbol{x}^1) \\ \vdots \\ m(\boldsymbol{x}^n) \end{matrix}, \begin{bmatrix} k(\boldsymbol{x}^1, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^1, \boldsymbol{x}^n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^n, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^n, \boldsymbol{x}^n) \end{bmatrix} \right)$$

$m(\cdot)$: mean function

$k(\cdot,\cdot)$: kernel function

$$m(\boldsymbol{x}) = \boldsymbol{0} \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \gamma \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}')^T \operatorname{diag}(\boldsymbol{\lambda})^{-2} (\boldsymbol{x} - \boldsymbol{x}') \right)$$

3. Likelihood is constructed base the assumption on the noise, i.e., i.i.d. Gaussian noise

$$p(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}) = N(\boldsymbol{f}^{1:n}, \sigma_\epsilon^2 \mathbf{I})$$

4. Joint distribution based on Bayes' rule :

$$p(f, \boldsymbol{y}^{1:n}) = \int p\big(f, \boldsymbol{f}^{1:n}\big) p\big(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}\big)\, d\boldsymbol{f}^{1:n} \quad \Rightarrow \quad \begin{bmatrix} \boldsymbol{y}^{1:n} \\ f \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

Property 4 : Conditionals of a GRV are Gaussians, more specifically, if

$$Z = \begin{bmatrix} Y_1 \\ - \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} Y_1 \\ - \\ Y_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & | & \Sigma_{12} \\ - - & & - - \\ \Sigma_{21} & | & \Sigma_{22} \end{bmatrix} \right)$$

where Y_1 is k-dim RV and Y_2 is an $n-k$ dim RV, then

$$Y_2 | \{Y_1 = y\} \sim N(\Sigma_{21}\Sigma_{11}^{-1}(y-\mu_1)+\mu_2, \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

4. Joint distribution based on Bayes' rule :

$$p(f, \boldsymbol{y}^{1:n}) = \int p\left(f, \boldsymbol{f}^{1:n}\right) p\left(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}\right) d\boldsymbol{f}^{1:n} \quad \Rightarrow \quad \begin{bmatrix} \boldsymbol{y}^{1:n} \\ f \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$

5. Conditionalization→ Posterior distribution on the function value $f = f(x)$ for x given $\boldsymbol{D}^n = \{\boldsymbol{x}^{1:n}, \boldsymbol{y}^{1:n}\}$

$$p(f | \boldsymbol{D}^n) = N\left(\mu(\boldsymbol{x}|\boldsymbol{D}^n), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^n)\right)$$

Mean : $\mu(\boldsymbol{x}|\boldsymbol{D}^n) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$

Variance : $\sigma^2(\boldsymbol{x}|\boldsymbol{D}^n) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

1. Given the data at nth iteration

Inputs	Latent function values	Observations
		$y^i = f^i + \epsilon^i$
$\boldsymbol{x}^{1:n} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^i, \dots, \boldsymbol{x}^n\}$	$\boldsymbol{f}^{1:n} = \{f^1, \dots, f^i, \dots, f^n\}$	$\boldsymbol{y}^{1:n} = \{y^1, \dots, y^i, \dots, y^n\}$

2. Prior on the function values $\boldsymbol{f}^{1:n}$ is represented as Gaussian Process (GP)

$$p(\boldsymbol{f}^{1:n}) = GP\big(m(\cdot), k(\cdot,\cdot)\big) : \quad \begin{bmatrix} f^1 \\ \vdots \\ f^n \end{bmatrix} \sim N\left(\begin{matrix} m(\boldsymbol{x}^1) \\ \vdots \\ m(\boldsymbol{x}^n) \end{matrix}, \begin{bmatrix} k(\boldsymbol{x}^1, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^1, \boldsymbol{x}^n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^n, \boldsymbol{x}^1) & \cdots & k(\boldsymbol{x}^n, \boldsymbol{x}^n) \end{bmatrix} \right)$$

$m(\cdot)$: mean function

$k(\cdot,\cdot)$: kernel function

$$m(\boldsymbol{x}) = \boldsymbol{0} \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \gamma \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}')^T \mathrm{diag}(\boldsymbol{\lambda})^{-2} (\boldsymbol{x} - \boldsymbol{x}') \right)$$

3. Likelihood is constructed base the assumption on the noise, i.e., i.i.d. Gaussian noise

$$p(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}) = N(\boldsymbol{f}^{1:n}, \sigma_\epsilon^2 \mathbf{I})$$

4. Joint distribution based on Bayes' rule :

$$p(f, \boldsymbol{y}^{1:n}) = \int p\big(f, \boldsymbol{f}^{1:n}\big) p\big(\boldsymbol{y}^{1:n} | \boldsymbol{f}^{1:n}\big)\, d\boldsymbol{f}^{1:n} \quad \Rightarrow \quad \begin{bmatrix} \boldsymbol{y}^{1:n} \\ f \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$

5. Conditionalization→ Posterior distribution on the function value $f = f(x)$ for x given $\boldsymbol{D}^n = \{\boldsymbol{x}^{1:n}, \boldsymbol{y}^{1:n}\}$

$$p(f | \boldsymbol{D}^n) = N\big(\mu(\boldsymbol{x} | \boldsymbol{D}^n), \sigma^2(\boldsymbol{x} | \boldsymbol{D}^n)\big)$$

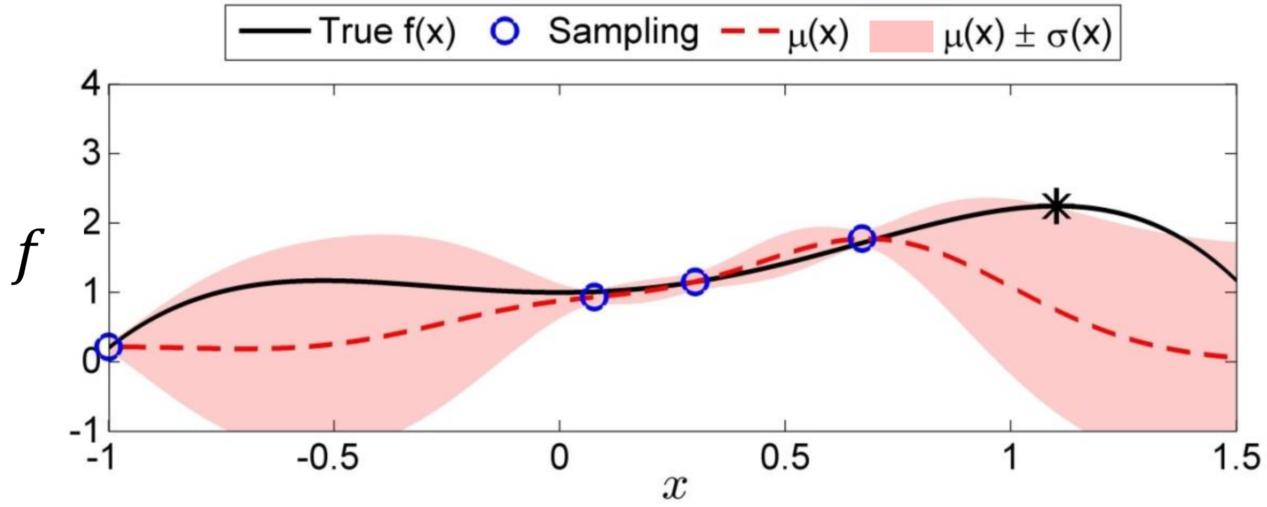
Mean : $\mu(\boldsymbol{x} | \boldsymbol{D}^n) = \boldsymbol{k}^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \boldsymbol{y}^{1:n}$

Variance : $\sigma^2(\boldsymbol{x} | \boldsymbol{D}^n) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \boldsymbol{k}$

Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

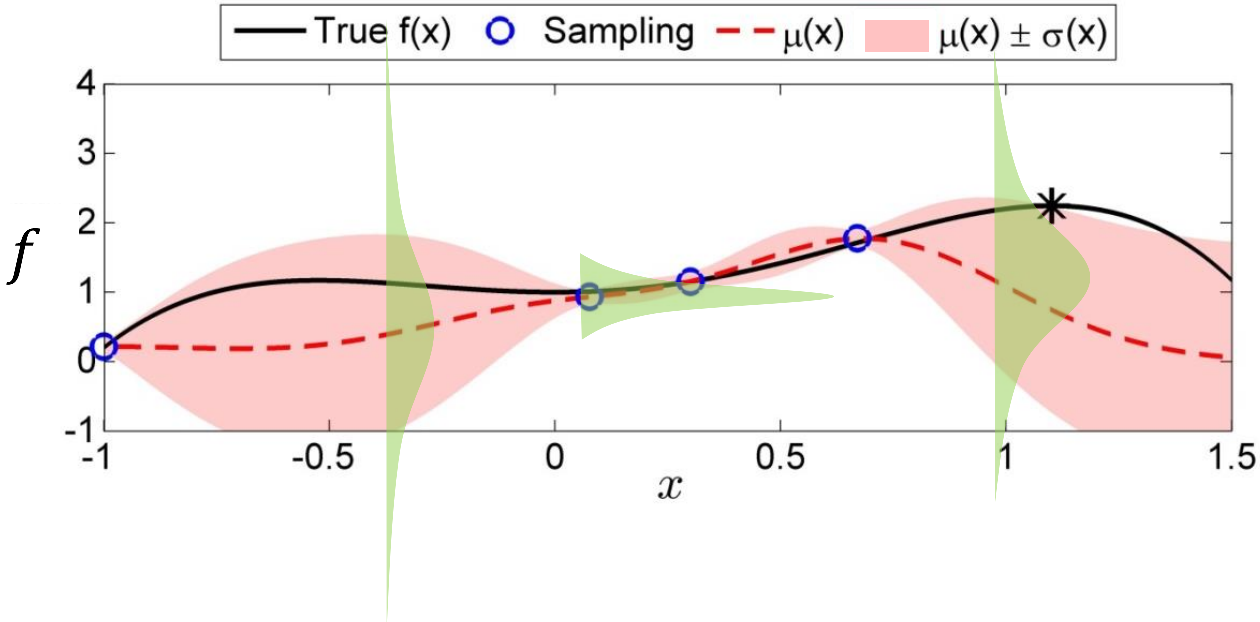
$$p(f|\boldsymbol{D}^n) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^n), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^n)\big)$$



Learning phase : Gaussian Process (GP) regression

Construct the distribution on unknown target value $f = f(\boldsymbol{x})$ corresponding $\boldsymbol{x}$

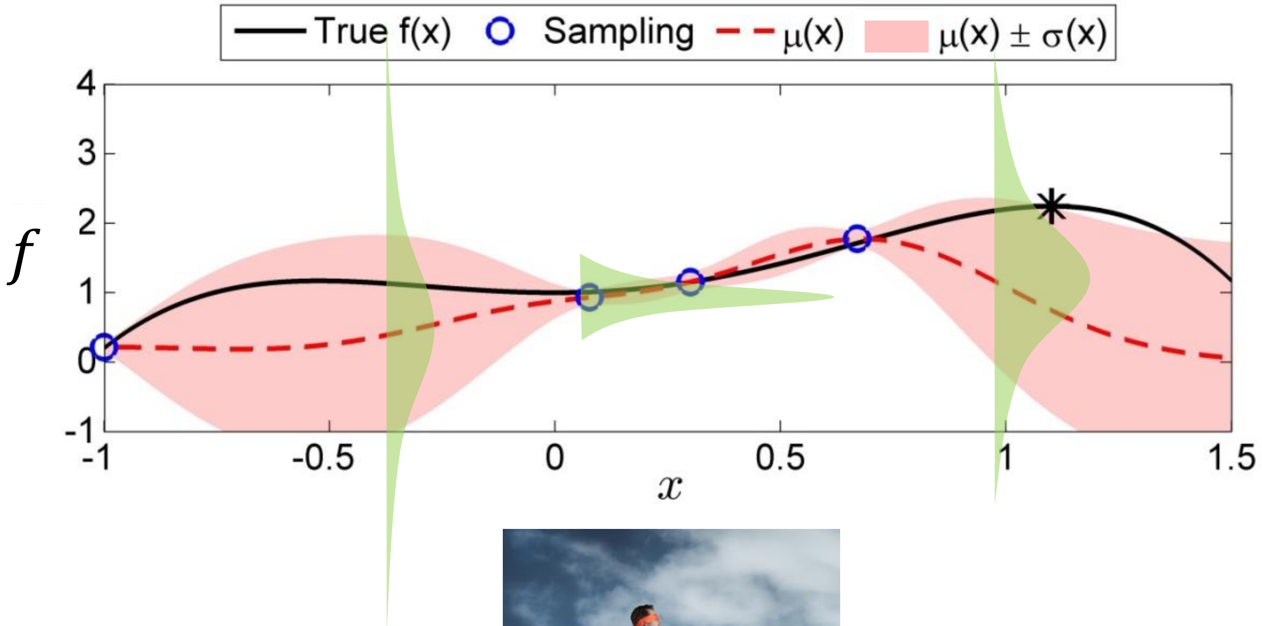
$$p(f|\boldsymbol{D}^n) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^n), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^n)\big)$$



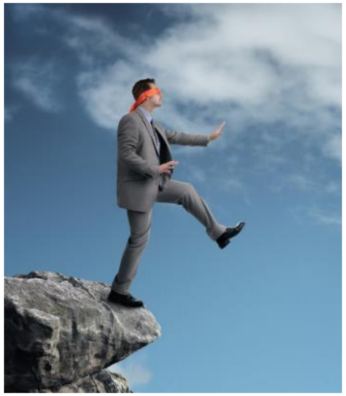
Optimization (Sampling) phase

How to select the next input ?

- **Learn target function (exploration)**
- **Improve target value (exploitation)**



Go to a graduate school to explore my intellectual capability?

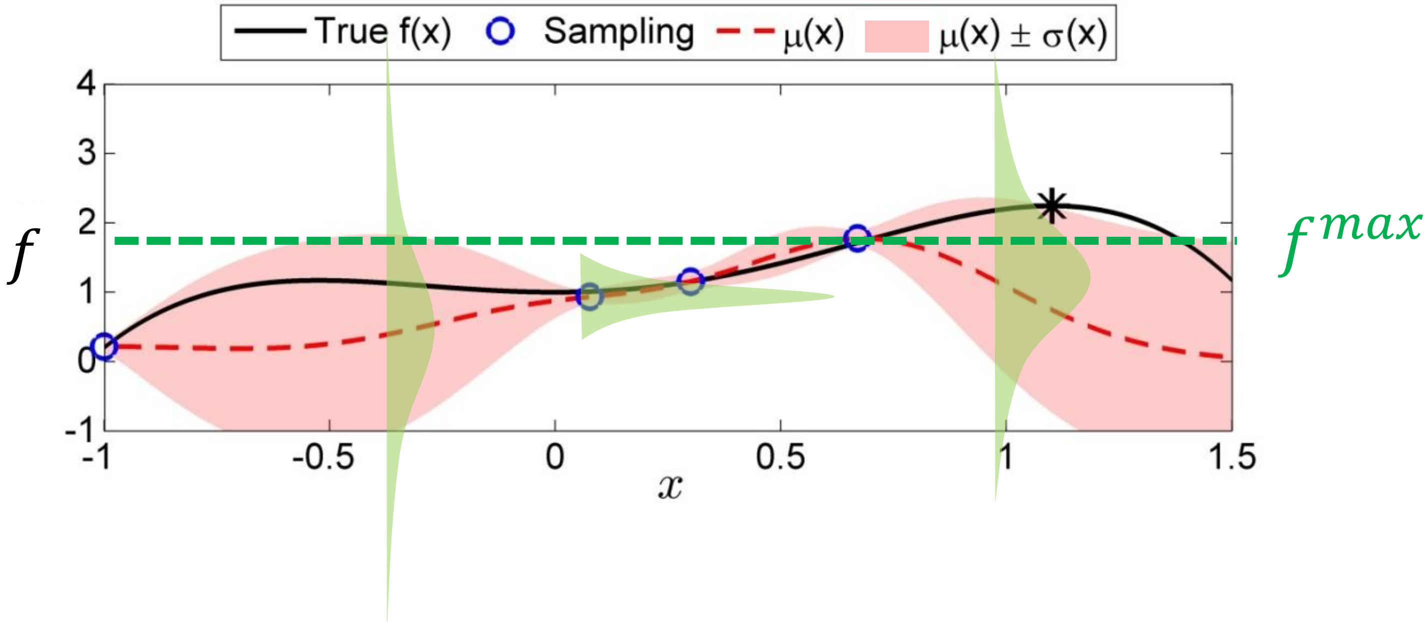


Go to a company to make money?

Optimization (Sampling) phase

How to select the next input ?

- **Learn target function (exploration)**
- **Improve target value (exploitation)**



Next sampling point $\boldsymbol{x}^{n+1}$ is determined by solving (Mockus *et al*, 1978)

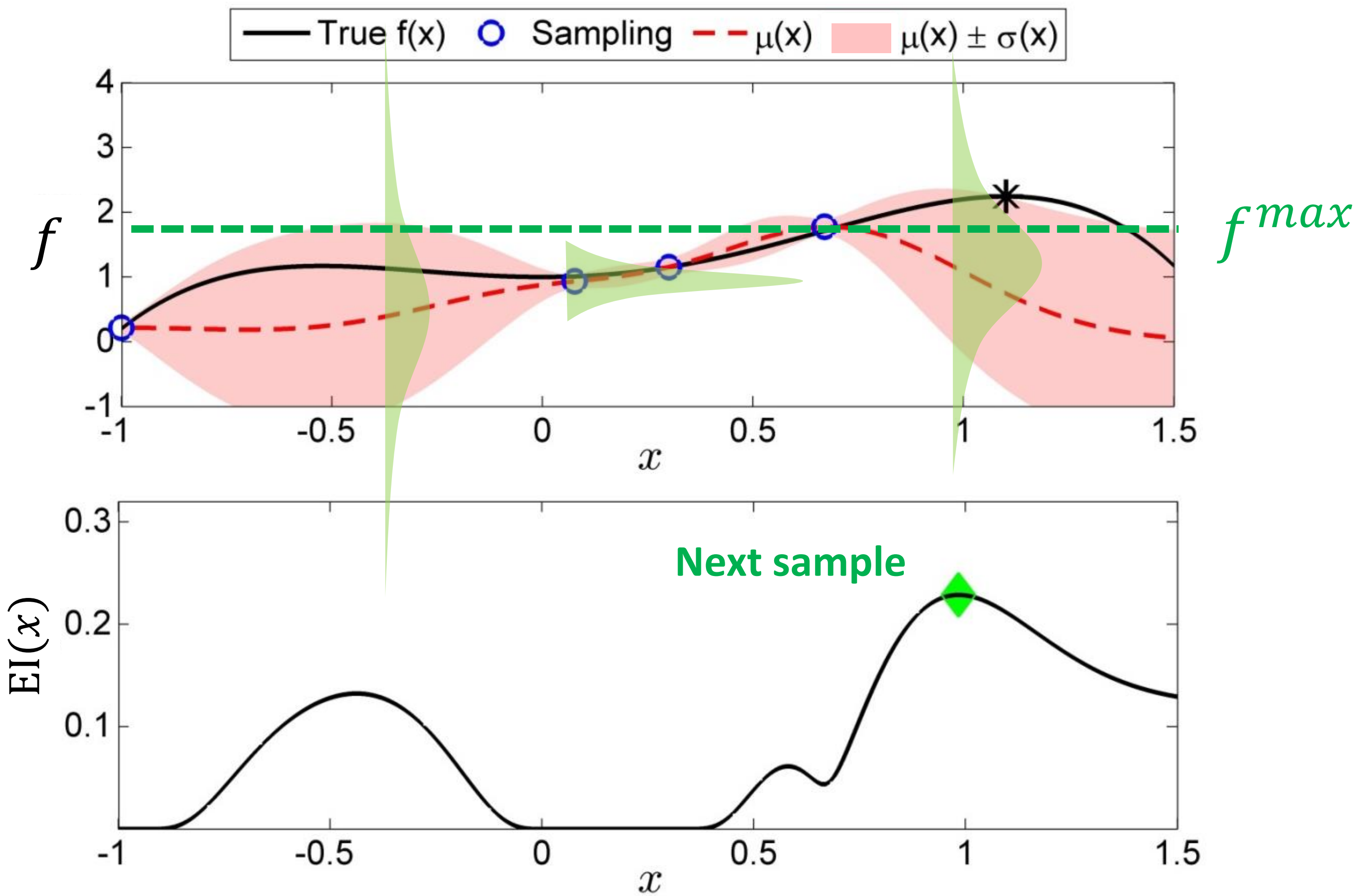
$$\boldsymbol{x}^{n+1} = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\} \,|\boldsymbol{D}^n]$$

DIRECT (Finkel, 2003), a gradient free optimization code, is used to solve this problem

Optimization (Sampling) phase

How to select the next input ?

- **Learn target function (exploration)**
- **Improve target value (exploitation)**

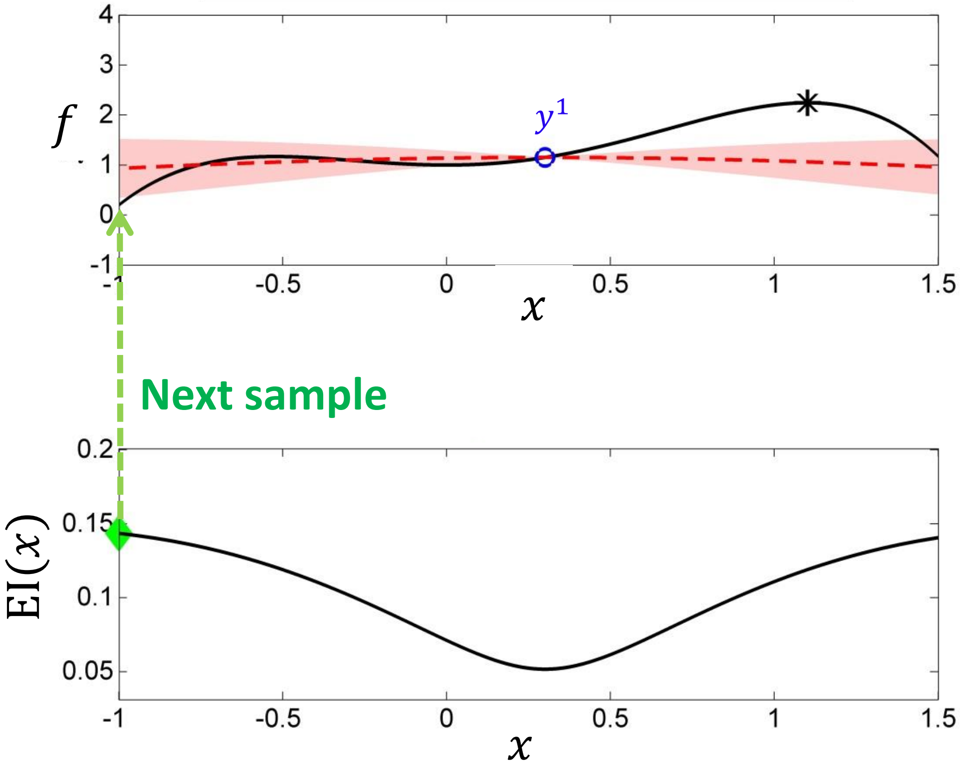


Illustrative example

$$\underset{x}{\text{maximize}}\ f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to } -1 \le x \le 1.5 \qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{matrix} y^1 \\ \\ \end{matrix}\begin{bmatrix} 1.15 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\left(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\right)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^2 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^2 = -1.00$$

Query the function value

$$y^2 = f(\boldsymbol{x}^2) + \epsilon,\ \epsilon \sim N(0, 0.01^2)$$

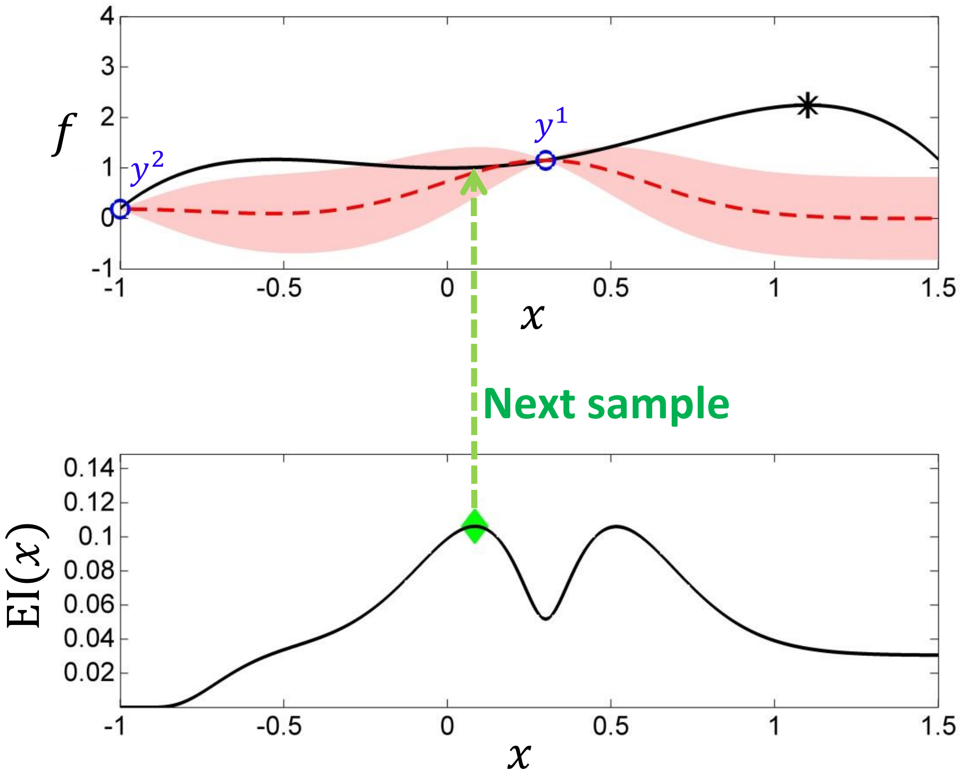
$$y^2 = 0.18$$

Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to } -1 \le x \le 1.5 \qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ \\ \end{array}\begin{bmatrix} 1.15 \\ 0.18 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^3 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^3 = 0.03$$

Query the function value

$$x^3 = f(\,\boldsymbol{x}^3) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

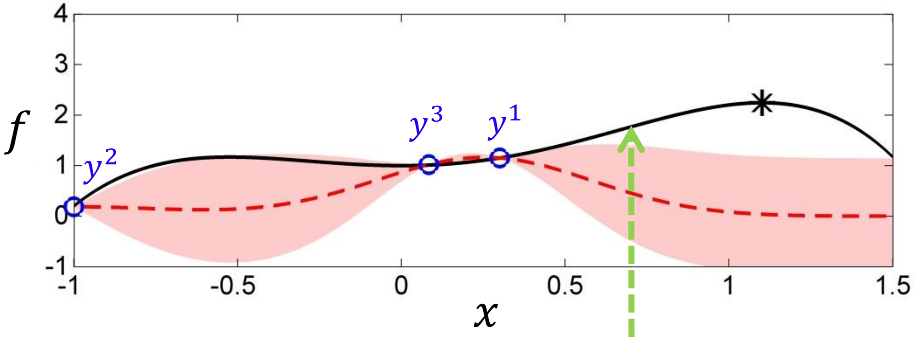
$$x^3 = 1.02$$

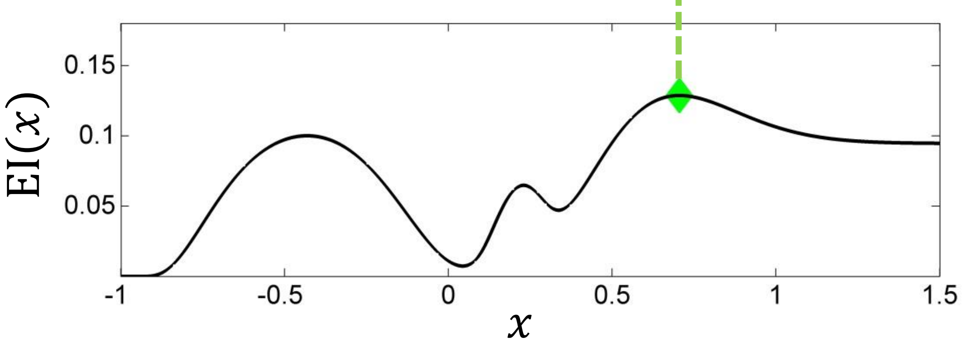
Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to } \; -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ \\ \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$





Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2\mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2\mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^4 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^4 = 0.70$$

Query the function value

$$y^4 = f(\,\boldsymbol{x}^4) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

$$y^4 = 1.76$$

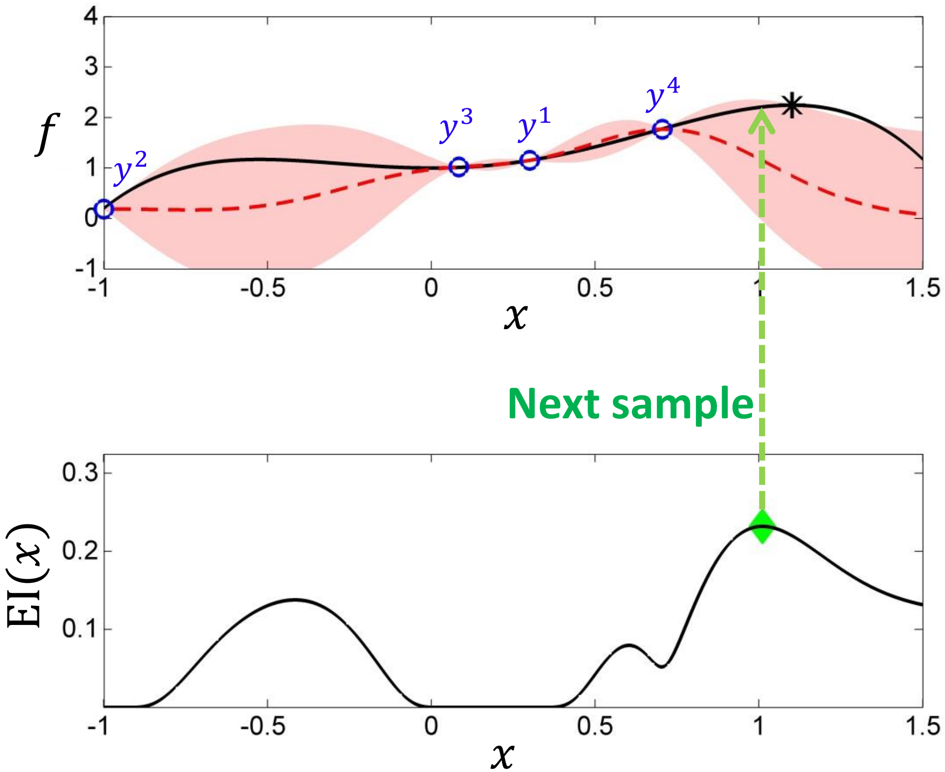
Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ f \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$

$n = 4$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^5 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^5 = 1.01$$

Query the function value

$$y^5 = f(\boldsymbol{x}^5) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

$$y^5 = 2.19$$

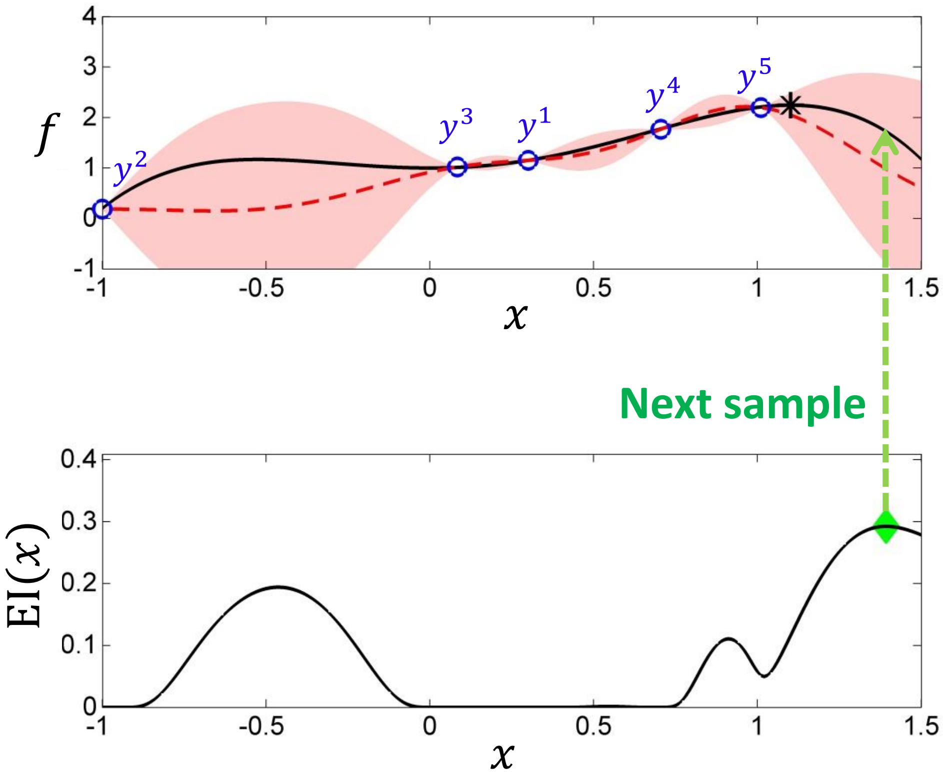
Bayesian Optimization

Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ y^5 \\ \\ \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ 2.19 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^6 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^6 = 1.39$$

Query the function value

$$y^6 = f(\,\boldsymbol{x}^6) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

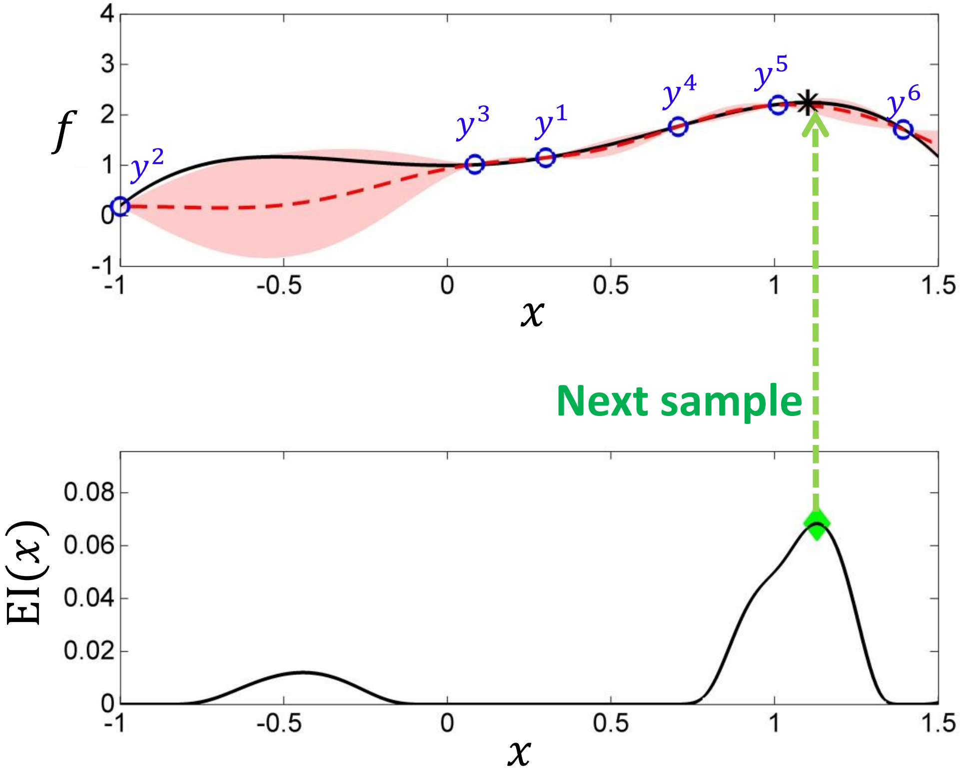
$$y^6 = 1.70$$

Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ y^5 \\ y^6 \\ \\ \end{array}\begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ 2.19 \\ 1.70 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^7 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^7 = 1.13$$

Query the function value

$$y^7 = f(\boldsymbol{x}^7) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

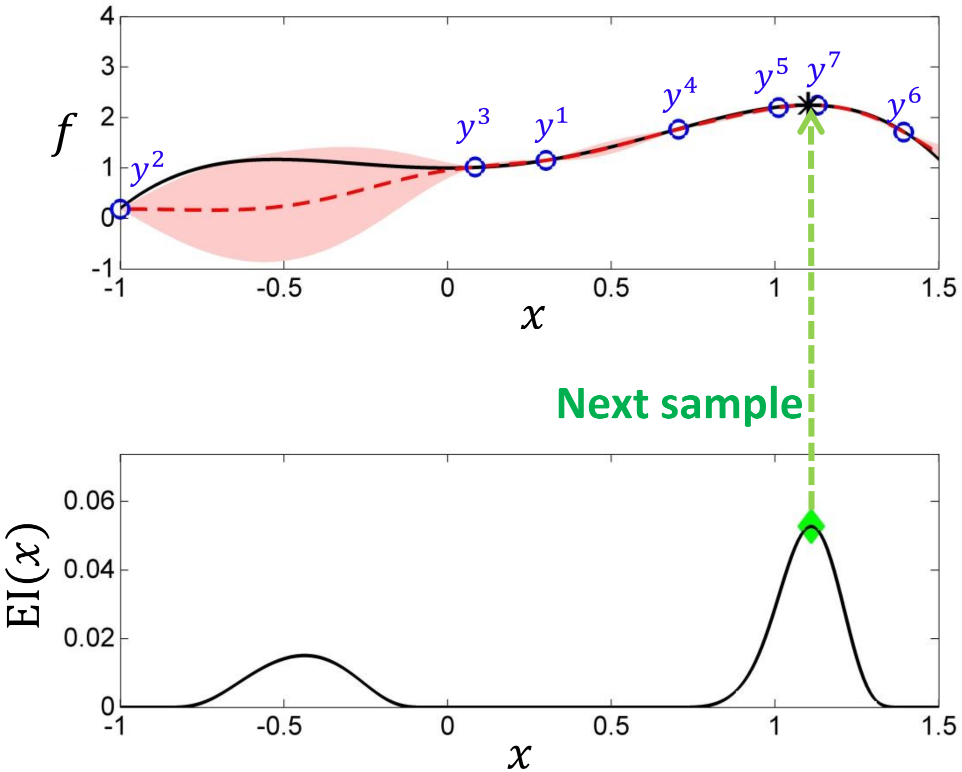
$$y^7 = 2.24$$

Illustrative example

$$\underset{x}{\text{maximize}}\, f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ y^5 \\ y^6 \\ y^7 \\ \\ \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ 2.19 \\ 1.70 \\ 2.24 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}\right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^8 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^8 = 1.11$$

Query the function value

$$y^8 = f(\boldsymbol{x}^8) + \epsilon,\ \epsilon \sim N(0, 0.01^2)$$

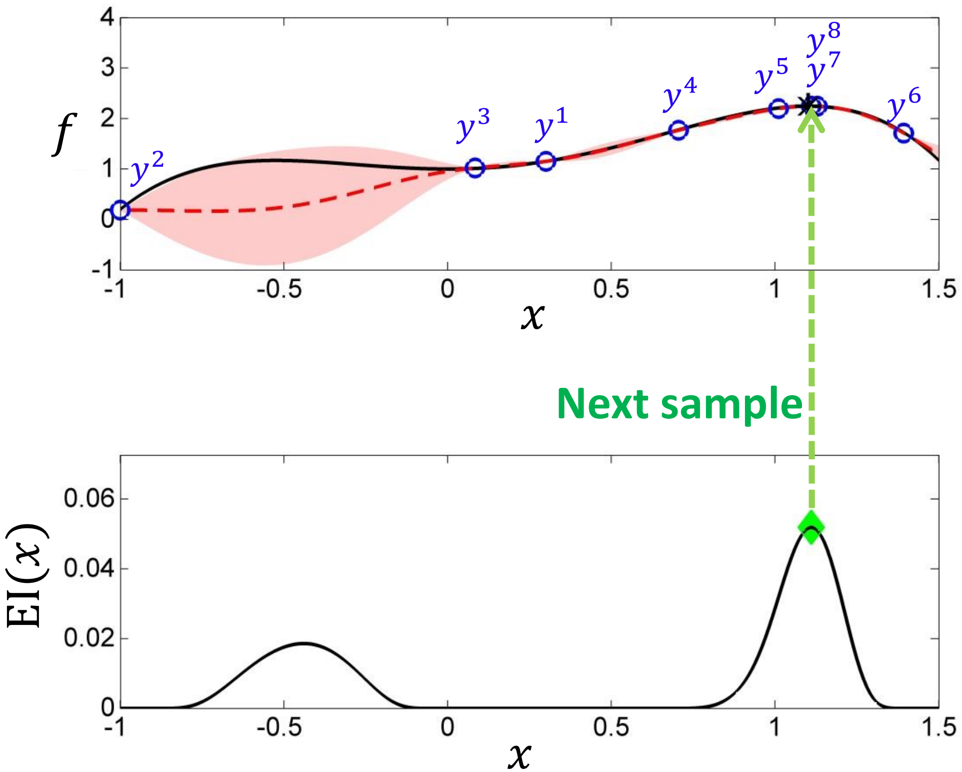
$$y^8 = 2.24$$

Illustrative example

$$\underset{x}{\text{maximize}}\; f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ y^5 \\ y^6 \\ y^7 \\ y^8 \\ f \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ 2.19 \\ 1.70 \\ 2.24 \\ 2.24 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^9 = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^9 = 1.11$$

Query the function value

$$y^9 = f(\boldsymbol{x}^9) + \epsilon,\ \epsilon \sim N(0, 0.01^2)$$

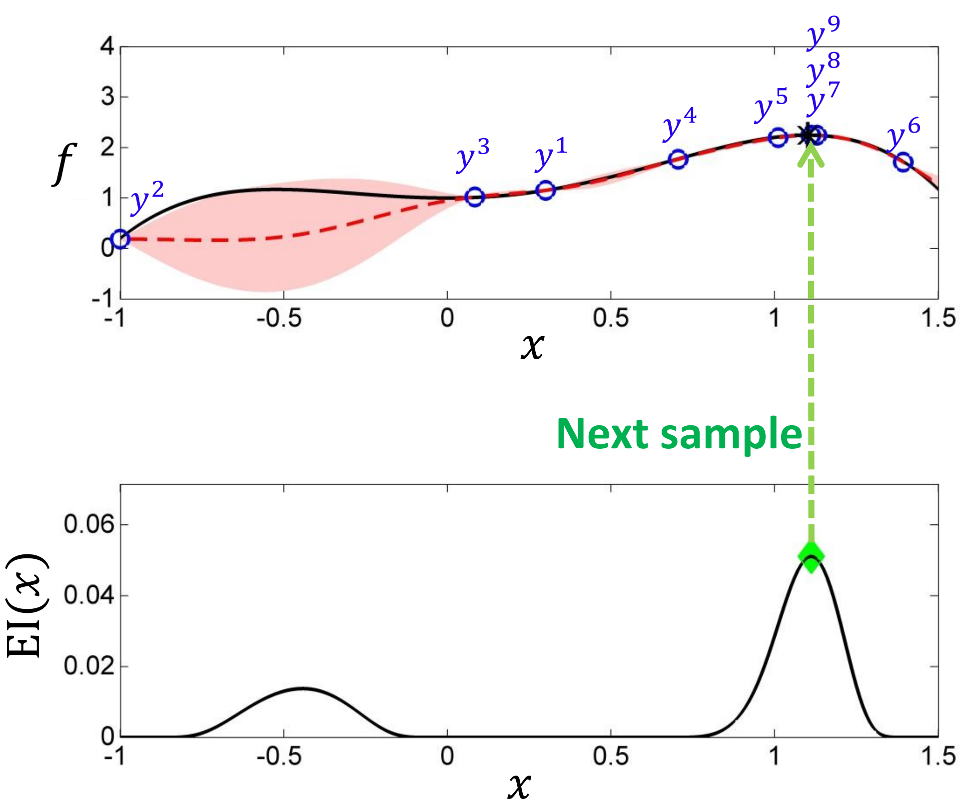
$$y^9 = 2.24$$

Illustrative example

$$\underset{x}{\text{maximize}}\ f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \leq x \leq 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1 \\ y^2 \\ y^3 \\ y^4 \\ y^5 \\ y^6 \\ y^7 \\ y^8 \\ y^9 \\ \\ \end{array} \begin{bmatrix} 1.15 \\ 0.18 \\ 1.02 \\ 1.76 \\ 2.19 \\ 1.70 \\ 2.24 \\ 2.24 \\ 2.24 \\ f \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_\epsilon^2 \mathbf{I} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^{10} = \arg\max_{\boldsymbol{z}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\} \,|\boldsymbol{D}^{1:n}]$$

$$\boldsymbol{x}^{10} = 1.11$$

Query the function value

$$y^{10} = f(\,\boldsymbol{x}^{10}) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

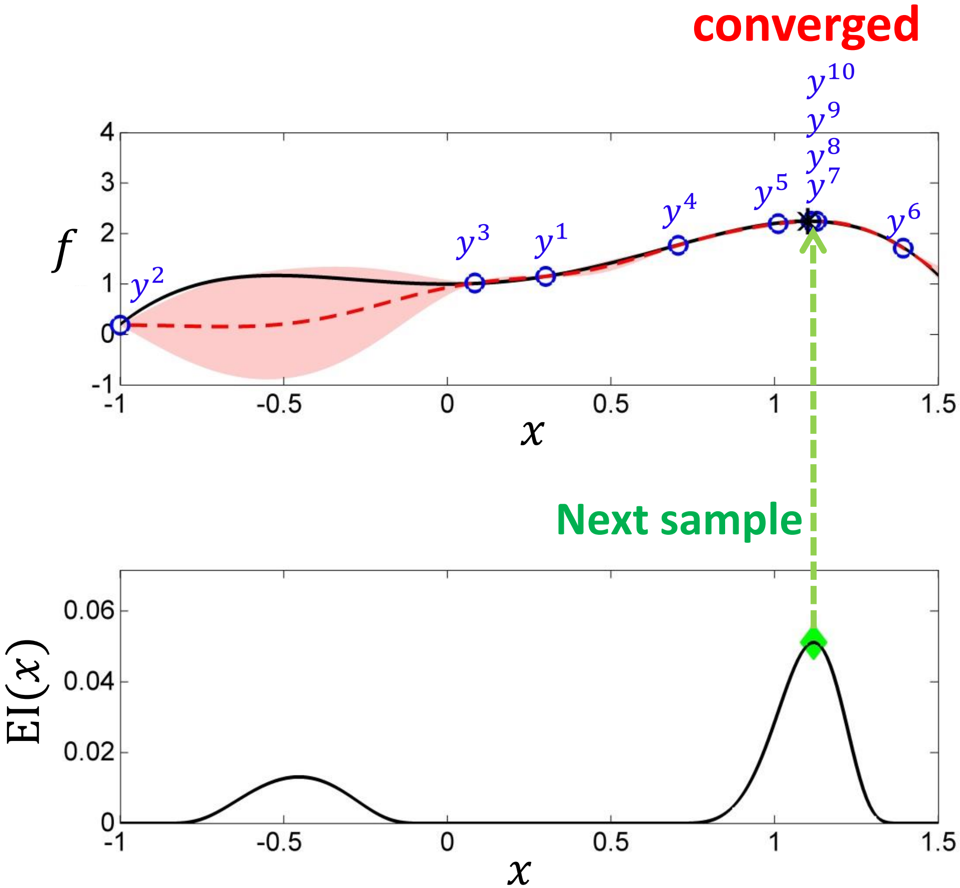
$$y^{10} = 2.24$$

Illustrative example

$$\underset{x}{\text{maximize}}\ f(x) = -1.3x^4 + 1\,x^3 + 1.5x^2 + 1 + \epsilon$$

$$\text{subject to} \quad -1 \le x \le 1.5 \qquad\qquad \epsilon \sim N(0, 0.01^2)$$

$$\begin{array}{c} y^1\\ y^2\\ y^3\\ y^4\\ y^5\\ y^6\\ y^7\\ y^8\\ y^9\\ y^{10}\\ \\ \end{array}\begin{bmatrix}1.15\\ 0.18\\ 1.02\\ 1.76\\ 2.19\\ 1.70\\ 2.24\\ 2.24\\ 2.24\\ 2.24\\ f\end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}+\sigma_\epsilon^2\mathbf{I} & \boldsymbol{k}\\ \boldsymbol{k}^T & k(\boldsymbol{x},\boldsymbol{x})\end{bmatrix}\right)$$



Construct probability distribution on the unknown function value

$$p(f|\boldsymbol{D}^{1:n}) = N\big(\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}), \sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n})\big)$$

$$\mu(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = \boldsymbol{k}^T(\mathbf{K}+\sigma_\epsilon^2\mathbf{I})^{-1}\boldsymbol{y}^{1:n}$$

$$\sigma^2(\boldsymbol{x}|\boldsymbol{D}^{1:n}) = k(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}^T(\mathbf{K}+\sigma_\epsilon^2\mathbf{I})^{-1}\boldsymbol{k}$$

Select the next trial action that maximizes

$$\boldsymbol{x}^{11} = \arg\max_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x}) \triangleq \mathrm{E}[\max\{0, f - f^{max}\}\,|\boldsymbol{D}^{1:n}]$$

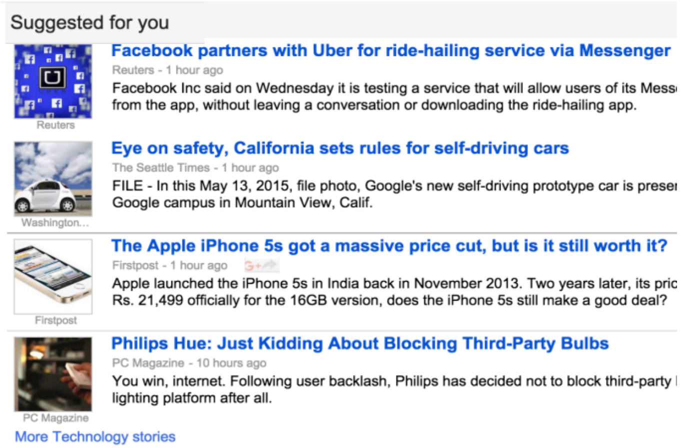
$$\boldsymbol{x}^{11} = 1.11$$

Query the function value

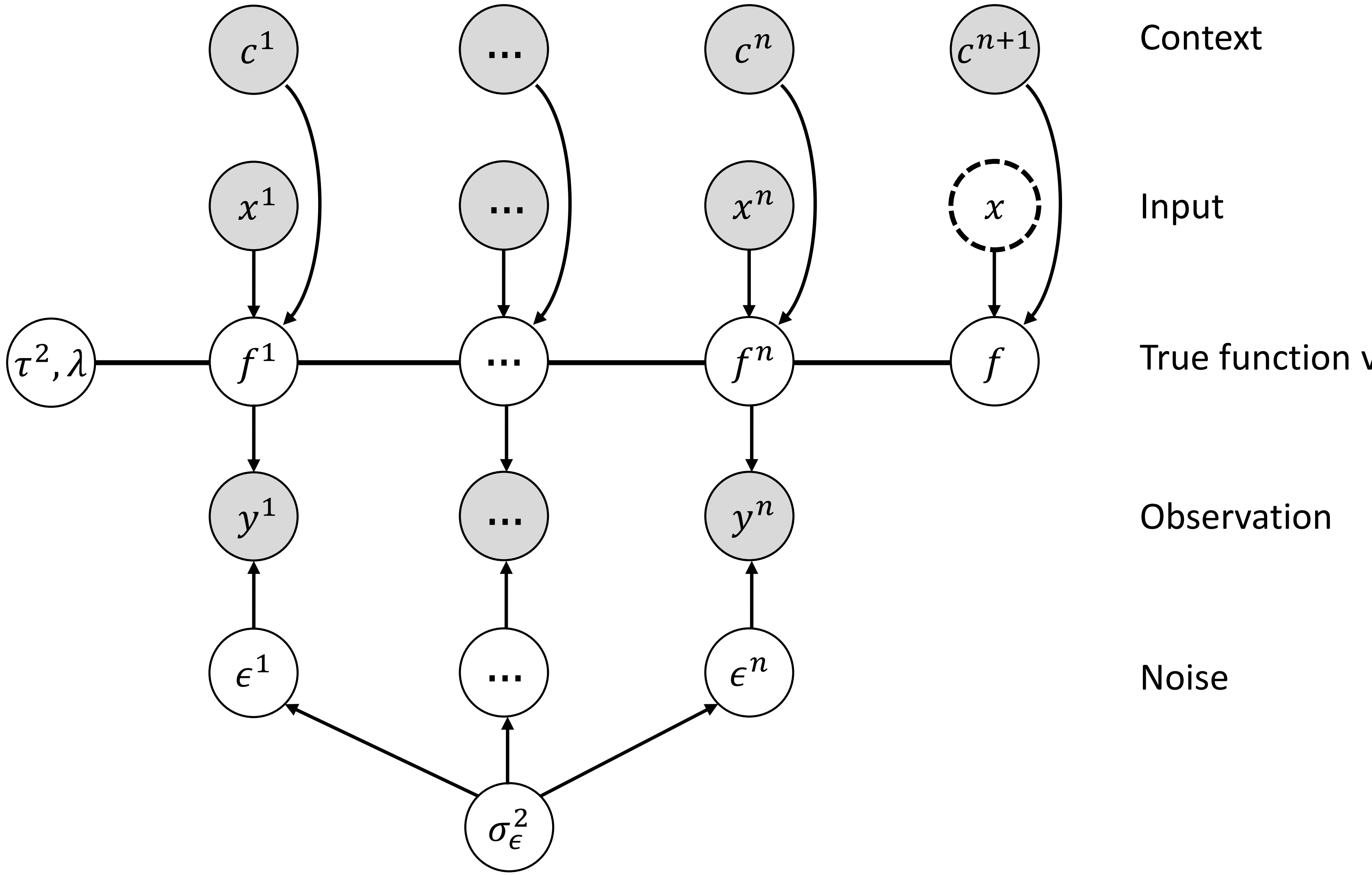
$$y^{11} = f(\boldsymbol{x}^{11}) + \epsilon, \epsilon \sim N(0, 0.01^2)$$

$$y^{11} = 2.24$$

Contextual Bandit Problem

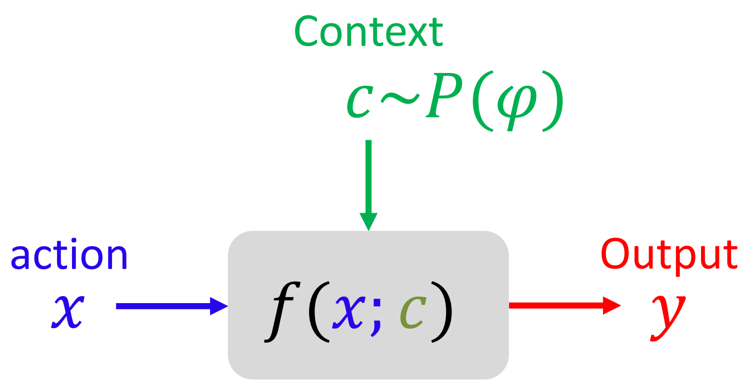


Finance : Portfolio optimization under unknown return profiles given varying economic condition
Health care : Choosing the best treatment among alternatives given different patient information
Internet shopping : Choosing the optimum price (sales v.s. profits) given varying season
Experiment design : Sequential experimental design given varying environmental condition

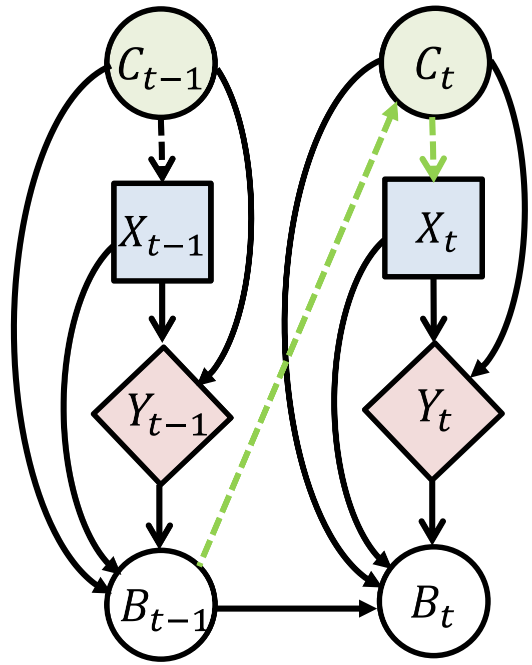


c^1
...
c^n
c^{n+1}
Context
x^1
...
x^n
x
Input
τ^2, λ
f^1
...
f^n
f
True function v
y^1
...
y^n
Observation
ϵ^1
...
ϵ^n
Noise
σ_ϵ^2

Contextual Bandit Problem



MDP over belief state



$B_t(f)$: Belief state about unknown function f at t

- Policy π maps all the history to new action:

$$\pi\colon\ [(c_1, x_1, y_1), (c_2, x_2, y_2), \dots, (c_{t-1}, x_{t-1}, y_{t-1}), c_t] \rightarrow x_t$$

- Find the optimal policy π^* that maximizes $E[\sum_{t=1}^{T} y_t]$ or $E[y_T]$

$$x^* = \pi^*(c) = \operatorname*{argmax}_x f(x; c)$$