

表面科学セミナー2020（実践編）

実践! マテリアルズインフォマティクス 実際の作業を通して身につける材料データ科学

マテリアルズインフォマティクスの基礎と応用 **（基礎）**

～Excelではじめる機械学習チュートリアル

SETソフトウェア(株) 松下 康弘 (理学博士)

本チュートリアルのポイント

- 可解問題（トイモデル）を通じて機械学習の本質的価値を理解する
初等物理の基本問題を機械学習で解く
- 線形モデルの延長線上で非線形モデルを理解する
教師あり学習と応答局面 / 教師なし学習と情報集約
- 効率よく学習データを作成する ～ 実験計画法(直交表)の活用
- 最適設計（逆問題）へのアプローチ ～ Excelソルバーの活用

注) トイモデル（英語: toy model、toyは「おもちゃ」の意味）とは、物理学のモデリングにおいて、メカニズムを簡潔に説明するのに役立つように、細部を捨象し、意図的に単純化したモデルのことである。(Wikipediaより引用)

$$\Delta h = a_0 + a_1 l + a_2 h + a_3 \theta^\circ + a_4 v_0$$

	A	B	C	D	E	F
1	l	h	θ°	v_0	Δh	Model
2	21	17	31	55	4.382	2.171665
3	11	10	14	47	7.257	14.1675
4	17	14	13	41	10.075	12.9394
5	17	12	33	33	0.960	0.042086
6	12	11	30	61	4.072	5.380621
7	22	20	18	84	12.852	11.33182
8	14	14	43	40	0.945	-1.13033
9	13	20	53	95	2.748	-0.22686
10	13	15	51	66	-1.054	-3.65292
11	25	11	27	53	-1.738	-4.28019
12	25	12	38	31	-7.532	-9.6844
13	23	15	12	72	10.111	9.384061
14	10	11	35	48	3.998	4.242777
15	18	20	32	55	8.752	6.812617
16	13	11	21	78	6.010	9.636088
17	12	16	58	85	-3.204	-5.6709
18	10	14	60	56	-3.321	-6.98915
19	24	15	11	68	10.335	9.043607
20	23	19	51	45	-9.403	-9.01274

回帰分析

入力元
 入力 Y 範囲(Y): \$E\$1:\$E\$50
 入力 X 範囲(X): \$A\$1:\$D\$50

ラベル(L) 定数に 0 を使用(Z) 95 %
 有意水準(O)

出力オプション
 一覧の出力先(S): \$H\$9
 新規ワークシート(B):
 新規ブック(V):

残差
 残差(B):
 残差グラフの作成(D):
 観測値と残差(T):
 観測値グラフの作成(I):

正規検定
 正規検定グラフの作成(N):

概要

回帰統計	
重相関 R	0.957691
重決定 R2	0.917173
補正 R2	0.913685
標準誤差	3.261873
観測数	100

分散分析表				
	自由度	変動	分散	F 値
回帰	4	11192.7	2798.176	262.991
残差	95	1010.783	10.63982	
合計	99	12203.49		

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	22.48492	2.167814	10.37216	2.65E-17	18.18126	26.78857	18.18126	26.78857
l	-0.86685	0.050075	-17.3112	3.9E-31	-0.96626	-0.76744	-0.96626	-0.76744
h	0.86528	0.106347	8.136376	1.55E-12	0.654154	1.076405	0.654154	1.076405
θ°	-0.55543	0.020877	-26.6056	8.13E-46	-0.59688	-0.51399	-0.59688	-0.51399
v_0	0.00726	0.015137	0.479588	0.632623	-0.02279	0.03731	-0.02279	0.03731

データ分析

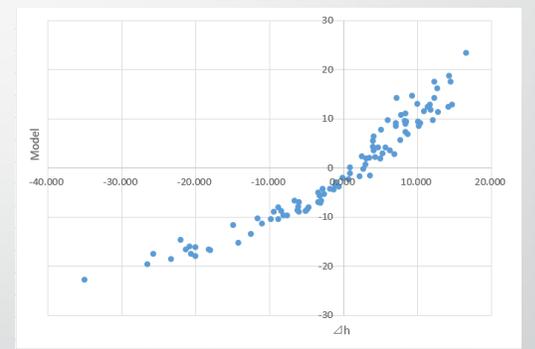
分析ツール(A)
 ヒストグラム
 移動平均
 乱数発生
 回帰分析

サンプル
 t 検定: 一对の標本による平均の検定
 t 検定: 等分散を仮定した 2 標本による検定
 t 検定: 分散が等しくないを仮定した 2 標本による検定
 z 検定: 2標本による平均の検定

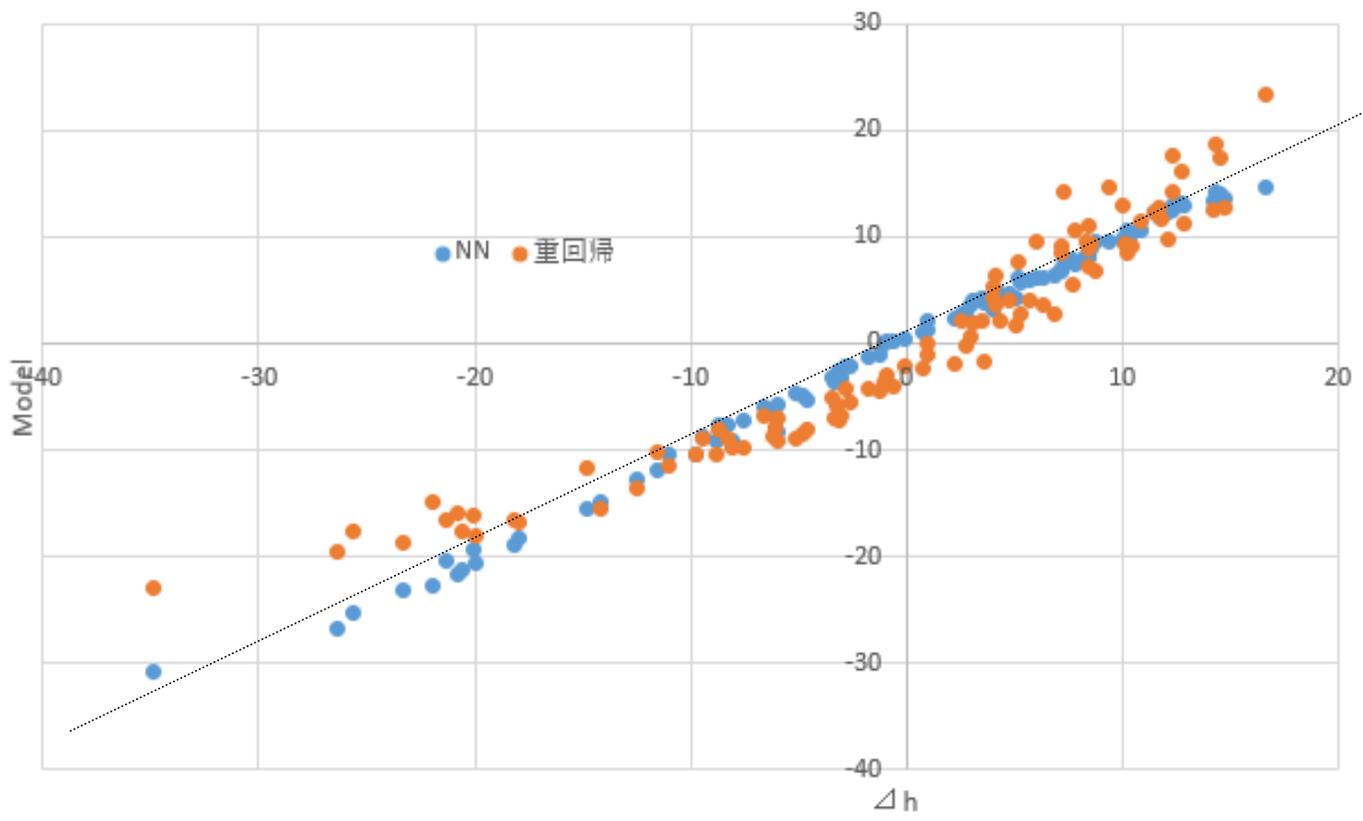
挿入

ページレイアウト 数式 データ 校閲 表示 開発 アドイン シヤチハク 電子印字

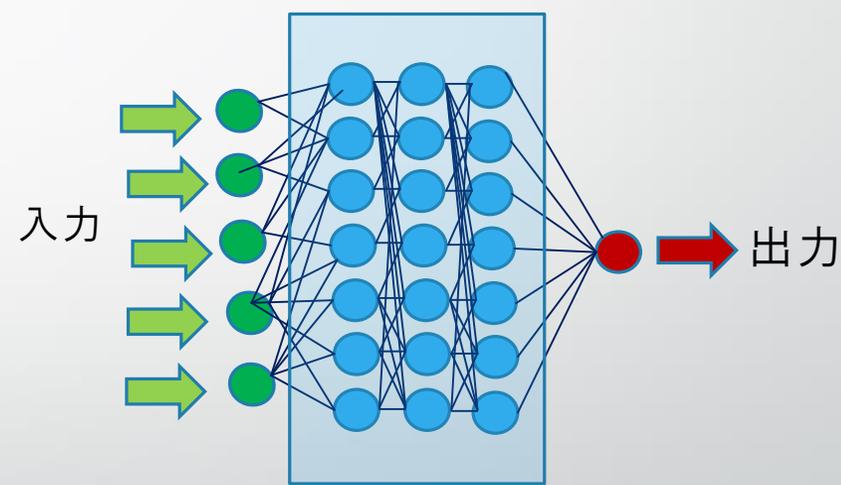
テーブル 画像 オンライン画像 ストア Bing マップ アイコン 個人用アドイン People Graph おすすめグラフ



Model1に対する重回帰モデルと多層ニューラルネットワークモデル(NN)の比較



	Δh	NN	重回帰
Δh	1		
NN	0.997706	1	
重回帰	0.957691	0.957469	1



ニューラルネットワークはデータ変換および交互作用の効果を中間層を通じて学習から獲得可能



	A	B	C	D	E	F
1		MIN	MAX		切片	22.48492
2	l	10	30		l	-0.86685
3	h	10	20		h	0.86528
4	θ°	10	60		θ	-0.55543
5	v0	30	100		v0	0.00726
6						
7	l	h	θ	v0	Δh	
8	17.33569	10	29.7796	59.27427	-6.31122E-06	
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

曲線
 $y=f(x)$

ソルバーのパラメーター

目的セルの設定:(I)

目標値: 最大値(M) 最小値(N) 指定値:(V)

変数セルの変更:(B)

制約条件の対象:(L)

- \$A\$8 <= \$C\$2
- \$A\$8 >= \$B\$2
- \$B\$8 <= \$C\$3
- \$B\$8 >= \$B\$3
- \$C\$8 <= \$C\$4
- \$C\$8 >= \$B\$4
- \$D\$8 <= \$C\$5
- \$D\$8 >= \$B\$5

制約のない変数を非負数にする(K)

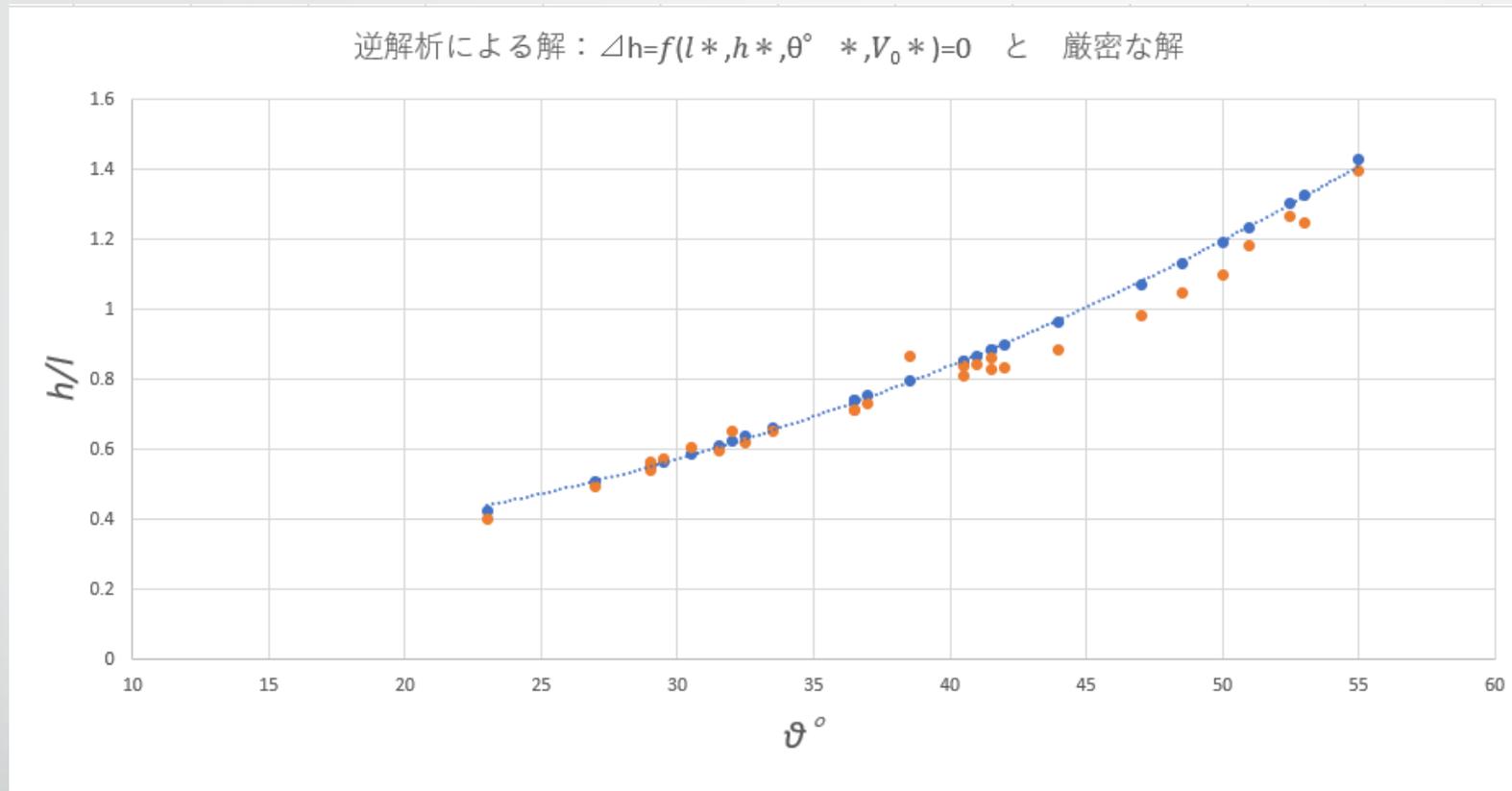
解決方法の選択: GRG 非線形

解決方法
滑らかな非線形を示すソルバー問題には GRG 非線形エンジン、線形を示すソルバー問題には LP シンプルクス エンジン、滑らかではない非線形を示すソルバー問題にはエボリューションナリー エンジンを選択してください。

$\Delta h=0$ を目標値として探索

青い曲線：古典力学から導出される関係式

オレンジ点：ニューラルネットワーク予測モデルから逆探索された最適条件データ



様々な実験データ



順問題

機械学習モデル
“ブラックボックス” (?)



逆問題

ドメイン固有の
普遍法則発見