

Multi-Sigma

スタートアップガイド

更新日：2022/10/20

目次

(重要) Multi-Sigma の使用条件	4
1. 学習データの作成.....	5
2. プロジェクトの作成.....	6
3. データのアップロード	7
4. タスクの作成.....	8
5. 前処理.....	9
6. AI 学習.....	10
7. 最適化	12
8. 最適化の際に説明変数に制約を与えたい場合.....	13
9. 未知の説明変数に対して結果を予測されたい場合	16
10. 要因分析	18
10. ベイズ最適化	20
11. 実験計画作成	22
12. 学習データの統計解析.....	24
13. AI の予測精度を上げるためのテクニック	26

13.1. 学習データの数の修正	26
13.2. アンサンブル解析	28
13.3. 自動設定後にモデルの数を増加	29
謝辞	30

(重要) Multi-Sigma の使用条件

- 1ユーザー1アカウントのご利用になります。
- 1アカウントで**同時に実行できる解析処理は、1つ**のみです。オートチューニング等の複数のジョブを同時に並列で実行することは出来ません (Limit exceeded エラーとなります)。
- 解析のためにアップロードできる**ファイルの容量は3MB**までです。
- 1アカウントで保存できる**データサイズの合計は500MB**までです。
- 各解析処理画面で実行ボタンを押してから、実行が始まる、あるいは結果が返ってくるまで時間がかかる場合がございます。反応がなくても、サーバーにはジョブが発行されておりますので、**実行ボタンを何度も押さないよう、お願いいたします。**

1. 学習データの作成

・作業に入る前に、お手持ちのデータのヘッダに各パラメータ名が記載されているのを確認の後、説明変数と目的変数の CSV ファイルに分けて保存しておいてください。

※1 インプットデータ、アウトプットデータには、それぞれ下記のようなパラメータ名のヘッダが必要です。パラメータ名は、英数字（記号、スペース不可）のみご利用いただけます。

また、オートチューニングを利用される場合は、20 程度データがあることが望ましいです。

※2 データ中に、欠損、文字情報などがあると、エラーになります。EXCEL などでは指数の e など表記されている場合、CSV 化する際に、文字情報として認識される場合があります。

インプットデータ例

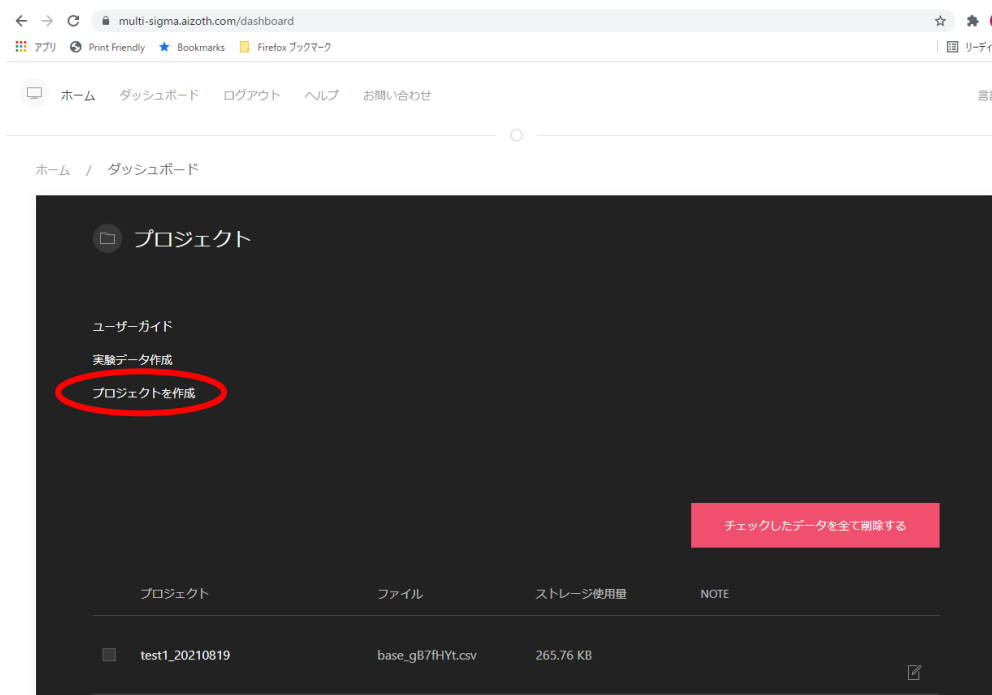
Input1	Input2	Input3
34.99273	0.096485	1.788765
68.31001	0.086094	3.270479
75.34887	0.056819	5.677714
41.38881	0.039424	1.606276
73.72395	0.027785	2.466618
27.19137	0.066883	9.497874
69.33889	0.000507	2.232949
14.19282	0.029894	0.119909
12.12865	0.055605	4.701682
92.03122	0.093505	8.898504
2.51042	0.041577	7.852931
28.8044	0.086406	6.84557
30.03625	0.07067	7.872244
6.675493	0.028	4.607661

アウトプットデータ例

Output1	Output2
69.22956	4.661119
85.99275	5.940414
60.45739	3.815804
56.46706	3.710678
78.26289	5.911408
53.23912	3.457923
59.339	3.956873
72.94924	5.041446
85.28639	6.34098
82.89526	5.793204
74.58631	4.903518
76.53725	5.47825
57.32129	3.9866
63.36183	4.760538

2. プロジェクトの作成

ダッシュボード画面にて、黒い画面領域の左上から3番目のリンク、「プロジェクトを作成」をクリック。



本稿では、ニューラルネットワーク解析+多目的遺伝的アルゴリズムをデフォルトとして説明を行いますので、プロジェクト作成画面では、「AI 解析」を選択ください。ベイズ最適化のご利用の方は、「10. ベイズ最適化」の章をご参照ください。



3. データのアップロード

プロジェクト作成画面にて、プロジェクト名を記入したのち、インプットファイルに説明変数の CSV ファイルを、アウトプットファイルに目的変数の CSV ファイルをアップロードし、「保存」をクリック。

The screenshot shows a web browser window with the URL `multi-sigma.ai/oth.com/business`. The page title is "プロジェクトを作成" (Create Project). The form contains the following elements:

- A search icon and the text "プロジェクトを作成" at the top of the form area.
- A text input field labeled "プロジェクト名" (Project Name).
- An "Input file" section with a "ファイルを選択" (Select File) button and the text "選択されていません" (None selected).
- An "Output file" section with a "ファイルを選択" (Select File) button and the text "選択されていません" (None selected).
- A blue "保存" (Save) button at the bottom right of the form.

→ダッシュボード画面にて、プロジェクトが作成されたのをご確認ください。

4. タスクの作成

作成されたプロジェクトをクリックした後、「タスクを作成」をクリックし、タスク作成画面にてタスク名をご記入ください。

(こちらのタスクが、個別の分析作業になります)

ホーム / ダッシュボード / プロジェクト: コンクリート_20220217

コンクリート_20220217 プロジェクト

タスクを作成

			プロファイリングを作成	チェックしたデータを全て削除する		
タスク名	データ前処理	AI学習	AI予測	要因分析	最適化	備考
<input type="checkbox"/> task_20220404 2022年4月4日10:16	設定	START	START	START	START	📝

※Kaggle よりダウンロードした下記のデータを用いて解析されています。

PRATHAM TRIPATHI, Calculate Concrete Strength

<https://www.kaggle.com/datasets/prathamtripathi/regression-with-neural-networking>

5. 前処理

作成されたタスク画面において、「前処理」をクリックし、まずは全てのパラメータに対して「Max-Min」（デフォルト設定）を選択して、「確定」をクリック頂ければと思います。

ご自身でデータの前処理をされている場合は、「none」をご選択ください。

列名	タイプ	最大	最小	平均	標準偏差	一括選択	一括選択	一括選択	一括選択
Strength	OUTPUT	82.6	2.33	36.2	17.1	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
Cement	INPUT	540.0	102.0	288.0	105.0	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
Slag	INPUT	359.0	0.0	70.3	86.0	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
FlyAsh	INPUT	200.0	0.0	50.6	62.7	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
Water	INPUT	247.0	122.0	181.0	21.6	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
Superplasticizer	INPUT	32.2	0.0	5.89	6.1	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
CoarseAggregate	INPUT	1140.0	801.0	979.0	76.2	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
FineAggregate	INPUT	993.0	594.0	775.0	81.3	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None
Age	INPUT	365.0	1.0	47.6	66.2	<input type="radio"/> Normalization	<input checked="" type="radio"/> Max-Min	<input type="radio"/> Abs-max	<input type="radio"/> None

※1 前処理はやり直しが出来ませんので、もし他の前処理方法を試されたい場合は、こちらのタスク作成から作業をやり直すことになります。

※2 前処理の選択ですが、特に目的変数については、同じ前処理方法（特に目的変数が複数の場合は Max-Min）を選択いただくことをお勧めします。前処理方法によって、前処理後の数値の大きさに差異が生じます。目的変数によって、前処理後の数値の大きさに差異が生じると、各目的変数の予測精度に差異が生じる可能性がございます。

※3 データの中身が、全て 0、あるいは同じ数値の場合、前処理方法ではそれぞれ None、Abs-max が自動的に選択され、変更することは出来ません。こちらは前処理エラーを回避するための措置ですので、そのままご利用ください。

6. AI 学習

AI 学習画面にて、もし、簡単に解析を試されたい場合は、デフォルトの設定のまま画面下の「保存」をクリックください。

画面左のコンソール部分で、「AI 学習」のタブが「Done」になりましたら、分析作業が終わります（4～5分程度かかります）。

The screenshot displays the 'AI設定' (AI Settings) page. On the left sidebar, there are navigation options: '一覧へ戻る', 'task_20220404', 'ファイルサイズ 122.86 KB', '入力ファイル: Input_Demo_Learning_20210728.csv', '出力ファイル: Output_Demo_Learning_20210728.csv', 'データの処理' (with '設定を見る'), 'AI学習' (with 'START'), 'AI予測' (with 'START'), '要因分析' (with 'START'), and '最適化' (with 'START'). The main content area is titled 'AI設定' and includes a '自動設定' button. The settings are as follows:

- Number of iteration: 5
- Number of outputs: 1 (Note: この値は変更できません。)
- Number of epoch: 1000
- Number of hidden layers: 1
- Number of neurons of hidden layers: 10
- Dropout of input layer: 0 (Note: この値は、Batch Normalizationがオフのときに編集可能です。)
- Dropout of other layers: 0 (Note: この値は、Batch Normalizationがオフのときに編集可能です。)
- Size of training data: 837
- Size of batch(≤ Size of training data): 84
- Number of patience: 500
- Number of split: 0.1
- Batch normalization: ON
- Check random: ON

A red dashed circle highlights the 'Dropout of input layer' and 'Dropout of other layers' fields, with a red arrow pointing to a '※1' symbol. Another red dashed circle highlights the 'Batch normalization' dropdown menu. A red circle highlights the '保存' (Save) button at the bottom right.

もし、精度よく解析されたい場合は、「自動設定」をクリック頂き、数時間お待ち頂ければ（データ数や説明変数の数により5時間～半日程度で変動）、精度良く予測するためのAIの学習条件を自動で探索致します。

こちら、作業中にアプリを閉じても、PC をシャットダウンしても処理は続きます。一方、一度処理をスタートしてしまいますと、その時間は他の解析作業が出来なくなってしまうので、その日の仕事を終わられる前にクリックして、翌日結果を確認するような使い方がおすすめです。

※1 Dropout と Batch Normalization は併用できません。Batch Normalization の使用時（On の時）は、Dropout は 0（Dropout を使用しない）が強制的に設定され、デフォルトでは Batch Normalization が設定されています。自動設定では Batch Normalization が使用されます。

7. 最適化

6の画面下部に、無事AIのモデルが作成されましたら、左側コンソールの「最適化」をクリック頂き、最適化画面にうつってください。

下記の最適化画面において、各目的変数の最大化/最小化/非制御/目標値を選択いただき、画面中断あたりのAIモデルの中で、最も誤差の「小さい」モデルを選択した上で、画面下の「保存」をクリックしてください。

非制御を選択頂くと、その目的変数は無視して最適化されます。

目標値を選択いただくと、テキストフォームに入力いただいた数値に近い解を、優先的に探索いたします。

入力ファイル:
input_CFD総括データ(学習20).csv
出力ファイル:
output_CFD総括データ(学習20).csv

データ前処理
設定を見る

AI学習
DONE

AI予測
START

要因分析
START

最適化
DONE

最適化設定

出力値

タイプ

Thrust 最大化 最小化 非制御 目標値 30.0

DI 最大化 最小化 非制御 目標値 0.005

入力値

タイプ

	最大値	最小値
Mizo <input checked="" type="radio"/> 実数 <input type="radio"/> 整数	18.0	3.0
Angle <input checked="" type="radio"/> 実数 <input type="radio"/> 整数	179.0	29.0

他の設定値は、デフォルトのままです。ただし、目的変数の数が多かったり、解の収束が悪い場合は、「Size of generation」を10から100までの範囲で変更し、世代数を増加することで、より最適化の精度が向上する可能性があります。

こちら、分析作業終わりましたら、左側コンソールの最適化のタブが「Done」になりますので、そちらをクリックしてください。(1~2分程度かかります)

最適化画面において、結果をダウンロードするか、あるいは画面下の表示部分において、縦軸、横軸のパラメータを選択し、結果を表示ください。

※ 各説明変数の最大値、最小値に、アップロードされた説明変数の範囲の外の数値を入力することも可能ですが、予測精度は保証されません。

8. 最適化の際に説明変数に制約を与えたい場合

目的変数に制約をかけたい場合は、下記の通り、最適化画面の制約の文字の右側にある「+」の部分をクリックすると、制約条件を設定するフォームが表示されます。

The screenshot shows the optimization interface. On the left is a sidebar with navigation buttons. The main area is divided into sections: '入力値' (Input Values), '制約' (Constraints), and '設定' (Settings). Under '入力値', there are four rows of input fields for 'Mizo', 'Angle', 'Depth1', and 'Depth2'. Each row has radio buttons for '実数' (Real) and '整数' (Integer), and two numerical input boxes for '最大値' (Maximum) and '最小値' (Minimum). Under '制約', there is a '+' button circled in red. Under '設定', there is a section for 'アンサンブルモデル' (Ensemble Model) with a list of models and their scores.

タイプ	最大値	最小値
Mizo <input type="radio"/> 実数 <input checked="" type="radio"/> 整数	18	3
Angle <input type="radio"/> 実数 <input checked="" type="radio"/> 整数	179	10
Depth1 <input checked="" type="radio"/> 実数 <input type="radio"/> 整数	0.24	0.05
Depth2 <input checked="" type="radio"/> 実数 <input type="radio"/> 整数	0.24	0.06

制約 +

設定

アンサンブルモデル

model10.h5	0.0997696186087748
model9.h5	0.0839986925985935

「+」をクリックすると、下記の通り、各説明変数の係数と合計値、および合計値に対する条件（以上/同値/以下）を選択する画面が展開されますので、ご希望の制約条件を設定してください。

下記画面例： $10 \times \text{Mizo} + 1 \times \text{Angle} = 200$ （赤字が設定）

The screenshot shows the constraint setting interface. The '制約' (Constraints) section is highlighted with a red box. It contains a table for setting coefficients and a section for setting the total value and its condition.

タイプ	係数
Mizo	10
Angle	1
Depth1	0
Depth2	0

合計値 以上 同値 以下

合計値: 200

設定

アンサンブルモデル

model10.h5	0.0997696186087748
model9.h5	0.0839986925985935
model8.h5	0.0497686787644512

下記は、前述の条件で探索された結果をダウンロードしたのですが、I列の合計値の結果を見て頂くと、一部完全に制約条件を満たさないものが散見されるものの、概ね制約条件を満たす結果が探索されています。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Generation	Thrust	DI	Mizo	Angle	Depth1	Depth2		=10*Mizo + 1*Angle								
2	1	18.89416	0.000748	3	70	0.239258	0.069979		100								
3	1	18.7202	0.000787	3	70	0.239063	0.075263		100								
4	1	18.51352	0.000793	5	50	0.238535	0.062867		100								
5	1	17.88462	0.001139	3	70	0.228145	0.095139		100								
6	1	17.5795	0.001259	6	40	0.220556	0.073454		100								
7	1	16.55758	0.001201	3	70	0.239731	0.143334		100								
8	1	16.87138	0.001352	6	41	0.223595	0.096274		101								
9	1	16.79149	0.001818	4	60	0.198952	0.10365		100								
10	1	16.47593	0.001476	5	50	0.222745	0.117286		100								
11	1	16.69535	0.002474	6	39	0.157094	0.072315		99								
12	1	16.56554	0.00207	4	60	0.18634	0.103696		100								
13	1	16.5347	0.001812	5	50	0.201235	0.103601		100								
14	1	15.94224	0.00138	4	60	0.235264	0.152612		100								
15	1	16.08887	0.0015	5	50	0.226285	0.131235		100								
16	1	15.70253	0.001547	5	49	0.228558	0.144422		99								
17	1	16.51924	0.002384	4	60	0.164839	0.092758		100								
18	1	15.19656	0.001722	4	60	0.221382	0.170816		100								
19	1	14.09834	0.001687	4	60	0.238091	0.220246		100								
20	1	12.85099	0.001746	4	60	0.055383	0.202099		100								
21	1	12.98589	0.001756	6	41	0.22886	0.231467		101								
22	2	18.89416	0.000748	3	70	0.239258	0.069979		100								
23	2	18.82199	0.000742	4	60	0.239109	0.062867		100								

※なお、係数の数値が全て0の時は、説明変数に制約がかかりません（デフォルトの設定）。また、係数および合計値の一部に負の値が設定された場合や、あるいは制約を満たす条件が存在しない場合は、制約条件を満たさない結果が出力されることがあります。特に、係数の大きさに差異がある場合は、制約がかかりにくくなりますので、結果をダウンロードして、ご確認ください。

下記の事例では、Mizoの係数に-5が設定されています。

制約

タイプ	係数
Mizo	-5.0
Angle	1.0
Depth1	0.0
Depth2	0.0
合計値	<div style="display: flex; align-items: center; gap: 10px;"> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="radio"/> 以上 <input checked="" type="radio"/> 同値 <input type="radio"/> 以下 </div> <input style="width: 100px;" type="text" value="100.0"/> </div>

下記は、上記の条件で探索された結果をダウンロードしたのですが、I列の合計値の結果をみて頂くと、条件に近い結果が探索されてはいるものの、制約を完全に満たしきれていないことが確認されます。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Generation	Thrust	DI	Mizo	Angle	Depth1	Depth2		=-5*Mizo + 1*Angle								
2	1	16.92854	0.00085	8	142	0.233843	0.097841		102								
3	1	17.36777	0.000931	3	102	0.238149	0.111287		87								
4	1	18.48882	0.000967	3	102	0.211554	0.060382		87								
5	1	18.30147	0.001022	3	102	0.210536	0.064989		87								
6	1	17.64908	0.001072	5	126	0.206185	0.075461		101								
7	1	17.62649	0.001273	3	102	0.202058	0.078649		87								
8	1	17.04289	0.001232	7	134	0.204265	0.086435		99								
9	1	17.31551	0.001801	3	102	0.161908	0.061103		87								
10	1	15.7793	0.001207	8	139	0.239367	0.143494		99								
11	1	17.15054	0.001535	3	102	0.187518	0.084454		87								
12	1	16.23496	0.001329	3	102	0.23203	0.149861		87								
13	1	15.4643	0.00131	12	159	0.237973	0.146313		99								
14	1	16.20991	0.001366	3	102	0.228953	0.148672		87								
15	1	16.30716	0.001582	3	102	0.203537	0.127798		87								
16	1	15.83315	0.001459	3	102	0.231024	0.163776		87								
17	1	16.75423	0.001979	3	105	0.151178	0.070266		90								
18	1	15.55581	0.001426	3	117	0.229826	0.165724		102								
19	1	16.30094	0.001863	12	161	0.155402	0.065945		101								
20	1	16.40918	0.00223	3	102	0.137641	0.072051		87								
21	1	11.41117	0.001412	12	160	0.065328	0.237396		100								
22	2	18.94501	0.000631	3	102	0.230978	0.060382		87								
23	2	18.8579	0.000695	3	102	0.227269	0.060382		87								

9. 未知の説明変数に対して結果を予測されたい場合

「AI 予測」画面において、予測モデルを選択し、未知の説明変数のリストの CSV ファイルをアップロードし、「保存」を推して頂くと、予測結果がダウンロード頂けます。

※ インプットデータ、アウトプットデータには、パラメータ名のヘッダが必要です。パラメータ名は、英数字（記号、スペース不可）のみご利用いただけます。

model5.h5	0.0592383643273682
model4.h5	0.0607962449048228
model3.h5	0.0563612167850125
model2.h5	0.060956345546087

Shift + クリックで複数のモデルを選択。
CTRL + クリック (Win) / CMD + クリック (Mac) で選択または選択解除。]

入力ファイル
[ファイルを選択] Allinput_2...7_Test.csv

保存

予測した結果に対して、後に実際の結果が得られましたら、そちらの結果を画面下部においてアップロードしますと、予測と実際の結果の誤差を得ることが出来ます。

※ こちらの誤差は、下記の相対誤差を用いて計算されているため、実測値が 0 の場合、あるいは予測値が非常に小さい数値の場合（例：0.00003）、実測値と予測値の数値が近い場合でも、誤差の数値が非常に大きな数値となる場合がございます。誤差の数値が大きな場合は、結果をダウンロードして、EXCEL 等で実測値と予測値を実際に比較されることをお勧めします。

$$\text{相対誤差} = |(\text{実測値} - \text{予測値}) / \text{予測値}|$$

一覧へ戻る

task_20220404

ファイルサイズ 406.02 KB

入力ファイル:
Input_Demo_Learning_20210728.csv

出力ファイル:
Output_Demo_Learning_20210728.csv

データ前処理

設定を見る

AI学習

DONE

AI予測

DONE

DONE 結果ファイルのダウンロード

検証

検証ファイル
現在: uploads/user/2/input_data_ai/1421/Output_Demo_Test_20210728.csv
変更: 選択されていません

Strength: **12.92%**

検証する

10. 要因分析

「要因分析」画面において、予測するモデルを選択いただき、「実行」をクリック頂くと、要因分析の結果が表示されます。

Slag	<input type="text" value="359.0"/>	<input type="text" value="0.0"/>
FlyAsh	<input type="text" value="200.0"/>	<input type="text" value="0.0"/>
Water	<input type="text" value="247.0"/>	<input type="text" value="122.0"/>
Superplasticizer	<input type="text" value="32.2"/>	<input type="text" value="0.0"/>
CoarseAggregate	<input type="text" value="1140.0"/>	<input type="text" value="801.0"/>
FineAggregate	<input type="text" value="993.0"/>	<input type="text" value="594.0"/>
Age	<input type="text" value="365.0"/>	<input type="text" value="1.0"/>

設定

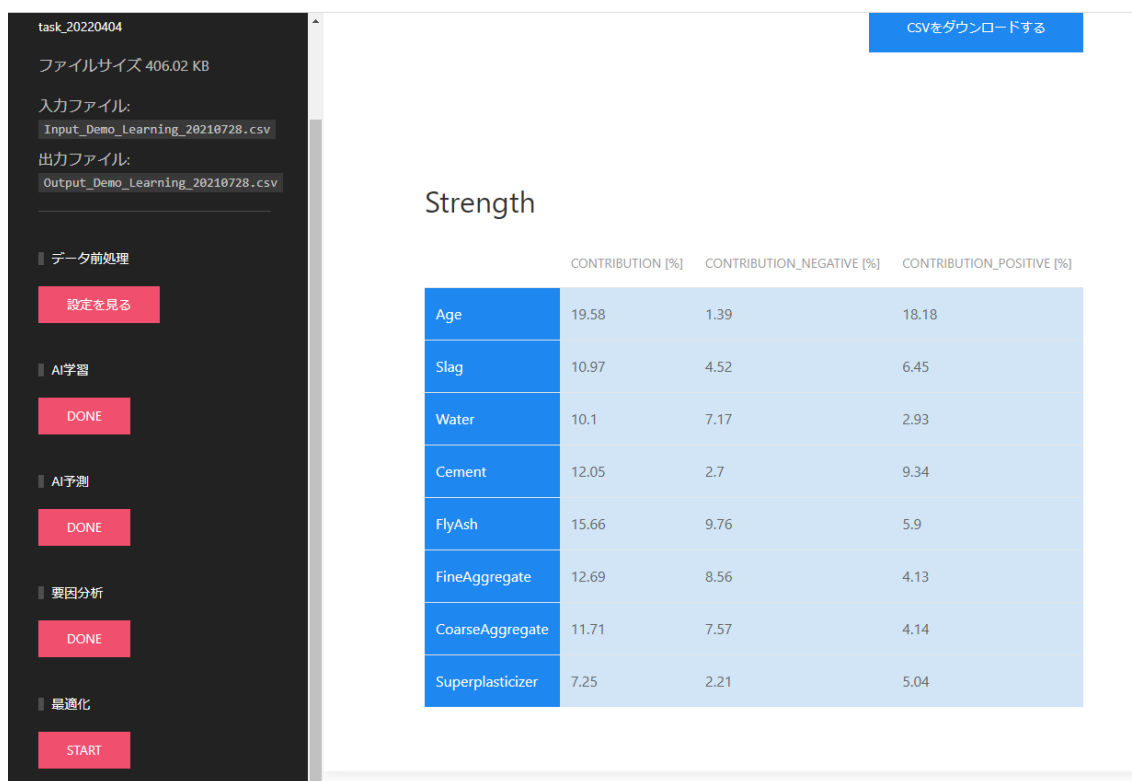
アンサンブルモデル

model5.h5	0.0592383643273682
model4.h5	0.0607962449048228
model3.h5	0.0563612167850125
model2.h5	0.060956345546087

Shift +クリックで複数のモデルを選択。
CTRL +クリック (win) CMD +クリック (Mac) で選択または選択解除。]

実行

結果において、Contribution が、各説明変数が各目的変数に対する影響の大きさを示しており、その右側の Contribution_Positive、Contribution_Negative が、説明変数による目的変数への正・負の影響の大きさを示しています。



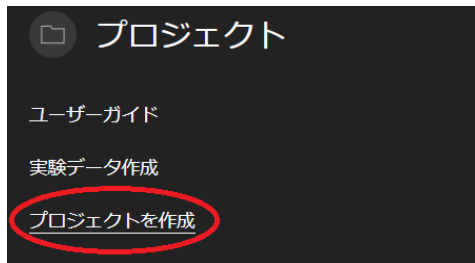
※1 Σ Contribution = 100%

Contribution = Contribution_Positive + Contribution_Negative

※2 説明変数の解析範囲を変えると、各説明変数の影響度の大きさが変わります。

10. ベイズ最適化

ダッシュボード画面において、「プロジェクトを作成」をクリックしてください。



解析手法で「ベイズ解析」を選択し、プロジェクト名を入力して保存を押してください。

A form titled 'プロジェクトを作成' (Create Project) with a search icon. It contains two input fields: '解析手法' (Analysis Method) with 'ベイズ解析' (Bayesian Analysis) selected, and 'プロジェクト名' (Project Name) with 'プロジェクト1' (Project 1) entered. A blue '保存' (Save) button is at the bottom right.

作成したプロジェクト名をクリックしてください。



入力ファイルと出力ファイルをアップロードして、保存を押してください。※

※ アップロードできるデータには下記のような上限があります。

説明変数： 10

目的変数： 2

データ数： 50

データ数+候補数： 55

A form titled 'Bayse Analysis' with a search icon. It has two sections: '入力ファイル' (Input File) and '出力ファイル' (Output File). Each section has a 'ファイルを選択' (Select File) button and the text '選択されていません' (Not selected). A blue '保存' (Save) button is at the bottom right.

表示したい入力パラメータの候補数を入力して、探索方式を選択してください。※

選択できる候補数の上限は、データ数と候補数の合計値が 55 個までです。

パラメーター

解析手法

表示する候補数

探索方式

改善期待値(Expected Improvement)

探索する入力パラメータの最小値、最大値、刻み幅を入力し、整数か実数かを選択してください。※

※ 刻み幅は、説明変数、目的変数、データの数によって入力できる数値の下限が変わります。下限値以下の数値を入力するとエラーが表示されます。

入力値

パラメーター	最小値	平均値	最大値	刻み幅	タイプ
Input1	<input type="text" value="3"/>	8.78	<input type="text" value="18"/>	<input type="text" value="1.0"/>	<input checked="" type="radio"/> 整数 <input type="radio"/> 実数
Input2	<input type="text" value="10"/>	89.90	<input type="text" value="180"/>	<input type="text" value="11.333333333333334"/>	<input type="radio"/> 整数 <input checked="" type="radio"/> 実数

出力パラメータの最適化の方向として最小化か最大化かを選択し、実行を押してください。

出力値

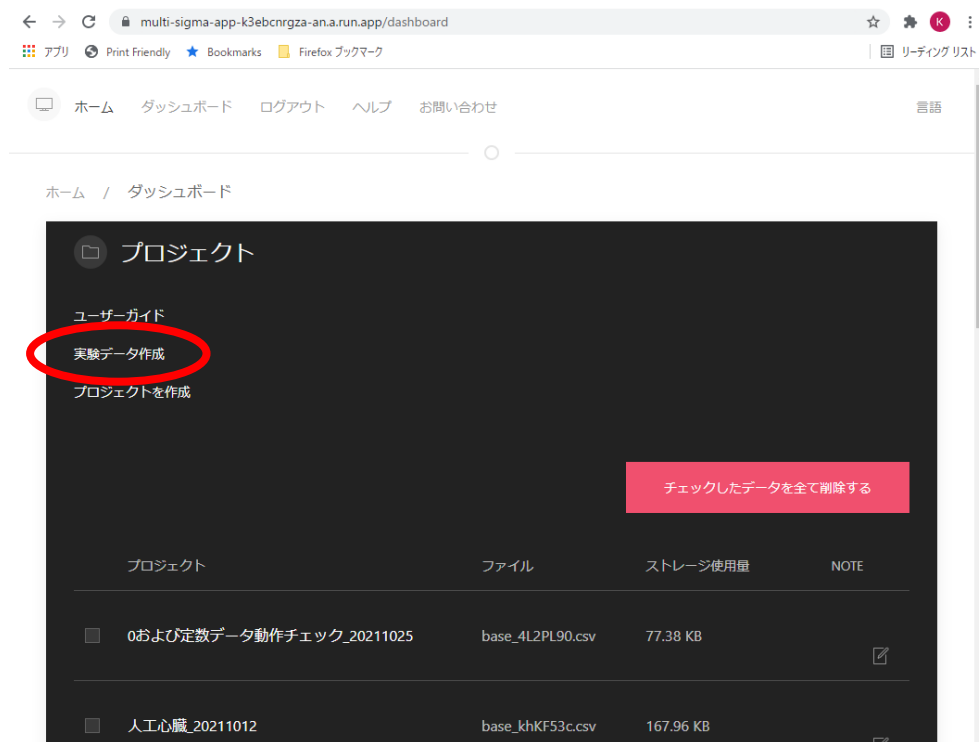
パラメーター	最小化	平均値	最大化	最適化
Output1	8.07	24.08	35.56	<input type="radio"/> 最小化 <input checked="" type="radio"/> 最大化
Output2	0.00424	0.01	0.012	<input checked="" type="radio"/> 最小化 <input type="radio"/> 最大化

実行

処理が完了すると、入力パラメータの候補と、それに対する出力パラメータの予測値が表示されます。

11. 実験計画作成

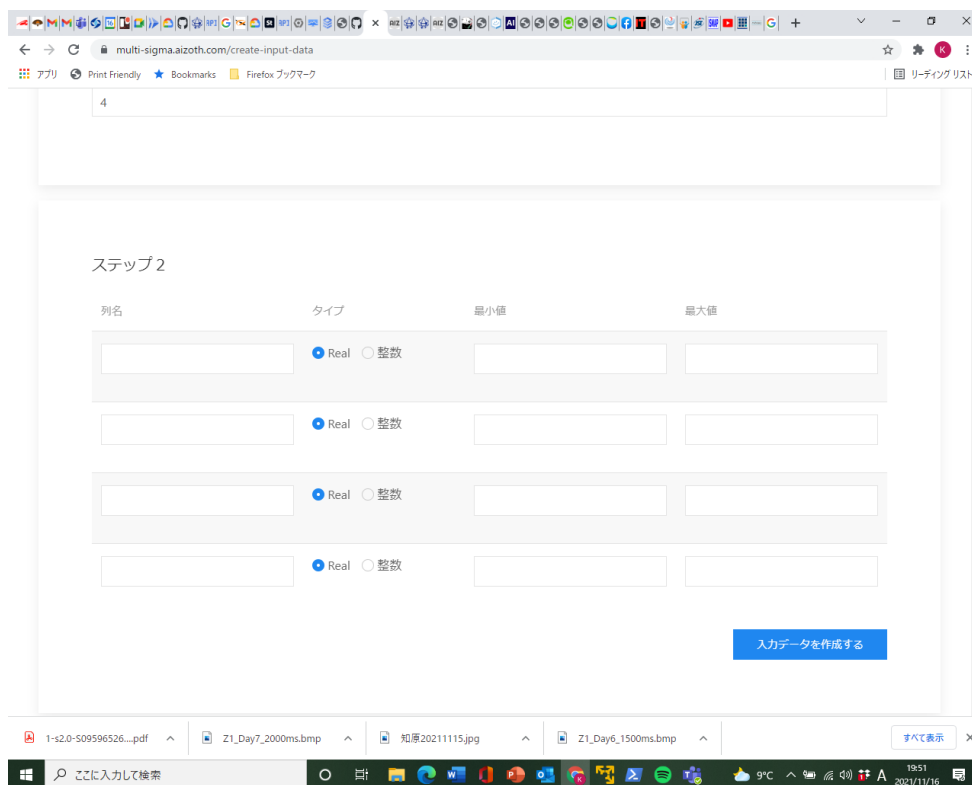
「ダッシュボード」画面において、「実験データ作成」をクリックしてください。



下記画面にて、データ数と説明変数の数を入力してください。



下記の画面より、各説明変数が実数値か整数値か、最小、最大の範囲を設定して、「入力データを作成する」をクリックすると、実験条件の CSV ファイルがダウンロードできるようになります。



12. 学習データの統計解析

プロジェクト作成後、タスク作成画面において、「プロファイリングを作成」ボタンを押してください。※

※ 説明変数の数が多くなると、プロファイリングの作成が終わらない可能性があります。

ホーム / ダッシュボード / プロジェクト: コンクリート_20220217

コンクリート_20220217 プロジェクト

タスクを作成

プロファイリングを作成

チェックしたデータを全て削除する

タスク名	データ前処理	AI学習	AI予測	要因分析	最適化	備考
<input type="checkbox"/> task_20220404 2022年4月4日10:16	設定を見る	DONE	DONE	DONE	START	

プロファイリングの作成には時間がかかります。しばらく（数分）時間が経過した後こちらの画面を参照頂くと、下記のようにデータの解析結果を参照できるようになりますので、こちらのボタンを押してください。

ホーム / ダッシュボード / プロジェクト: コンクリート_20220217

コンクリート_20220217 プロジェクト

タスクを作成

プロファイリングを開く

チェックしたデータを全て削除する

タスク名	データ前処理	AI学習	AI予測	要因分析	最適化	備考
<input type="checkbox"/> task_20220404 2022年4月4日10:16	設定を見る	DONE	DONE	DONE	START	

下記の通り、アップロード頂いたデータを統計解析した結果が表示されます。

Overview

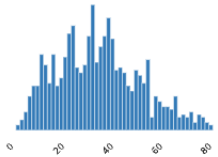
Overview Alerts (20) Reproduction

Dataset statistics		Variable types	
Number of variables	9	Numeric	9
Number of observations	930		
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	11		
Duplicate rows (%)	1.2%		
Total size in memory	65.5 KIB		
Average record size in memory	72.1 B		

Variables

Strength
Real number (R₂₀)
HIGH CORRELATION
HIGH CORRELATION

Distinct	845	Minimum	2.331807832
Distinct (%)	90.9%	Maximum	82.5992248
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	36.17349634	Memory size	7.4 KIB



Toggle details

※ 本機能は、下記の pandas-profiling の機能を用いて実装されています。

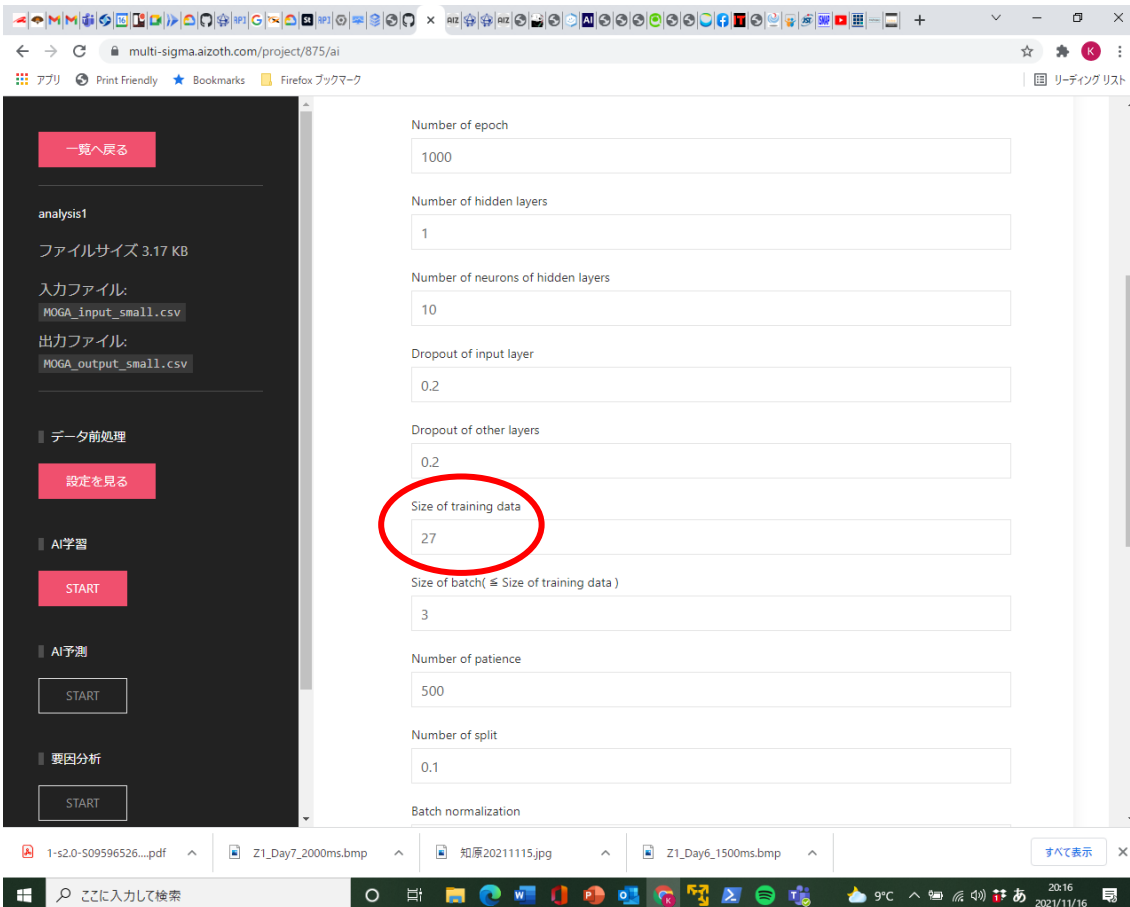
<https://github.com/ydataai/pandas-profiling>

13. AI の予測精度を上げるためのテクニック

13.1. 学習データの数の修正

下記の AI 学習のデフォルトの設定では、アップロードしたデータの 9 割を学習用に、残りの 1 割を AI の誤差を評価するための検証用のデータとして使用するよう設定されています（学習用に使用されるデータの数が、Size of training data の欄に表記）。

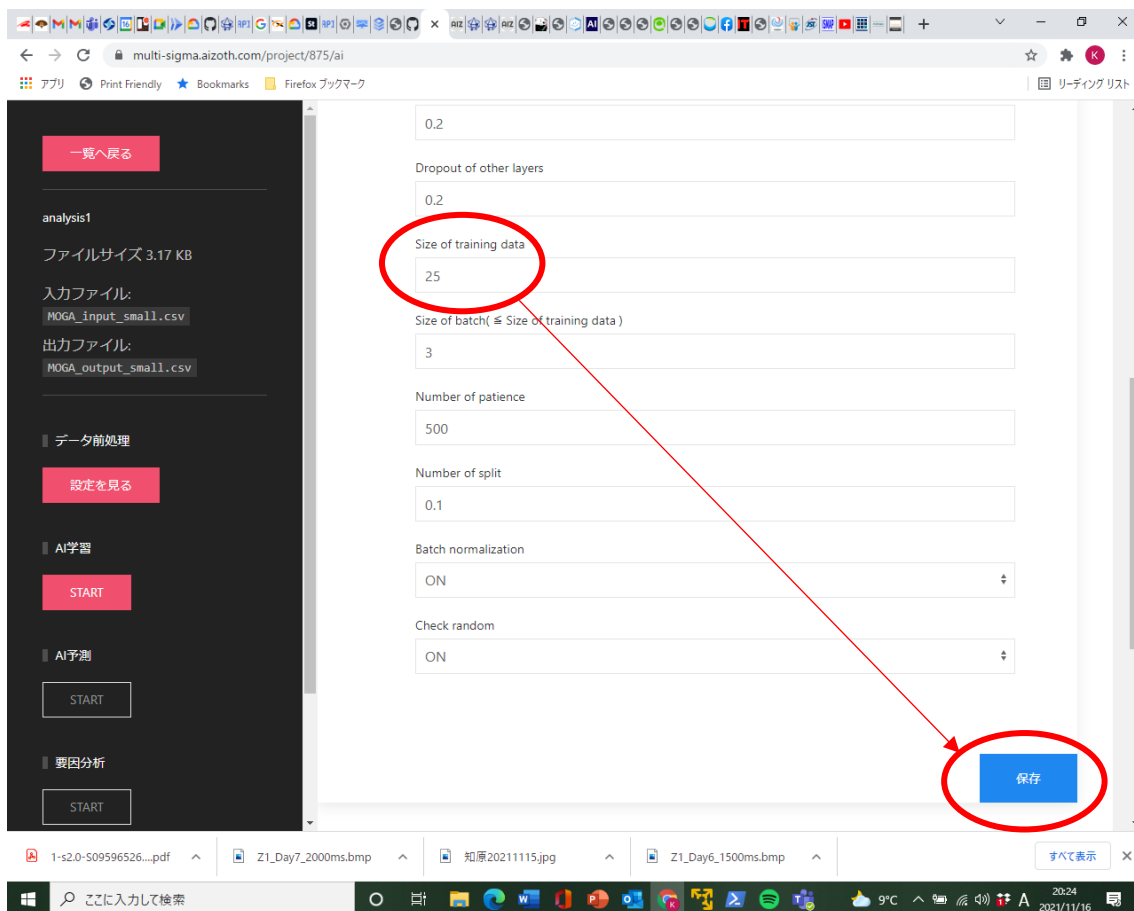
仮に、アップロードしたデータの数が 30 の場合、下記の AI 学習画面で、学習データの数が 27 で、残りの 3 つが検証用のデータとなりますが、検証用のデータが 3 つの場合、たまたまその 3 つのデータにのみ精度が良い AI が、誤差の小さい AI として評価されてしまう可能性があります。



The screenshot shows a web browser window displaying the AI learning interface. The URL is multi-sigma.aizoth.com/project/875/ai. The interface is divided into a left sidebar and a main settings area. The sidebar contains a navigation menu with options like 'データ前処理', 'AI学習', 'AI予測', and '要因分析'. The main settings area lists various parameters for training, including 'Number of epoch' (1000), 'Number of hidden layers' (1), 'Number of neurons of hidden layers' (10), 'Dropout of input layer' (0.2), 'Dropout of other layers' (0.2), 'Size of training data' (27), 'Size of batch' (3), 'Number of patience' (500), 'Number of split' (0.1), and 'Batch normalization'. The 'Size of training data' field is circled in red, highlighting the value 27. The bottom of the browser shows the Windows taskbar with the date 2021/11/16 and time 20:16.

そこで、少なくとも検証用のデータを5つ程度残すように、Size of training data の数値を変更し（30 データの場合、5 データを検証用としてのコスト、25 データを記入）、保存をクリックし、マニュアルで解析を実行します。

そうすると、学習用データの変更がシステムに反映されますので、そのあとに再度「自動設定」を実行頂くと、適切な検証データ数によって、ハイパーパラメータのチューニングが行われます。



13.2. アンサンブル解析

AI 予測、要因分析、最適化では、複数の AI を組合せて、複数の AI の解析結果の平均値を用いて解析するアンサンブル解析が可能です。特にデータの数が少ない場合、アンサンブル解析を用いることで、予測精度が向上する可能性があります。

AI 学習において、自動設定を実行すると、AI のモデルが 10 個生成されますので、10 個の AI の中で、誤差の小さい AI を **3 つ程度** 選択して解析することをお勧めします。

AI 予測において、検証用のデータを用いて、AI の組合せの精度を検証しながら、最適な組合せを探索してください。

The screenshot shows the multi-sigma.ai web interface. On the left is a dark sidebar with navigation buttons: '一覧へ戻る', 'analysis1', 'ファイルサイズ 250.39 KB', '入力ファイル: MOGA_input_small.csv', '出力ファイル: MOGA_output_small.csv', 'データ前処理' (with '設定を見る'), 'AI学習' (with 'DONE'), 'AI予測' (with 'START'), and '要因分析' (with 'START'). The main content area is titled 'inputs' and contains a table with columns 'タイプ', '最大値', and '最小値'. It lists three inputs: Input1 (Real, 92.0312, 2.51), Input2 (Real, 0.0964, 0.000506), and Input3 (Real, 9.985, 0.119). Below this is the '設定' (Settings) section, which includes 'アンサンブルモデル' (Ensemble Model) with a list of four models: model5.h5 (0.905509905530029), model4.h5 (1.24285138294792), model3.h5 (0.62915356411632), and model2.h5 (0.966792254159202). Instructions below the list state: 'Shift + クリックで複数のモデルを選択。CTRL + クリック (win) CMD + クリック (Mac) で選択または選択解除。'. Other settings include 'Size of generation' (10), 'Size of population' (100), and 'Number of crossover rate' (0.6). The browser's address bar shows 'multi-sigma.ai/oth.com/project/875/ga' and the Windows taskbar at the bottom shows the date '2021/11/16' and time '20:54'.

タイプ	最大値	最小値
Input1	92.0312	2.51
Input2	0.0964	0.000506
Input3	9.985	0.119

設定

アンサンブルモデル

model5.h5	0.905509905530029
model4.h5	1.24285138294792
model3.h5	0.62915356411632
model2.h5	0.966792254159202

Shift + クリックで複数のモデルを選択。
CTRL + クリック (win) CMD + クリック (Mac) で選択または選択解除。]

Size of generation

10

Size of population

100

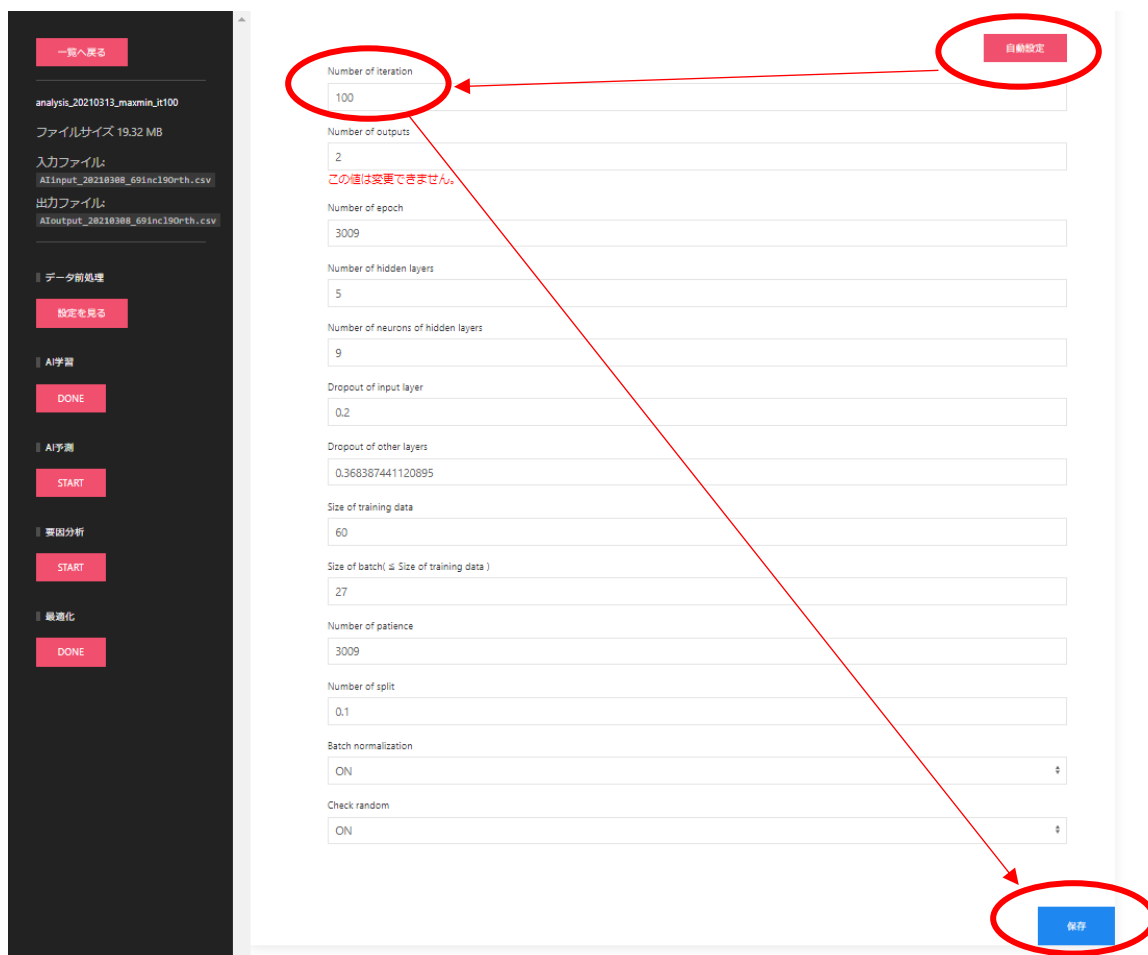
Number of crossover rate

0.6

13.3. 自動設定後にモデルの数を増加

ニューラルネットワーク解析では、ネットワークの重みと閾値の初期値を乱数で与えるために、全く同じ条件でモデルを作成しても、毎回誤差の大きさが異なります。したがって、複数モデルを作成して、その中で精度の高いモデルを使用することが重要ですが、作成するモデルの数を増やすと、より精度の高いモデルを作成できる可能性が高まります。

最初に、AI学習の画面で自動設定を行うと、自動で探索された条件で、10個のモデルがデフォルトで作成されます。次に、Number of iteration（モデルを作成する数）の数値を増加して（100個まで作成可能）、「保存」ボタンを押すと、自動設定によって探索された条件で、より多くのモデルを作成し、精度の高いモデルが作成できる可能性があります。



謝辞

- Multi-Sigma は、Keras、Tensorflow 等の Python をベースとしたオープンソースのライブラリを用いて開発されています。
- Multi-Sigma のベイズ最適化機能は、2020 年度 東京大学物性研究所 ソフトウェア高度化プロジェクトの支援を受け開発された PHYSBO※を用いて開発されています。
- 開発者の方々には、厚く御礼申し上げます。

※Tsuyoshi Ueno, Trevor David Rhone, Zhufeng Hou, Teruyasu Mizoguchi and Koji Tsuda, COMBO: An Efficient Bayesian Optimization Library for Materials Science, Materials Discovery 4, 18-21 (2016).

Available from <https://doi.org/10.1016/j.md.2016.04.001>