



# Inverse Estimation of trained Driverless AI model

- Driverless AIモデルの逆問題解析 -

H2O.ai

# “Inverse Estimation of trained Driverless AI model” App

H<sub>2</sub>O.ai

- 予測モデル( $y = f(X)$ )の作成
- スコアリングの実施

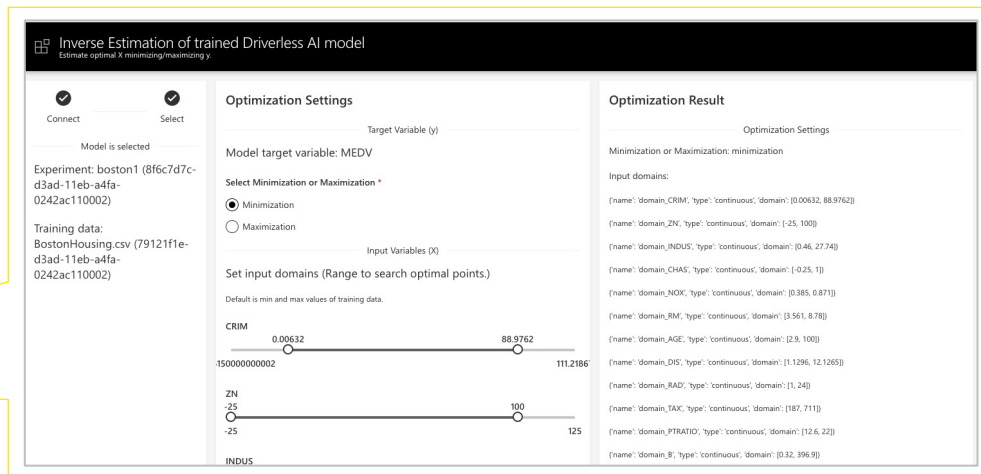
H<sub>2</sub>O  
DRIVERLESSAI

- スコアリング用インプットの連携

- スコアリング結果の連携



- App画面
- 最適化計算の実施

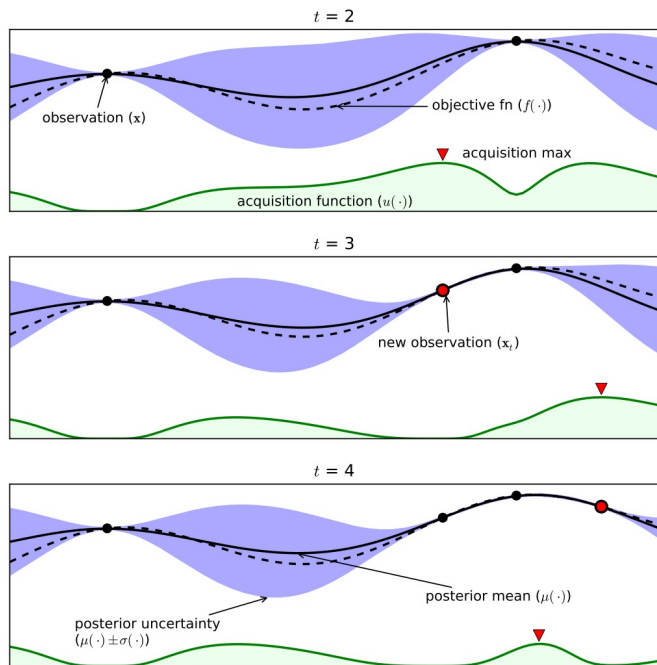


## 【逆問題解析】

作成した予測モデル( $y = f(X)$ )において、ターゲット変数( $y$ )を最大化・最小化する入力変数( $X$ )の値を計算する

# 【参考】 ベイズ最適化

## 1次元インプット、最大値探索の例



-- (黒の破線、 $f(\cdot)$ )

➤ 探索対象のBlack Box関数 (Driverless AIモデル)

— (黒の実線、 $\mu(\cdot)$ )

➤ 代理関数 (Black Box関数に代わってその形状を求め、最大値を探索する)

➤ 代理関数にはガウス過程 (関数が分布として得られ、推定の不確かさが表現できる ( $\mu(\cdot) \pm \sigma(\cdot)$ )) が用いられ、新しいデータ (インプットとそれに対応するBlack Box関数のアウトプット) が取得されるたびに分布が更新される

— (緑の実線、 $u(\cdot)$ )

➤ Acquisition関数 (最大になる値に対応するインプットが、次に探索するインプットの値 (そのインプットの時のBlack Box関数のアウトプットを計算))

➤ 代理関数の最大値と、その分布の不確かさ (紫の影の部分) を元に関数の形状が決定

➤ 関数の形状を決める方法として、Maximum Probability of Improvement, Expected Improvement, Lower/Upper Confidence Boundといったものがある

現時点でのデータ (2つの黒い点) からガウス過程を推定し、それに基づいた Acquisition関数を計算。Acquisition関数が最大となるインプットを算出 (赤い三角)

新しく観測したものを加えたデータ (2つの黒い点と1つの赤い点) からガウス過程を推定、それに基づいた Acquisition関数を計算。Acquisition関数が最大となるインプットを算出 (赤い三角)

上記を繰り返す。

(イテレーション回数の上限や収束制約等に達するまで繰り返しを実施)

# Step1: Driverless AIへの接続

**Inverse Estimation of trained Driverless AI model**  
Estimate optimal X minimizing/maximizing y.

1 Connect 2 Select

Connect to DAI

ⓘ This app is not connected to DAI!

Driverless AI URL \*

http://52.202.108.180

Driverless AI Username \*

h2oai

Driverless AI Password \*

.....

Connect

- Driverless AIと接続

**Inverse Estimation of trained Driverless AI model**  
Estimate optimal X minimizing/maximizing y.

✓ Connect 2 Select

Select a Model

Driverless AI Models \*

boston1

Select

- 学習済みモデル(Experiment)の指定

# Step2: 最適化計算の設定

fx Inverse Estimation of trained Driverless AI model  
Estimate optimal X minimizing/maximizing y

Connect ☒ Select ☒

Model is selected

Experiment: cc3-test  
(575939ba-afc7-11ec-98cf-0242ac110002)

Training data:  
UCL\_Credit\_Card3.csv  
(e04af12e-9dec-11ec-a1da-0242ac110002)

Optimization Settings

Target Variable (y)

Model target variable: LIMIT\_BAL

Select Minimization or Maximization \*

☒ Minimization ☐ Maximization

Input Variables (X)

Categorical columns

Select levels of categorical columns to search. ([nan] is missing value in training data.)

SEX \*

M x F x

EDUCATION \*

university x graduate school x high school x

MARRIAGE \*

married x single x others x nan x

Numerical columns

Set input domains of numerical columns. (Range to search optimal points.)  
Default is min and max values of training data.

AGE

21  79

PAY\_1

-2  8

PAY\_2

-2  8

PAY\_3

-2  8

PAY\_4

-2  8

- 選択された Experimentと学習データの情報

- カテゴリカル型数値型特徴量(X)の探索カテゴリの選択

- 数値型特徴量(X)の探索範囲の設定

- Target変数(y)の最小化・最大化の選択

Other settings

acquisition type

Expected Improvement

max iteration

15

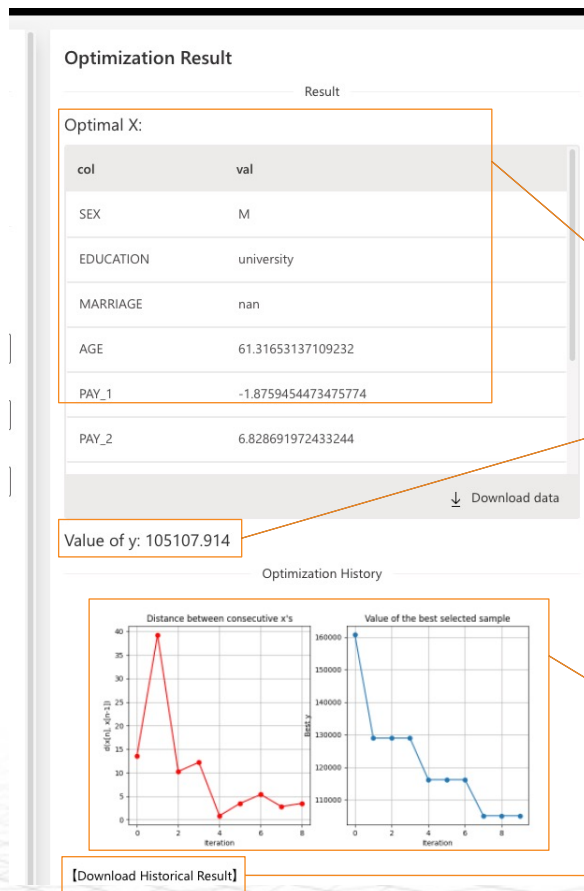
max run time (sec)

600

Optimize

- 最適化のパラメータ
  - Acquisition関数
  - 実行時間制約(最大イテレーション回数、最大計算時間)

# Step3: 最適化の結果



- 特徴量変数の最適化計算結果 ( $\hat{X}$ )

- 最適化計算結果 ( $\hat{X}$ ) におけるターゲット変数の値

- 計算履歴のプロット  
([https://gpyopt.readthedocs.io/en/latest/GPyOpt.plotting.html#GPyOpt.plotting.plots\\_bo.plot\\_convergence](https://gpyopt.readthedocs.io/en/latest/GPyOpt.plotting.html#GPyOpt.plotting.plots_bo.plot_convergence))

- 途中計算含めた、計算結果のダウンロード