bayes_gp_plain.py

**目的: 複数条件（記述子）と最適化する物性値 (目的関数) を
ガウス過程を用いたベイズ最適化により推定する**

(i) usage: python bayes_gp_plain.py infile max_num_probes num_rand_basis score_mode interval

max_num_probes: Use a number to reach convergence

num_rand_basis: Use a large number so as to reproduce training data

score_mode : [EI|PI|TS]

interval : # of cycles to update hyper parameters

ex: python bayes_gp_plain.py data_simple.xlsx 1 200 EI 0

Input file: Excel (.xlsx) or CSV (.csv) file

Output file: Excel (.xlsx) file

Input file: Excel (.xlsx) or CSV (.csv)

data_simple.xlsx (data_simple.csv)

target	x	y
	-1	-1
	-0.8	-1
-2.33	-0.6	-1
-2.05	-0.4	-1
	-0.2	-1
-1.73	0	-1
	0.2	-1
	0.4	-1
	0.6	-1
-2.05	0.8	-1
-2.33	1	-1

1st column: Target function
2nd and right columns: Descriptors

Target functions can leave blank data:
Non-blank data will be used for training
All data will be used to estimate prediction

IMPORTANT NOTE
If descripts or target functions vary in exponential scale, take logarithm of those values to more linearlize

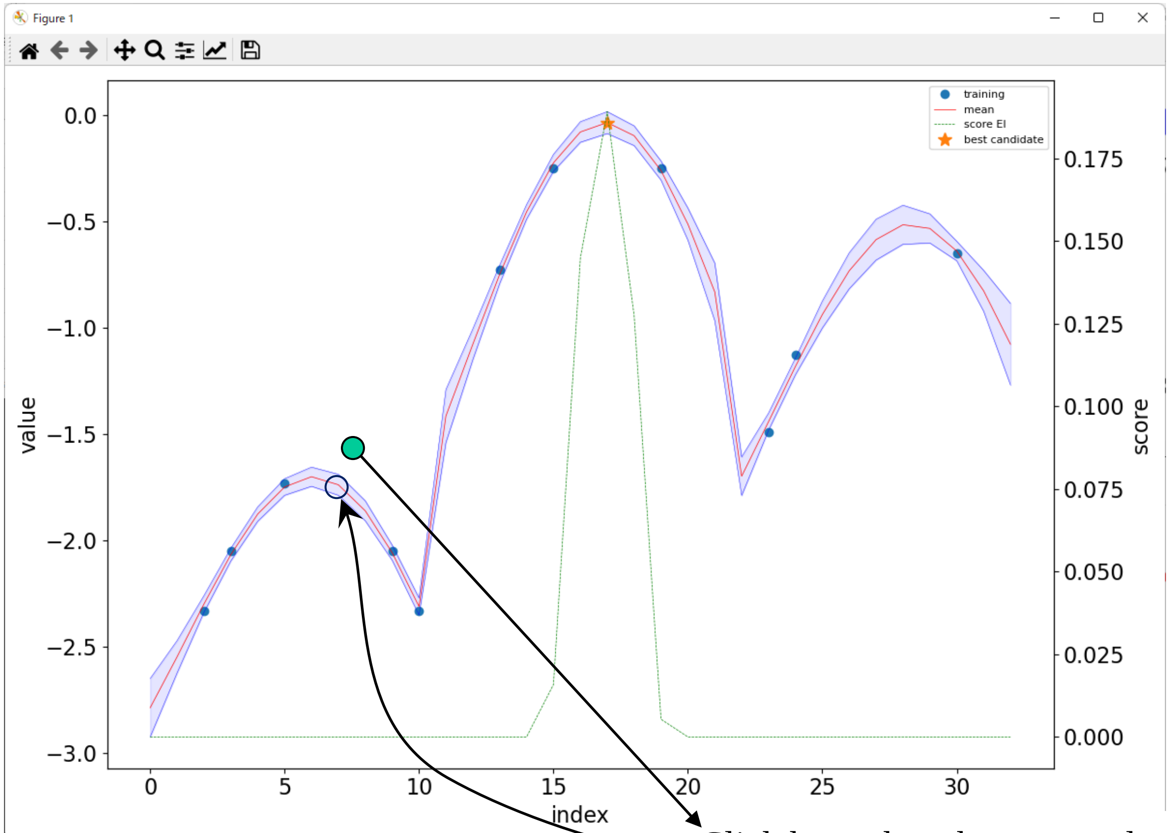
IMPORTANT NOTE
At least ONE BLANK target fuction data is required. Or the program will be terminated by ERROR in base_search()

Maximum search

python bayes_gp_plain.py

Equivalent to the default: **python bayes_gp_plain.py data_simple.xlsx 1 200 EI**

Read data_simple.xlsx, max_num_probes = 1, num_rand_basis = 200, score_mode = ‘EI’, and iterval = 0



Click here, then the nearest data information are shown in the console output:

clicked at idx = 7 / line 9: **descriptors = [0.4 -1.]** given target value = nan

predicted target value = -1.7141204429199943 +- 0.05214082204618646

Input file for minimum search

data_target_min.xlsx

min:target	x	y	-z
	-1	-1	1
	-0.8	-1	1
-2.33	-0.6	-1	1
-2.05	-0.4	-1	1
	-0.2	-1	1
-1.73	0	-1	1
	0.2	-1	1
	0.4	-1	1
	0.6	-1	1
-2.05	0.8	-1	1

Header labels can have the following control tags at the top (case insensitive).

'max:' (default): Find maximum score
'min:': Find minimum score
Target function is converted to –t
(t denotes the target function)
'=value': Find closest data to 'value'
(*value* is given by a floating point number)
Target function is converted to $-(t - value)^2$
Note: $(t - value)^2$ does not follow the Gauss process.
This option loses theoretical basis
'-': Ignore this column
neither as target function nor descriptors

1st column: Target function, converted to –*t*
2nd and 3rd columns: Descriptors
4th column: with '-' so is ignored

IMPORTANT NOTE
At least ONE BLANK target fuction data is required. Or the program will be terminated by ERROR in base_search()

Input file for value search

data_target_value.xlsx

=-0.5:target	x	y	-z
	-1	-1	1
	-0.8	-1	1
-2.33	-0.6	-1	1
-2.05	-0.4	-1	1
	-0.2	-1	1
-1.73	0	-1	1
	0.2	-1	1
	0.4	-1	1
	0.6	-1	1
-2.05	0.8	-1	1

1st column: Target function, converted to $-(t - (-0.5))^2$
2nd and 3rd columns: Descriptors
4th column: with '-' so is ignored

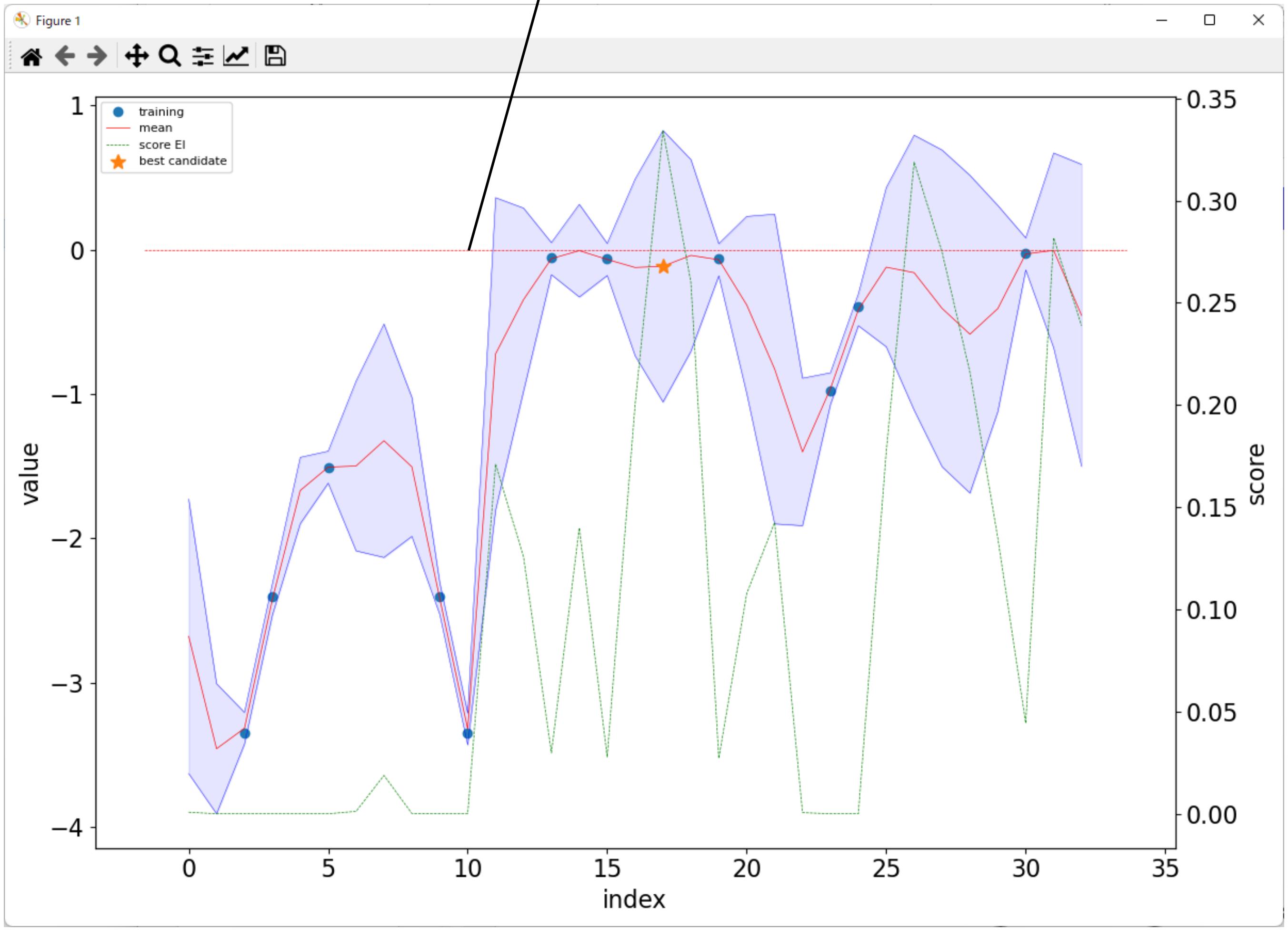
IMPORTANT NOTE
At least ONE BLANK target fuction data is required. Or the program will be terminated by ERROR in base_search()

Value search

python bayes_gp_plain.py data_target_value.xlsx

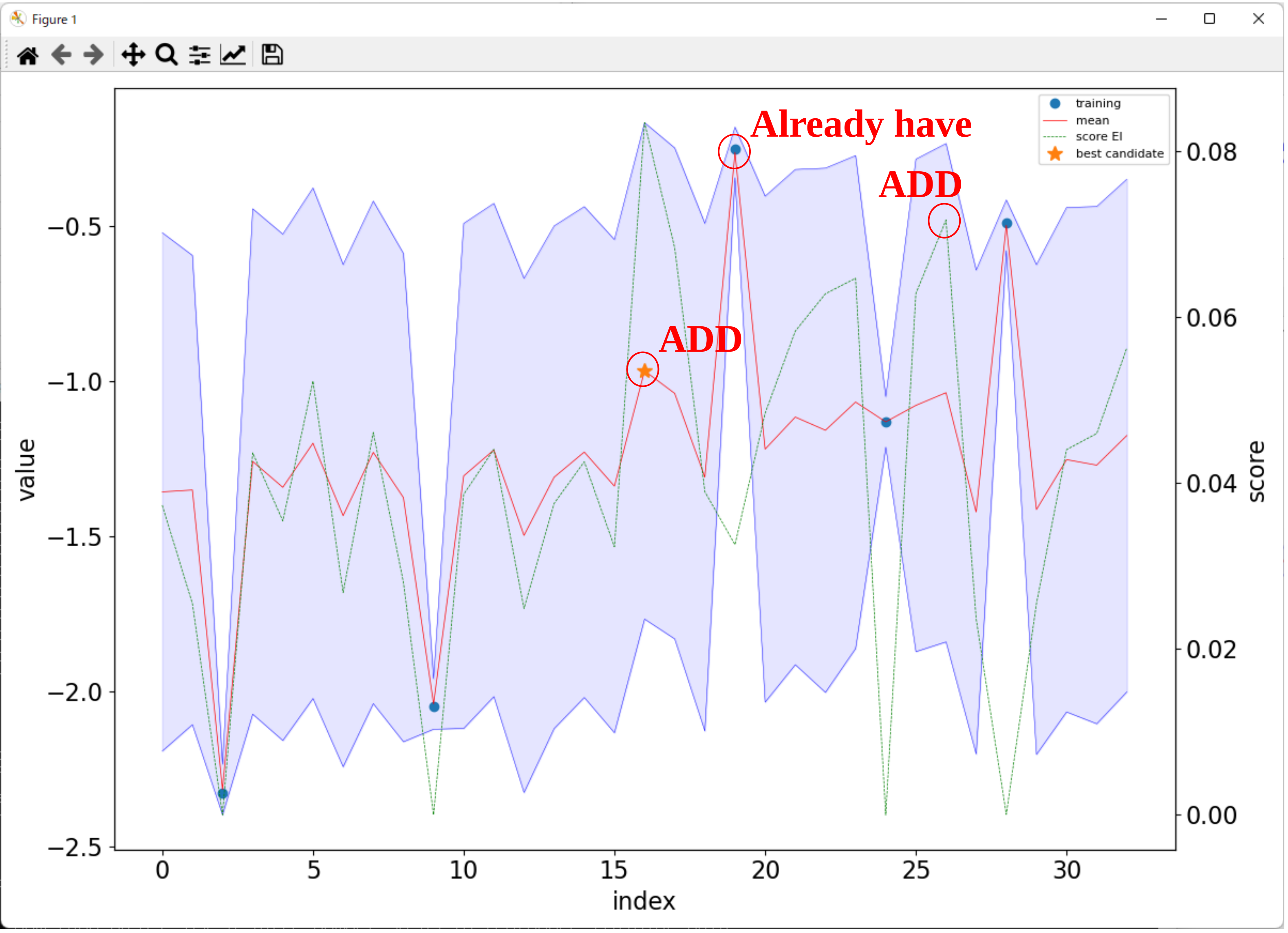
Zero line corresponds to the target value of the target function

Note: The values cannot be positive as they are $(t - value)^2$
so the positive prediction values and the symmetric standard deviations
have no theoretical validness



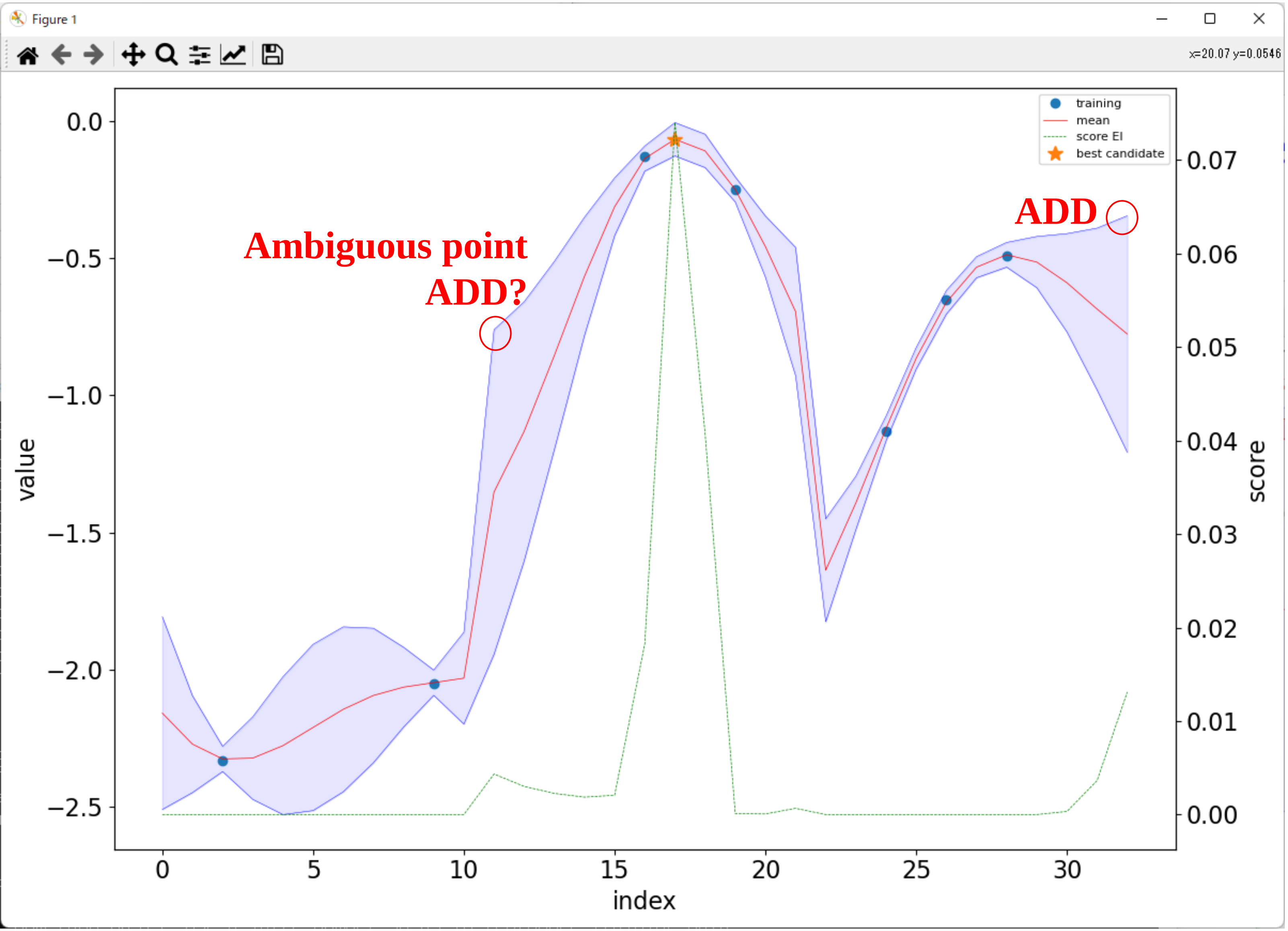
Variation of predictions

python bayes_gp_plain.py data_test.xlsx



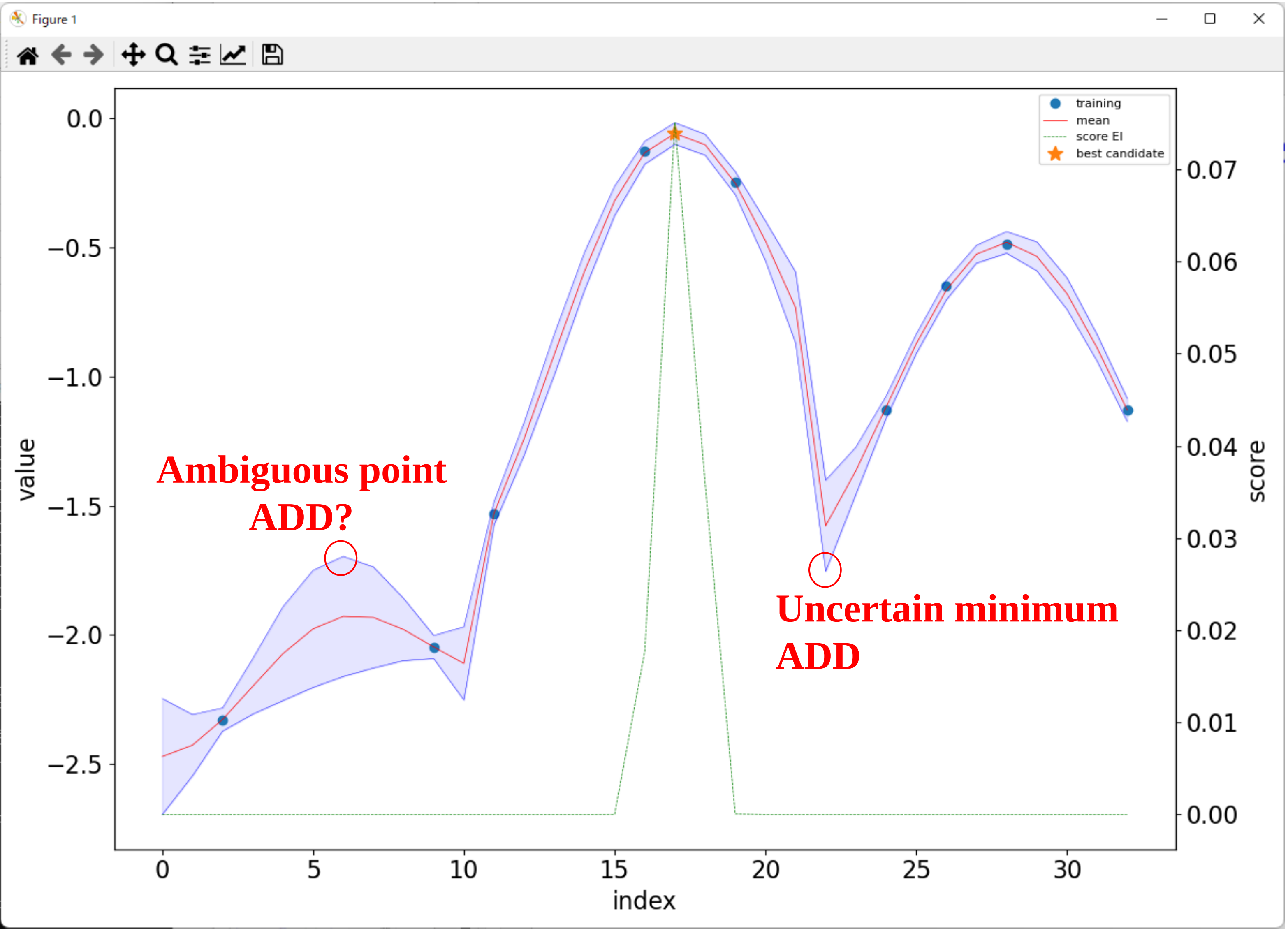
Variation of predictions

python bayes_gp_plain.py data_test.xlsx



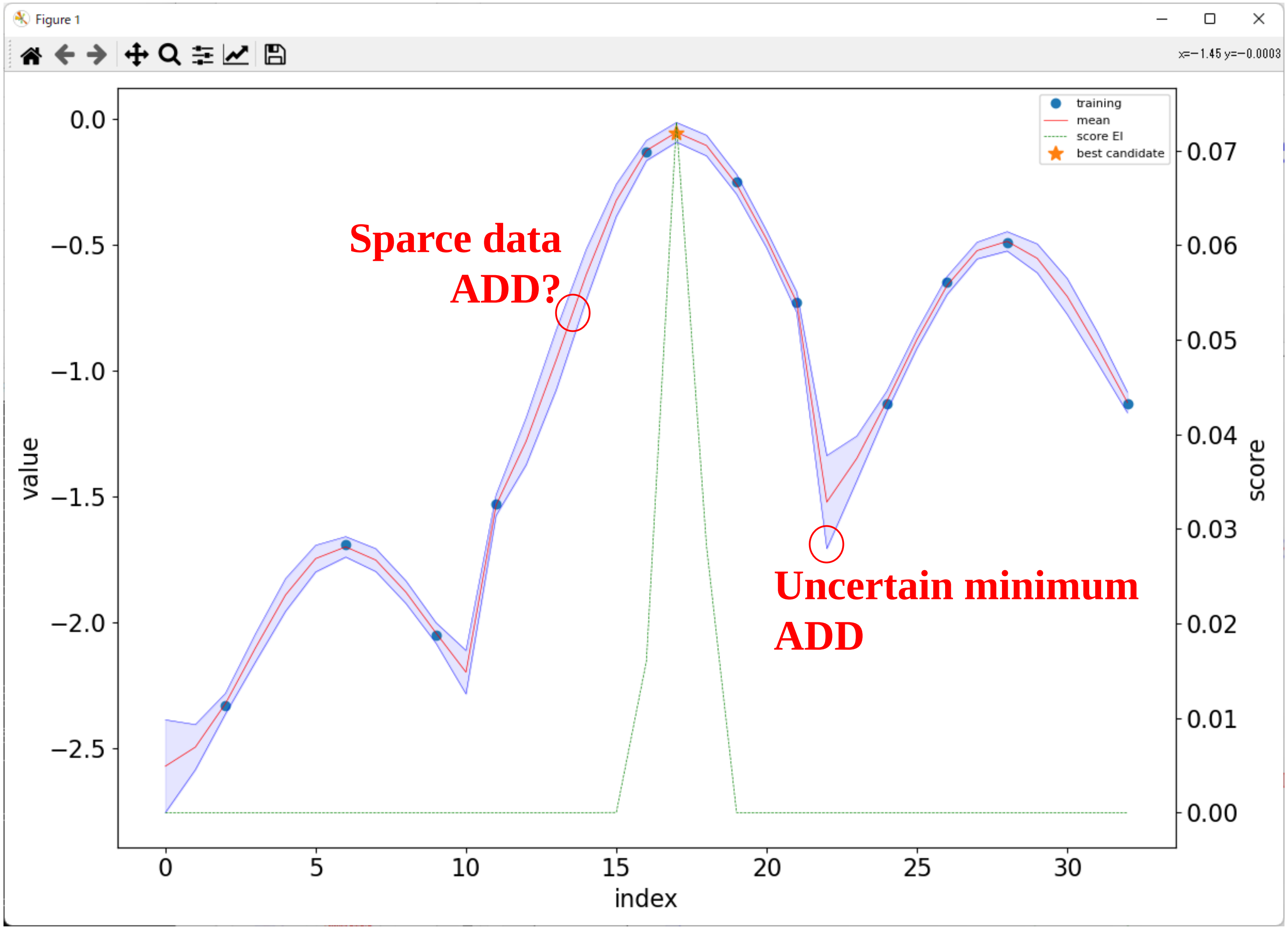
Variation of predictions

python bayes_gp_plain.py data_test.xlsx



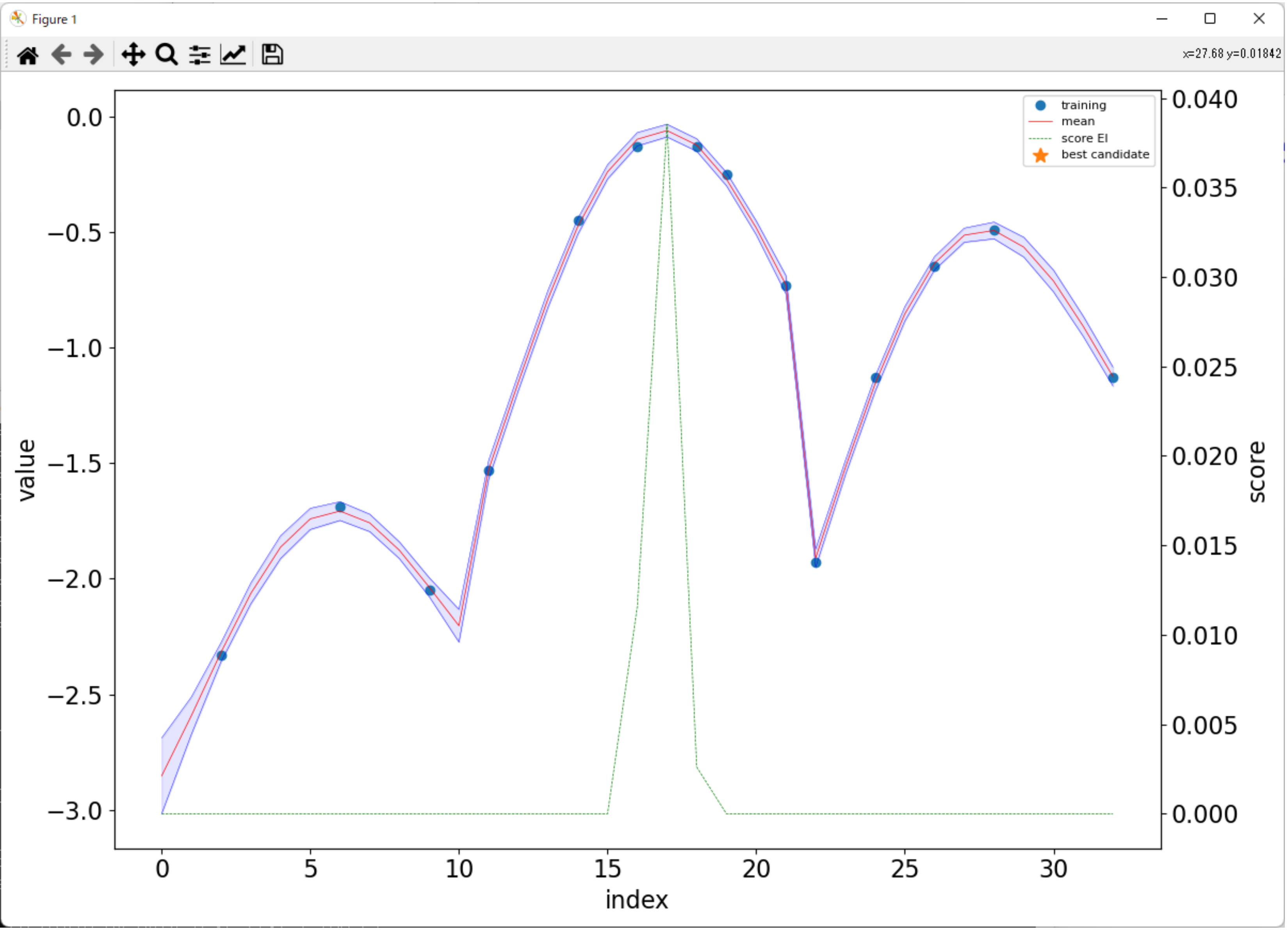
Variation of predictions

python bayes_gp_plain.py data_test.xlsx



Variation of predictions

python bayes_gp_plain.py data_test.xlsx



Theory

ベイズ推定・ベイズ最適化

強化学習の１つとして使われる

ガウス過程にベイズ推定を取り込み、最適化に利用

・観測データ $\{X'_i, y_i\}$、誤差 ε_i が正規分布に従うとする

・予測値をガウス基底で展開

$$t_i = \sum_k w_k \exp\left(-\sum_{k',k''} A_{k,k',k''}\left(x_{i,k'} - \mu_{k,k'}\right)\left(x_{i,k''} - \mu_{k,k''}\right)\right)$$

$A_{k,k,k'} = \frac{1}{2\sigma_{k,k',k''}{}^2}$　　$\mu_{k,k'}$, $\sigma_{k,k',k''}{}^2$: k番目の基底の平均と共分散行列

・w_kが正規分布に従うと仮定

・ベイズ推定により、w_kの平均、分散が求まる

＝＞ 予測値 t_i も正規分布に従う

・ガウス過程であれば、カーネルトリックを利用して

t_i の平均、分散を（比較的）簡単に求めることができる

ガウス過程 (Gauss process)

観測データ $\{X'_i, y_i\}$ をガウス基底で展開

観測データに誤差 ε_i があることを考慮すると、

$$t_i = \sum_k w_k \phi_k(X'_i)$$

$$y_i = t_i + \varepsilon_i$$

とあらわされる。一連の関数 $\phi_k(X'_i)$、ε_i、y_i 、w_k、t_i が
正規分布に従うとき、これをガウス過程と呼ぶ

Wikipedia:

確率過程 $\{X_t\}_{t\in T}$ は、任意に(有限個の) $X_{t1}, \ldots, X_{tk}$ を選んで作った線型結合(あるいはより一般に、$\{X_t\}_{t\in T}$ を標本関数 X_t 全体からなる連続濃度の函数空間と見たときの、任意の線型汎関数)が正規分布に従うとき、ガウス過程という。

統計学

頻度主義 (frequentism) 統計学

確率が結果 (データ) を決める

ベイズ主義 (Bayesianism) 統計学

データは定数、確率分布 (確率分布関数に含まれる定数) が分布をもつ

客観確率 (objective probability)

実験または理論的考察（思考実験）から求められ、客観的な観測結果と比較できるランダムな事象についての確率

主観確率 (subjective probability)

人間の主観的な信念あるいは信頼の度合

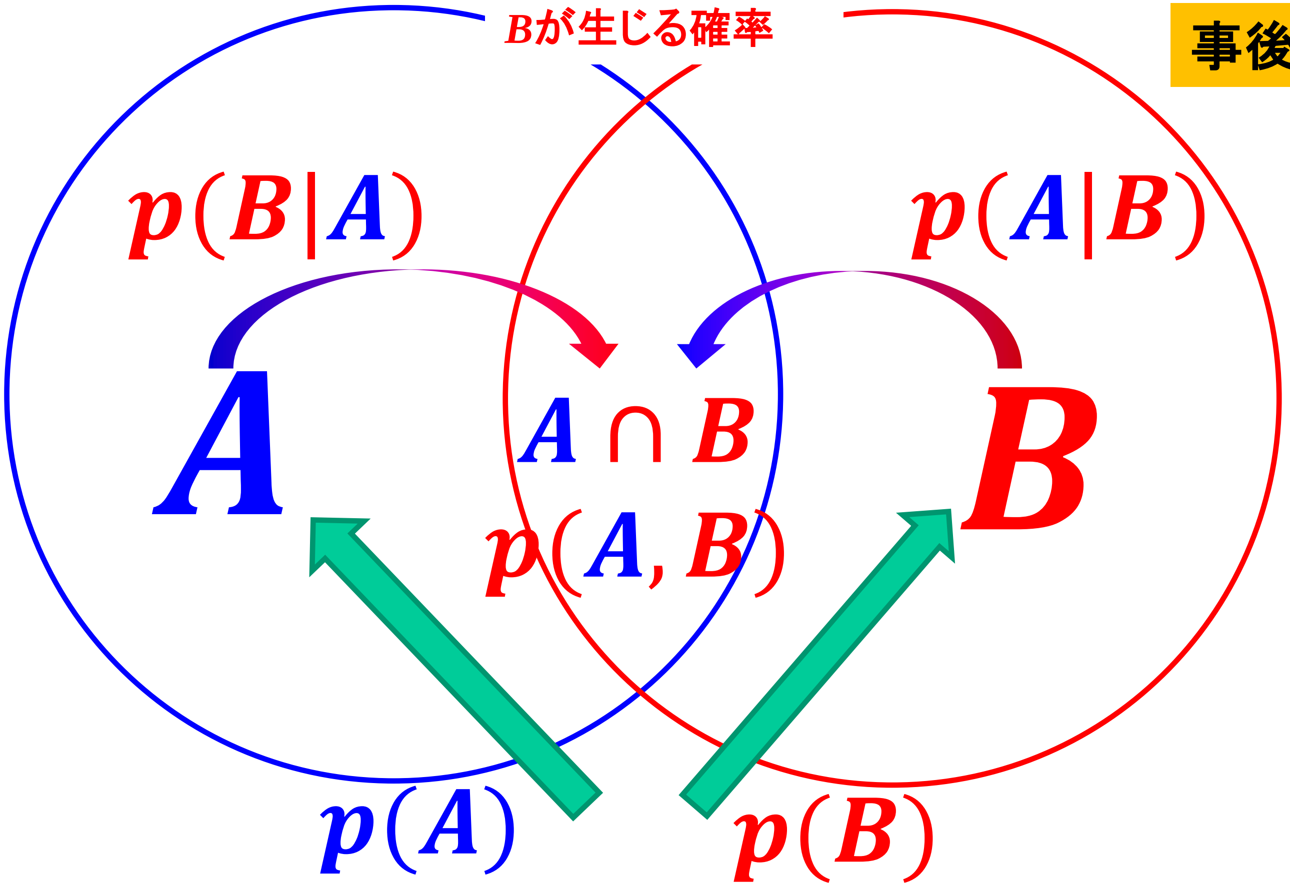
確率に関するベイズの定理

$$p(A,B) \quad = \quad p(B|A)p(A) \quad = \quad p(A|B)p(B)$$

同時確率 事後確率 事前確率

AとBが同時に生じる確率 Aが生じた場合にBが生じる確率 Aが生じる確率

事後確率 ＝ 条件付き確率



ベイズの定理: $p(A|B) = p(B|A)p(A)/p(B)$

事後確率 事前確率

ベイズの定理と材料研究

$$p(A,B) \quad = \quad p(B|A)p(A) \quad = \quad p(A|B)p(B)$$

AとBが同時に生じる　Aが生じた場合にBが生じる確率　Aが生じる確率

ベイズの定理: $p(A|B) = p(B|A)p(A)/p(B)$

B が生じた場合にAが生じる確率　事前確率

A: 実験条件　B: 測定結果　と考えると、

$p(B|A)$: A の条件で実験を行った結果、B が得られる確率

ベイズの定理により、

B が得られる確率の高い条件 A を $p(A|B)$ から推定できる

問題: $p(B) = \sum_A p(B|A)$ を決めるためにはすべての実験をやらないといけない

↑今日はこの問題は忘れて説明を続ける

主観確率 (subjective probability) の問題

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

殺人事件が起こり、容疑者 X を捕まえた。Xの血液鑑定をしたところ、血液の特徴が一致した。
Xが犯人である確率 [事後確率 $p(\text{犯人}|\text{一致})$] はいくらか。

条件付確率: 偶然血液の特徴が一致する確率は10^{-5}

$p(\text{一致}|\text{犯人でない}) = 10^{-5}$　　　$p(\text{一致}|\text{犯人}) = 1$

(0) 単純に考えて、$p(\text{犯人}) = 1 - p(\text{一致}|\text{犯人でない}) = 0.99999$ としていいか？
日本人口 12,000万人のうち、血液型が一致するのは1,200人もいる　=> それじゃ、約1/1200？
世界人口 700,000万人のうち、血液型が一致するのは70,000人もいる　=> それじゃ、約1/70000？

事後確率 $p(\text{犯人}|\text{一致}) = p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{\sum_A p(B|A)p(A)}$

$$= \frac{p(\text{一致}|\text{犯人})p(\text{犯人})}{p(\text{一致}|\text{犯人})p(\text{犯人})+p(\text{一致}|\text{犯人でない})\{1-p(\text{犯人})\}} = \frac{p(\text{犯人})}{p(\text{犯人})+10^{-5}\{1-p(\text{犯人})\}}$$

事前確率の仮定:

(1) 理由不十分の原則　: $p(\text{犯人}) = 1/2$

$p(B) = p(\text{一致}) = 0.50005$ => $p(\text{犯人}|\text{一致}) = 0.99999$: 犯人確定

(2) 一人当たりの殺人率: $p(\text{犯人}) = 10^{-5}$

$p(B) = p(\text{一致}) = 1.99999 \times 10^{-5}$ => $p(\text{犯人}|\text{一致}) = 0.5$

事前確率で「犯人をみつける確率は非常に低い」という思い込み (主観) があるため、
事後確率が下がる

(3) 犯人を含んでいるとみられる地域には3700万人の人が住んでいる: $p(\text{犯人}) = 1/37000000$

$p(B) = p(\text{一致}) = 1.003 \times 10^{-5}$ => $p(\text{犯人}|\text{一致}) = 0.0027$

3700万人の対象者中、血液の特徴が一致する人は370人いる。$p(B) \sim 1/370$

ベイズ更新 (Bayesian updating)

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

メール A が迷惑メール A_1 である確率を求める。メールの特徴をB としたとき:

迷惑メールA_1である確率 $p(A_1|B)$、迷惑メールでないA_2である確率: $p(A_2|B)$

=> $p(A|B)$ が得られているとする。

ここに、ほかの特徴 C による判定を加える。

$$p(A,B,C) = p(A|B,C)p(B,C) = p(B,C|A)p(A)$$

=> $p(A|B,C) = \frac{p(B,C|A)p(A)}{p(B,C)}$

A, Bが独立であれば$p(B,C) = p(B)p(C)$、

Aが与えられた条件で Bと Cが独立であれば $p(B,C|A) = p(B|A)p(C|A)$

=> $p(A|B,C) = \frac{p(B|A)p(C|A)p(A)}{p(B)p(C)} = \frac{p(C|A)}{p(C)}\frac{p(B|A)p(A)}{p(B)}$

$p(A,B) = p(A|B)p(B) = p(B|A)p(A)$ から、

$$p(A|B,C) = \frac{p(C|A)}{p(C)}p(A|B)$$

※ B の情報から A が迷惑メールであるかどうかの確率 $p(A|B)$ がわかっている場合に、

C による情報 $\frac{p(C|A)}{p(C)}$ を加えて、Aが迷惑メールである確率を更新 (改善) できる

確率に関するベイズの定理

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$

A, B: 事象 (event)

$p(A)$: 事象Aが観察される確率

$p(A, B)$: 同時確率 (joint probability)。事象AとBが同時に観察される確率

$\sum_{i,j} p(A_i, B_j) = 1$

$\sum_i \ p(A_i, B_j) = p(B_j)$: 周辺確率 (marginal probability)

$p(B|A)$: 条件付き確率 (conditional probability)。

事象Aが観察されたという条件の下で、Bが観察される確率

$p(B_j|A_i) = \frac{p(A_i, B_j)}{p(A_i)}$

$\sum_j \ p(B_j|A_i) = 1$

乗法の定理 (multiplication theorem of probability): $p(A_i, B_j) = p(B_j|A_i)\, p(A_i) = p(A_i|A_j)\, p(B_j)$

全確率の公式 (law of total probability): $p(B_j) = \sum_i \ p(A_i, B_j)p(A_i)$

ベイズの定理 (Bayes' theorem): $p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{p(B_j)} = \frac{p(B_j|A_i)p(A_i)}{\sum_i \ p(B_j|A_i)p(A_i)}$

$p(A_i)$　: 事前確率 (prior probability)

$p(A_i|B_j)$: 事後確率 (posterior probability)

$p(A_i|B_j) = p(B_j|A_i)$: AとBは互いに独立である(independent) => $p(A_i, B_j) = p(A_i)p(B_j)$

確率分布に関するベイズの定理

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

母数 θ を確率変数として扱う

$f(\theta)$: 確率変数の確率分布。事前確率分布 (prior probability distribution)

$f(\theta, x) = f(x|\theta)f(\theta) = f(\theta|x)f(x)$: 確率変数が$\theta$をとり、データが x をとる同時確率

$f(\theta|x)$: データによる母数の条件付き分布。事後確率分布 (posterior probability distribution)

データが確定している場合のθの条件付き確率を

θの関数として尤度関数 (likelihood function) と呼ぶ

$f(x|\theta)$: 確率変数がθをとるとき、データが x を取る条件付き確率。

一般に 確率密度分布関数 と呼ばれる

$$f(x) = \int f(x|\theta)\, f(\theta) d\theta$$

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}$$: 確率分布に関するベイズの定理

$\Rightarrow$ 観測データ $\{x_i, y_i\}$ があるとき、分布関数の母数θ の分布がわかる

例えばフィッティング変数 w_i が正規分布

$$f(w_i|x_i, y_i) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

に従う場合、w_i の期待値 μ と標準偏差σ が求まる

線形回帰

線形回帰:

観測データ $\{X_i, y_i\}$ を $y = a_0 + \sum_i a_i x_i$ でフィッティングし、
係数 a_i (i = 0, 1, …, N) を求めること

$X_i = \{x_{i,j}\}$: 記述子 (descriptor) の組　　実験条件、計算条件に対応
y_i: 目的関数　　実測値、計算値に対応

記述子 $x_{i,j}$、目的関数 y_i は平均 $\mu_{i,j}$ と標準偏差 $\sigma_{i,j}$ を使って
標準化する: 平均 0、標準偏差 1になる

$$x'_{i,j} = \frac{x_{i,j} - \mu_{i,j}}{\sigma_{i,j}} \quad => y' = \sum_i a_i X'_i$$ でフィッティングすればよい

単回帰: 記述子が1種類だけ
重回帰: 記述子が複数

最尤推定法

誤差 $\varepsilon_{\mathrm{i}} = f(x_{\mathrm{i}}, a_{\mathrm{i}}) - y_{\mathrm{i}}$ が分散 σ_{i} の正規分布に従うとする。データ$(x_{\mathrm{i}}, y_{\mathrm{i}})$に対するパラメータ (a_{i}) の尤度関数は

$$P(a_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{{\varepsilon_i}^2}{2{\sigma_i}^2}\right]$$

尤度を最大化するパラメータを求めるのが「最尤推定法」。

$$max\, P(a_i) = max \ln P(a_i) = min \sum_i \frac{{\varepsilon_i}^2}{{\sigma_i}^2}$$: 最小二乗法に一致する

線形最小二乗法

誤差の自乗和 (L2 norm) $S = \Sigma(f(x_i) - y_i)^2$ **を最小化**

$$f(x_i) = a + bx_i$$

$$S = \Sigma(a + bx_i - y_i)^2$$

$$dS/da = 2\Sigma(a + bx_i - y_i) = 2an + 2b\Sigma x_i - 2\Sigma y_i = 0$$

$$dS/db = 2\Sigma x_i(a + bx_i - y_i) = 2a\Sigma x_i + 2b\Sigma x_i^2 - 2\Sigma x_i y_i = 0$$

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

線形回帰を非線形問題へ

重回帰: 観測データ $\{X_i, y_i\}$ を $\mathrm{y} = \sum_i a_i X'_i$ でフィッティング

＝＞ $\mathrm{y} = \sum_k a_k f_k(X'_i)$ でフィッティングしても、

未知変数 a_i に関しては線型方程式のまま

=> 線形回帰の方法がそのまま使える

例: $f_k(x) = x^k$: 多項式回帰

$f_k(x) = \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$: ガウス基底

線形最小二乗法: 一般関数の場合: 行列表示

$$f(x) = \sum_{k=1}^{n} a_k f_k(x) \qquad S = \sum_{i=1}^{N} \left(y_i - \sum_{k=1}^{n} a_k f_k(x_i) \right)^2$$

$$\frac{dS}{da_l} = -2 \sum_{i=1}^{N} f_l(x_i) \left(y_i - \sum_{k=1}^{n} a_k f_k(x_i) \right) = 0$$

$$\begin{pmatrix} \sum f_1(x_i) f_1(x_i) & \sum f_1(x_i) f_2(x_i) & \sum f_1(x_i) f_3(x_i) & \cdots & \sum f_1(x_i) f_N(x_i) \\ \sum f_2(x_i) f_1(x_i) & \sum f_2(x_i) f_2(x_i) & \sum f_2(x_i) f_3(x_i) & & \sum f_2(x_i) f_N(x_i) \\ \sum f_3(x_i) f_1(x_i) & \sum f_3(x_i) f_2(x_i) & \sum f_3(x_i) f_3(x_i) & & \sum f_3(x_i) f_N(x_i) \\ \vdots & & & \ddots & \\ \sum f_N(x_i) f_1(x_i) & \sum f_N(x_i) f_2(x_i) & \sum f_N(x_i) f_3(x_i) & & \sum f_N(x_i) f_N(x_i) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \sum y_i f_1(x_i) \\ \sum y_i f_2(x_i) \\ \sum y_i f_3(x_i) \\ \vdots \\ \sum y_i f_N(x_i) \end{pmatrix}$$

係数に関して線形であれば、1度の行列計算で最終解が得られる

例: $f(x) = a + b \log x + c/x$

$f(x, y) = a + bxy + cy/x$