

4. 機械学習



アウトライン

1. 機械学習
2. 機械学習の現状
3. 教師あり学習と教師なし学習
4. クラスタ分析（教師なし学習の例）
5. 線形近似（教師あり学習の例）
6. 最適化
7. 最適化の用途

4-1 機械学習

機械学習

- **機械学習**は、**コンピュータ**による**学習**
- **学習**のため、**データを使用**
- **学習**による**上達**の**能力を持つ**

知的能力の**向上**

知識の**獲得**

ルール**の獲得**

別の方法「知識やルールなどを人間が書き、コンピュータに与える」。その限界を超えることが期待できる

人工知能の種類

人工知能

機械学習

学習による上達的能力を
持つ人工知能

知的な IT システム

ルールや知識を人間が書いた
人工知能

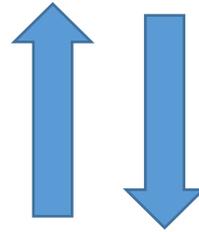
機械学習のメリット, デメリット

機械学習

学習による上達的能力を
持つ人工知能

メリット

「ルールや知識を、人間がプログラムで書かねばならないことの限界を突破



デメリット

学習不足, 過学習, などの注意点がある. 完璧に学習が成功するわけではない.

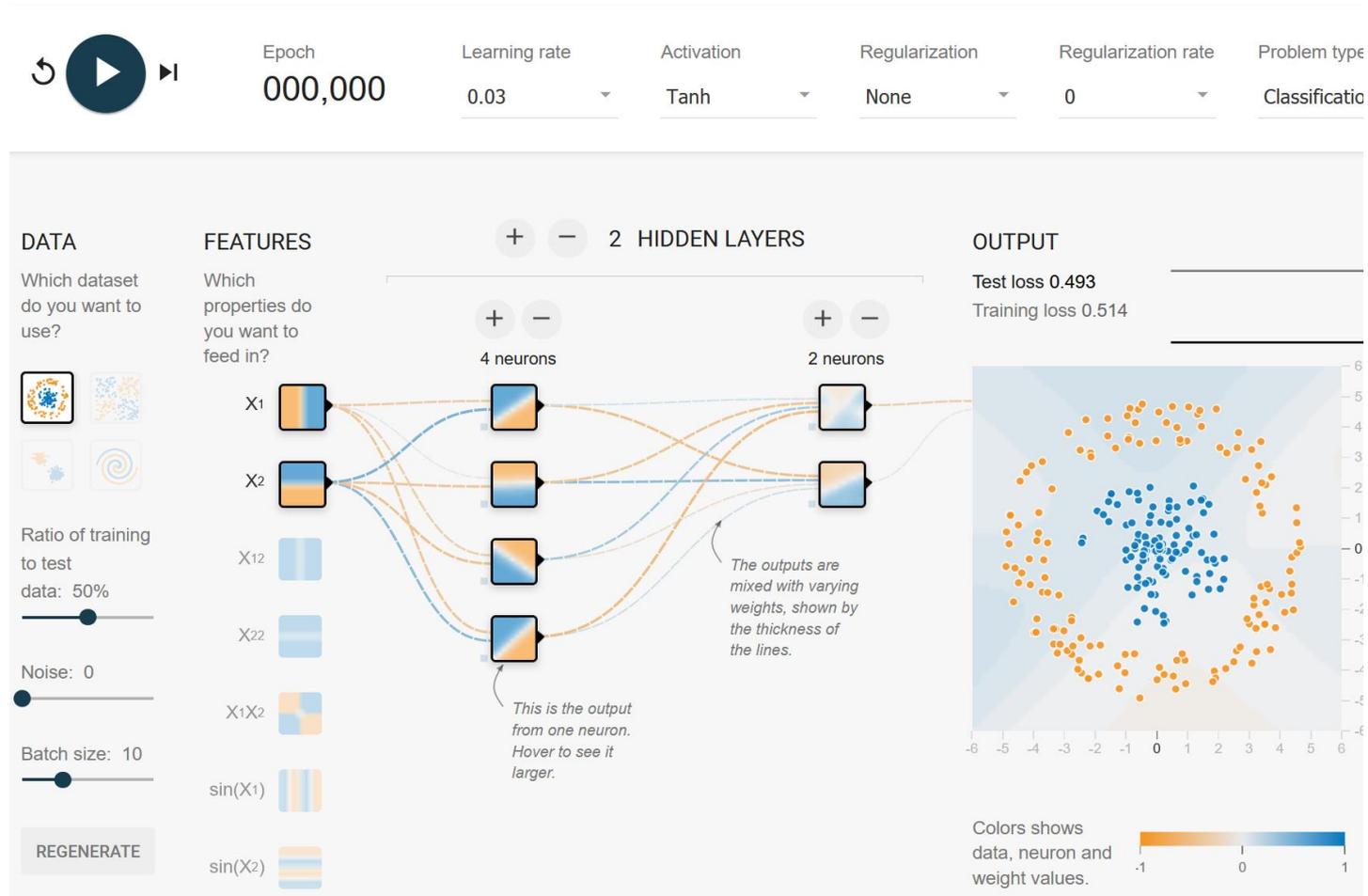
知的な IT システム

ルールや知識を人間が書いた
人工知能

機械学習のメリット

- **学習**のため、**データを使用**
- ルールや知識をプログラムとして書く必要が**ない**
→ 「機械学習は、データさえあれば、すぐに試すことができる」と言う人も
- **データの追加**による**能力向上**が期待できる
- 「すでに人間の能力を超えている」と考える人も
速度、いつでも使える、
結果が安定している

学習能力をコンピュータに組み込んでおき、 あとでデータを与えて学習させる



<https://playground.tensorflow.org>

機械学習

- **機械学習**では、データによる**学習**を行う
- **学習**に用いるデータのことを、**訓練データ**などという
- **学習を重ねることで上達する**

- 「**学習**によって、**未知のデータに対しても当てはまるパターンや規則**を、コンピュータが抽出している」という考え方もある

ここまでのまとめ

- **人工知能の学習**は、**ニューラルネットワークの結合の重み**が**変化する**もの。その技術は1950年以降変わっていない。
- より効率よく学習できる技術が進展。**学習による上達**で、ゲームなどを人間よりも上手にできるようになってきた
- 自動運転など、**さまざまな作業**について、人工知能が学習し、上達できる可能性が出てきた

4-2 機械学習の現状

機械学習の能力，応用分野

- 能力

分類，検出，識別，認識（文字認識，画像の認識，音声認識），予測，合成，翻訳，特徴抽出，ランク付け，情報推薦

- 応用

経済，金融，ロボット，医療，医学研究（遺伝子解析）など幅広い

人工知能の宿泊サイトでの応用

• 宿泊予約サイト Booking.com

情報推薦, 表示の改良, 好みの予測

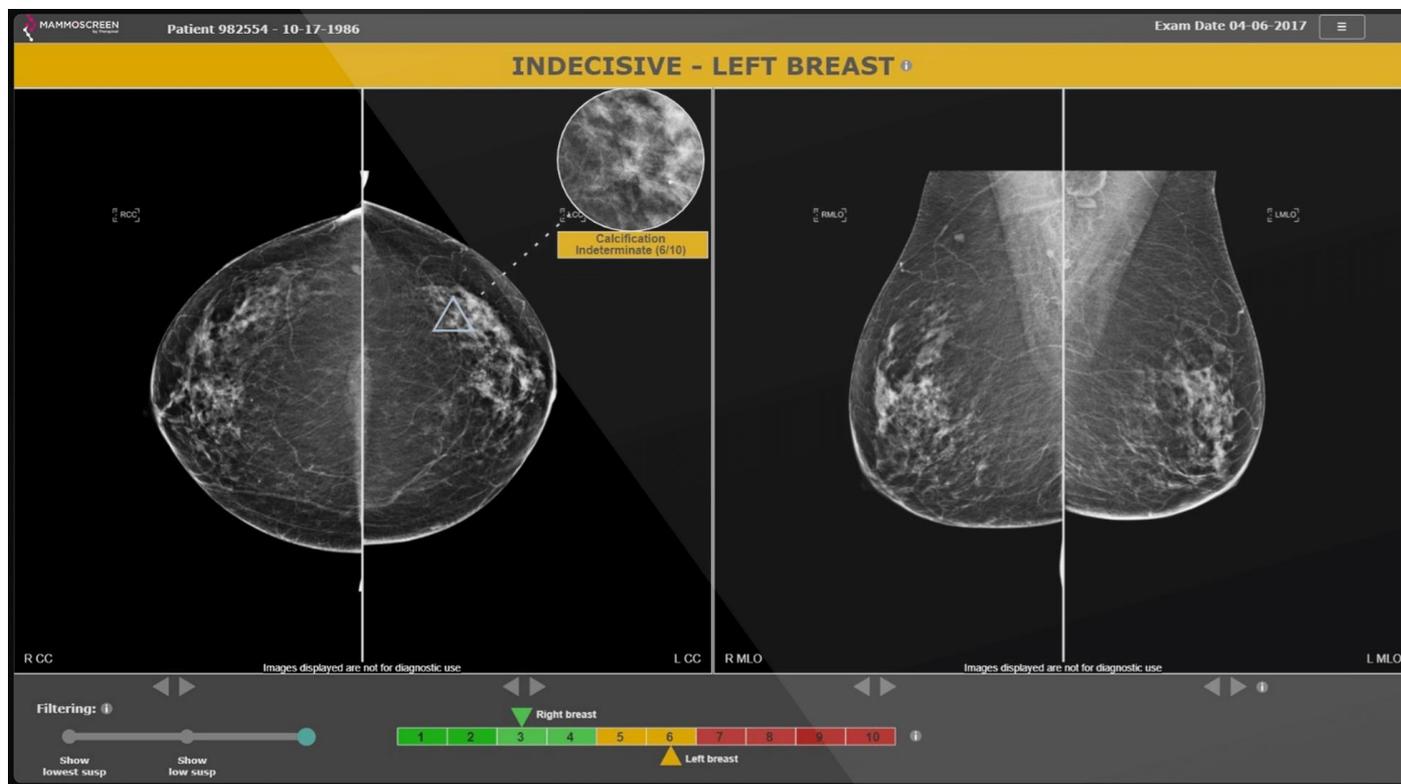
The screenshot displays the Booking.com search results page for hotels in Fukuoka. The search criteria are: 目的地 (Destination), チェックイン - チェックアウト (Check-in - Check-out), and 大人2名・子供0名・1部屋 (2 adults, 0 children, 1 room). A calendar overlay shows the selected dates for May and June 2022. Below the search bar, there are several hotel listings with their names, addresses, ratings, and prices per room per night.

ホテル名	住所	評価	料金 (1泊)
ダイワロイネットホテル福山駅前	福山市, 三之丸町	8.6	¥9,600
リッチモンドホテル福山駅前	福山市, 東桜町1-1	8.2	¥8,000
Hotel Trend Fukuyama Ekimae	福山市	7.9	¥7,500
カンデオホテルズ福山	福山市, 御船町2-8-20	7.8	¥8,500
福山プラザホテル	福山市, 住吉町1-40	7.4	¥6,800

医療での人工知能の応用のニュース

- 画像診断について，さまざまなニュースがインターネットで公開されている。

胸部のX線画像について，自分自身で，人工知能アプリを動かし，ヒントを得る



MAMMOSCREEN社のニュース

<https://www.mammoscreen.com/how-it-works>

医学での人工知能の応用のニュース

血液検査による癌の検査（人間の免疫機能によって破壊されたがん細胞の検出）に関するニュース



The screenshot shows a web browser displaying a news article from The Guardian. The browser's address bar shows the URL <https://www.theguardian.com/science/2020/mar/31/new-blood-test-can-detect-50-types-of-cancer>. The article is titled "New blood test can detect 50 types of cancer" and is categorized under "Science". A yellow banner indicates "This article is more than 2 years old". The article's sub-headline is "System uses machine learning to offer new way to screen for hard-to-detect cancers". Below the text is a photograph of two clusters of pink, spherical pancreatic cancer cells. The article text states: "Pancreatic cancer cells. The AI-based system detected 63% of stage I pancreatic cancers, and 100% at stage IV in the study. Photograph: Stocktrek Images, Inc./Alamy Stock Photo". The author's name, Nicola Davis, and their Twitter handle, @NicolaKSDavis, are also visible.

<https://www.theguardian.com/science/2020/mar/31/new-blood-test-can-detect-50-types-of-cancer>

人工知能とゲーム

能力の向上が続いている。

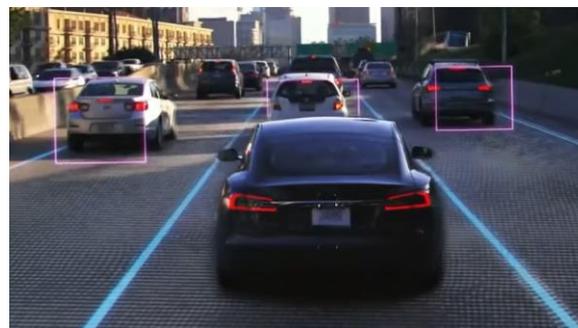
- 1997年：ディープ・ブルーが、**チェス**世界チャンピオンに勝利
- 2017年：人工知能 AlphaZero の登場。自己対戦による向上の能力を持つ
- 2017年5月：**囲碁**で、世界レベルの選手に 3戦全勝



麻雀AIについての、インターネットテレビ局「ABEMA」の番組

自動運転のニュース

- 自動運転について、さまざまなニュースがインターネットで公開されている。



6 Autonomous Vehicles & Companies to watch in 2021-2022 | Self Driving Cars

<https://www.youtube.com/watch?v=bdFi3RToOBk>

人工知能による画像の合成



さまざまな種類の画像を，安定して，高精細に生成することが可能に

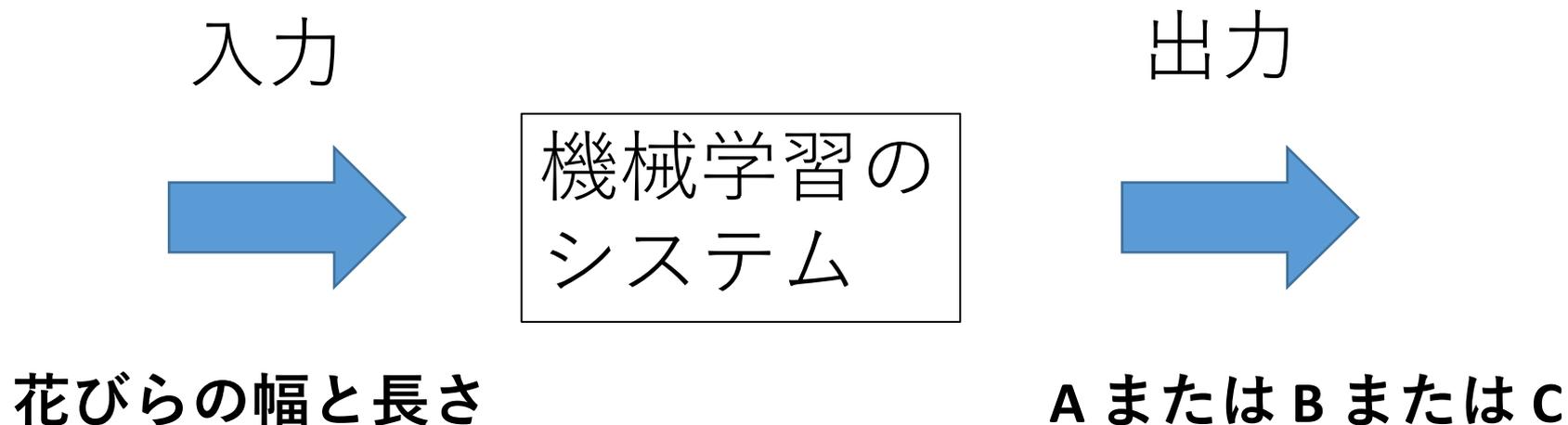
Large Scale GAN Training for High Fidelity Natural Image Synthesis

[Andrew Brock](#), [Jeff Donahue](#), [Karen Simonyan](#), <https://arxiv.org/abs/1809.11096v2>

4-3 教師あり学習と 教師なし学習

機械学習の分類の例

- ① 「花びらの幅と長さ」から、花の種類を推論したい
- ② 花の種類が A, B, C の3種類の場合は、
「花びらの幅と長さ」を入力とし、
A または B または C を出力する分類システム



機械学習のさまざまな手法

機械学習は、分類以外にもさまざまなある

教師あり学習

- ◇ 回帰
- ◇ 分類

教師なし学習

- ◇ クラスタ分析
- ◇ 次元削減
- ◇ エンコード
- ◇ 量子化 など

- 知的能力の向上
- 知識の獲得
- ルールの獲得

ポイント：**教師あり学習**と**教師なし学習**の**2種類**

教師あり学習と教師なし学習

教師あり学習

訓練データを使う

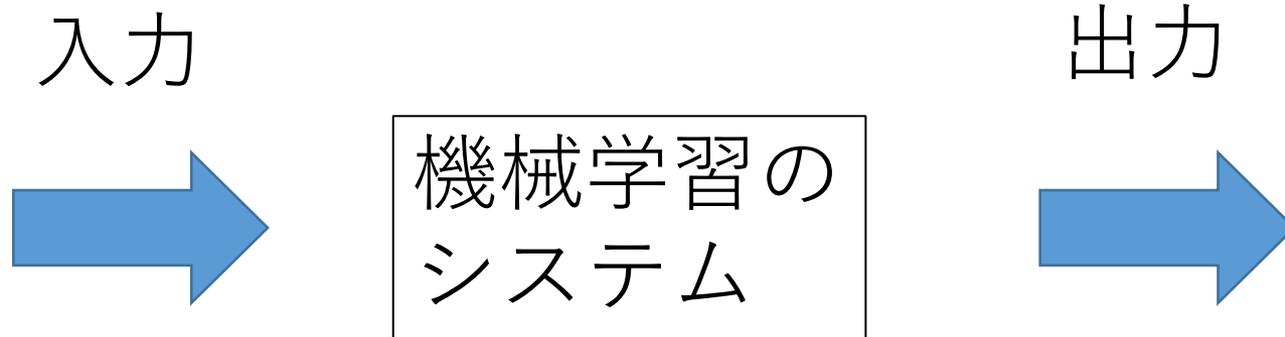
教師なし学習

訓練データは使わない。データは使う。

訓練データ

入力と出力のデータ :

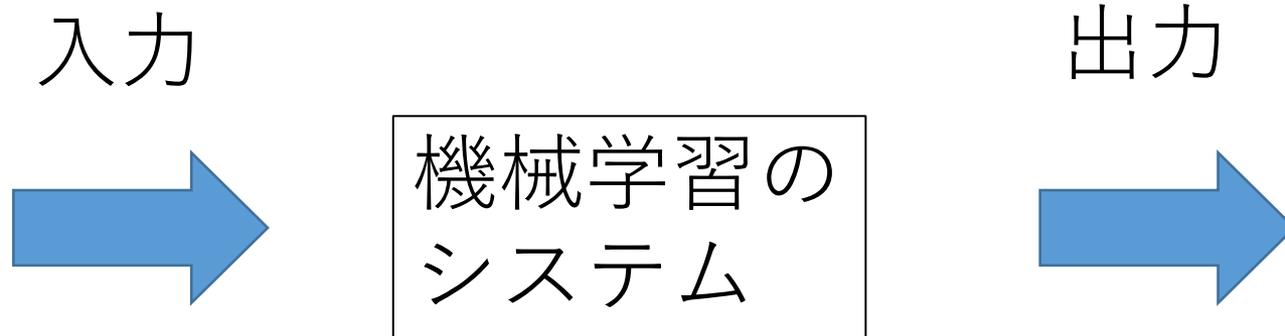
どういう入力のあるときに、**どういう出力が得られるべきかの正解のデータ**



訓練データ

入力と出力のデータ :

どういう入力のために、どういう出力が得られるべきかの正解のデータ



訓練データ

花の種類 A, B, C を推論するシステム.

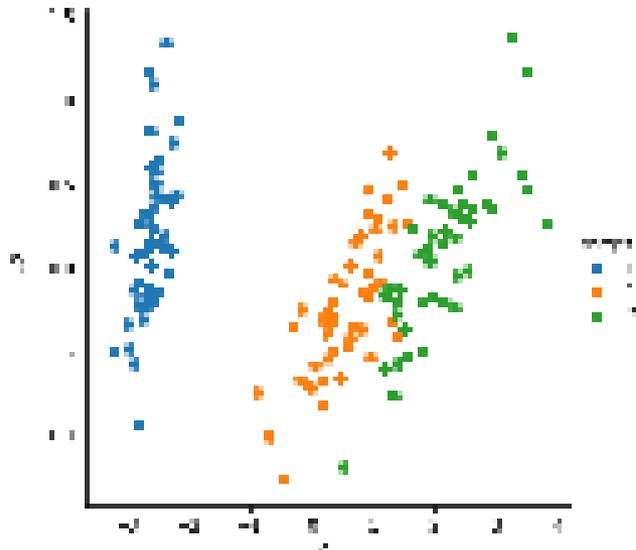
訓練データは, **入力** (花びらの幅と長さ) と **出力** (A または B または C) の両方



訓練データ

花の種類 A, B, C を推論するシステム.

訓練データは, **入力** (花びらの幅と長さ) と **出力** (A または B または C) の両方

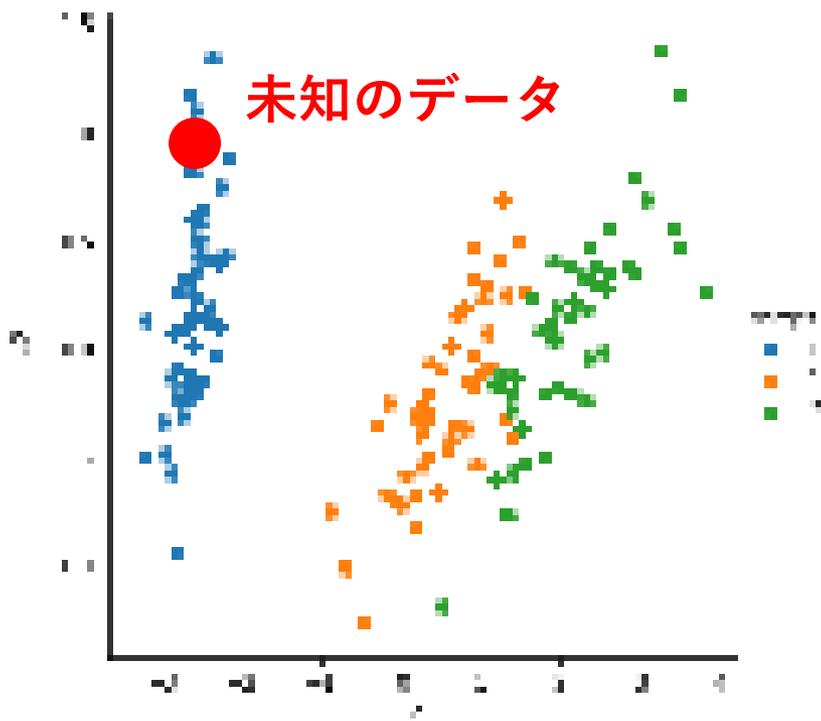


縦軸と横軸は, 花びらの幅と長さ

色 (青, オレンジ, 緑) は花の種類

教師あり学習

- **訓練データ**を使う
- 「**未知のデータを扱える能力を獲得**」と考えることもできる



この花の種類が何か？
を推論できる能力を獲得

4-4 クラスタ分析 (教師なし学習の例)

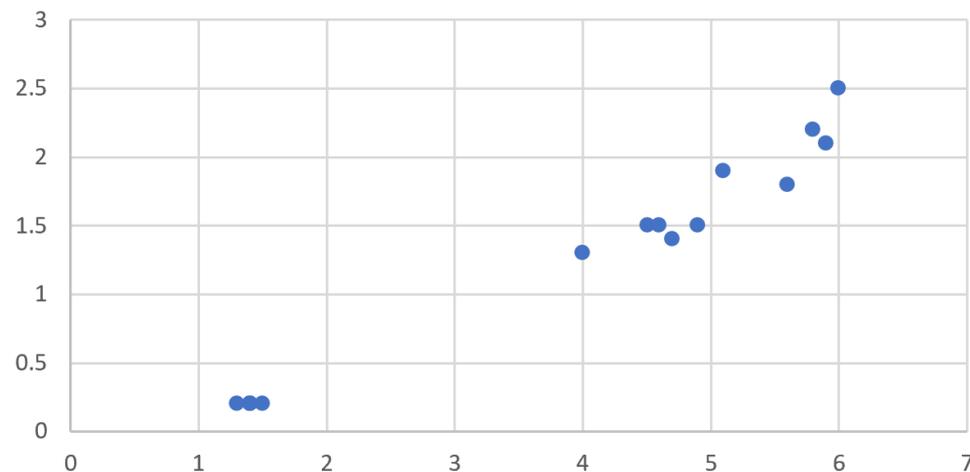
Excel の散布図

元データ

	A	B	C
1	petal_length	petal_width	species
2	1.4	0.2	setosa
3	1.4	0.2	setosa
4	1.3	0.2	setosa
5	1.5	0.2	setosa
6	1.4	0.2	setosa
7	4.7	1.4	versicolor
8	4.5	1.5	versicolor
9	4.9	1.5	versicolor
10	4	1.3	versicolor
11	4.6	1.5	versicolor
12	6	2.5	virginica
13	5.1	1.9	virginica
14	5.9	2.1	virginica
15	5.6	1.8	virginica
16	5.8	2.2	virginica



散布図



この2列で散布図

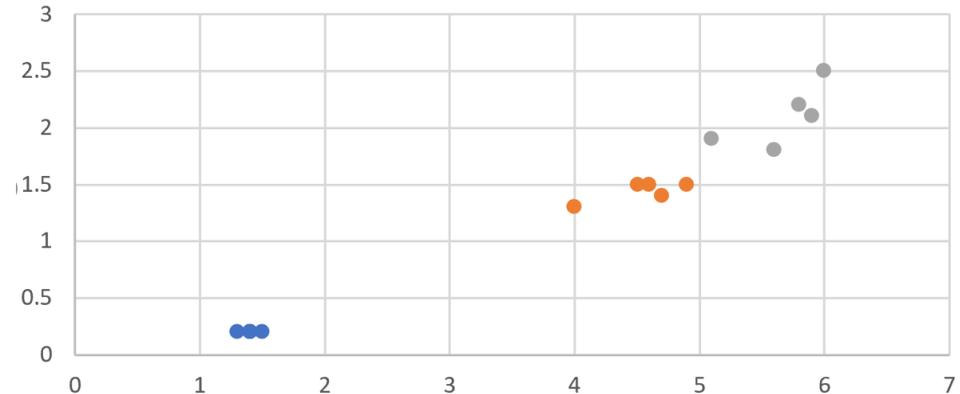
Excel の散布図 (色分け)

元データ

	A	B	C	D	E
1	petal_length	petal_width			species
2	1.4	0.2			setosa
3	1.4	0.2			setosa
4	1.3	0.2			setosa
5	1.5	0.2			setosa
6	1.4	0.2			setosa
7	4.7		1.4		versicolor
8	4.5		1.5		versicolor
9	4.9		1.5		versicolor
10	4		1.3		versicolor
11	4.6		1.5		versicolor
12	6			2.5	virginica
13	5.1			1.9	virginica
14	5.9			2.1	virginica
15	5.6			1.8	virginica
16	5.8			2.2	virginica



散布図

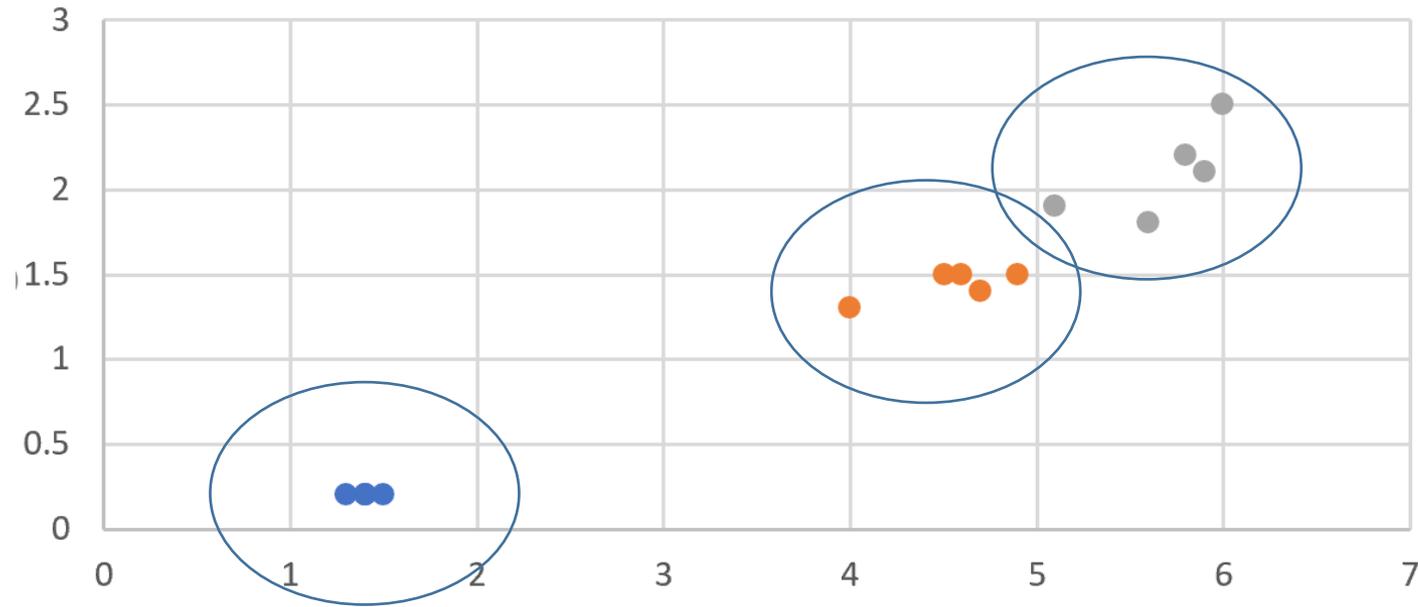


この **4 列** で散布図

setosa は B列, versicolor は C列,
virginica は D列 (違う列)

クラスタ分析

- データを使う
- データの密集（クラスタという）を見ることによる分析

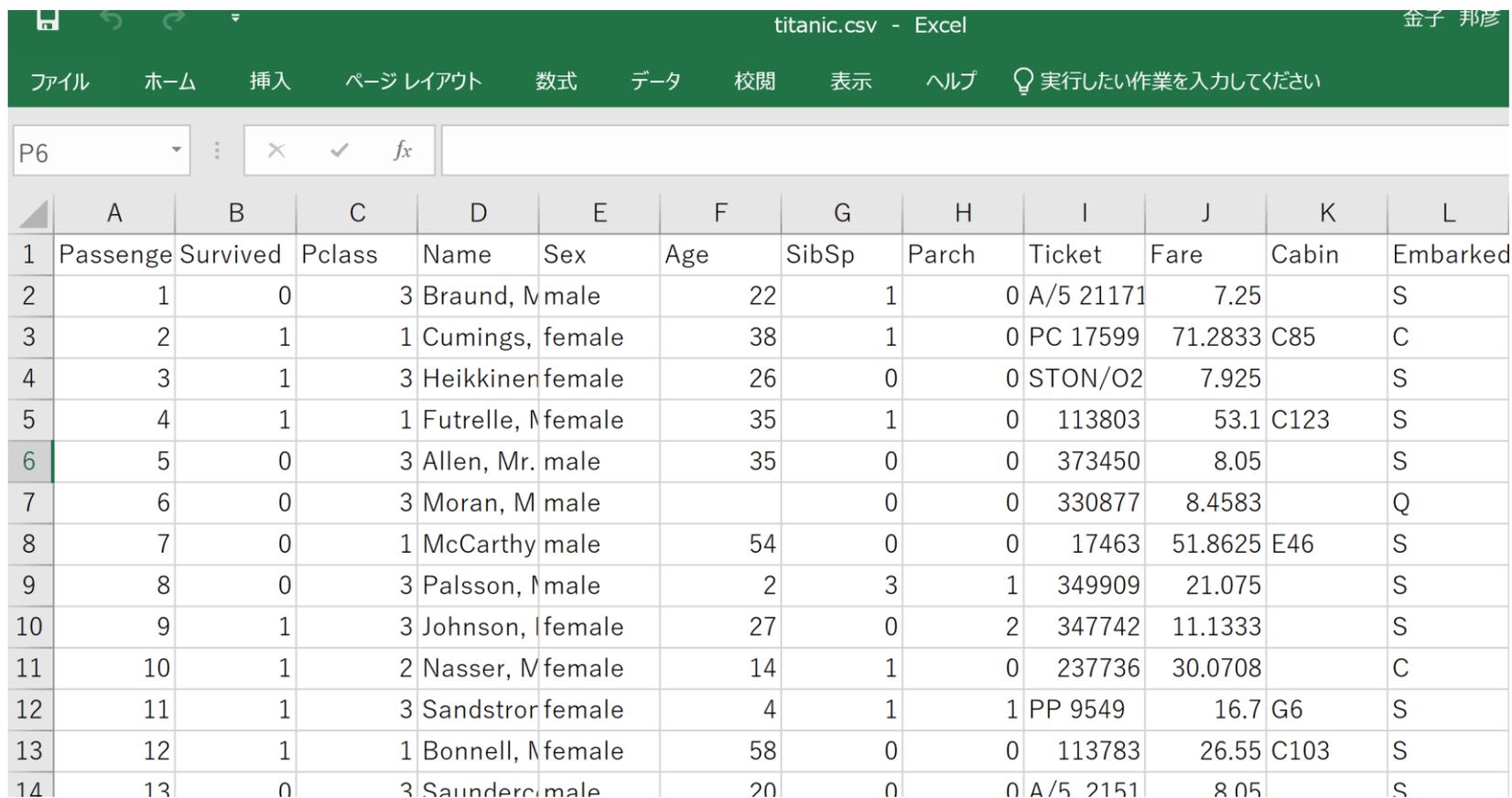


Titanic データセット

タイタニック号のデータ

救出、客室種類、性別、年齢、料金、家族の有無など

1309名分



	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenge	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cummings,	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen	female	26	0	0	STON/O2	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, M	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, I	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, M	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstror	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saunders	male	20	0	0	A/5 2151	8.05		S

① Titanic データセットから、**救出 (survived)**,
年齢 (age), **料金 (fare)** の列だけを抜き出し

	A	B	C
1	Survived	Age	Fare
2	0	22	7.25
3	1	38	71.2833
4	1	26	7.925
5	1	35	53.1
6	0	35	8.05
7	0		8.4583
8	0	54	51.8625
9	0	2	21.075
10	1	27	11.1333
11	1	14	30.0708
12	1	4	16.7
13	1	58	26.55
14	0	20	8.05
15	0	39	31.275
16	0	14	7.8542
17	1	55	18

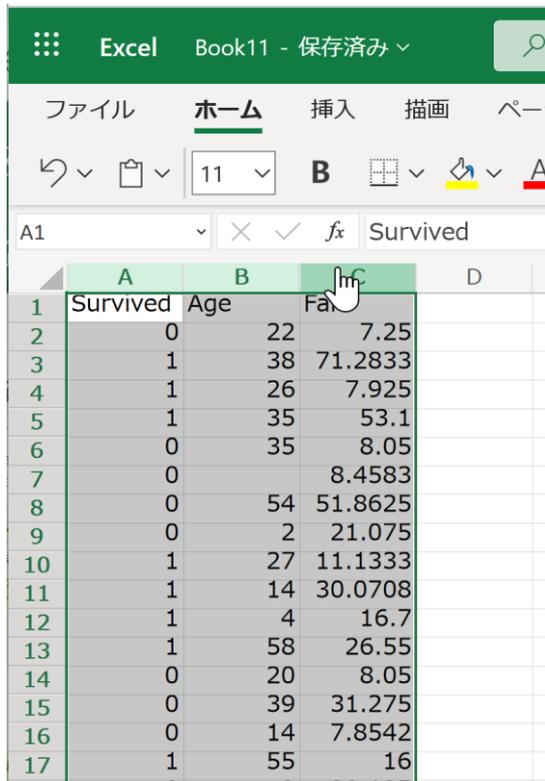
② 救出が 0 の行が**先に**，救出が 1 の行が**後**に来るように**並べ替え**

- A, B, C 列を昇順で並べ替える操作
(このとき, A 列の値を基準として, 全体を並べ替え)

オンライン版 Excel での並べ替え

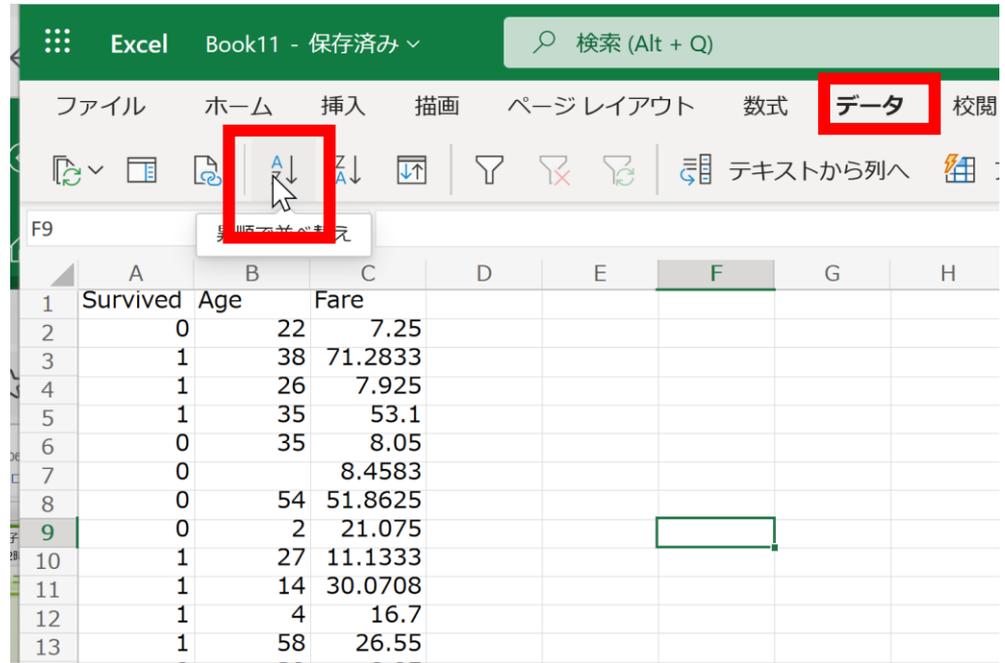
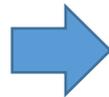
A, B, C 列を昇順で並べ替える操作

(このとき, A 列の値を基準として, 全体を並べ替え)



The screenshot shows the Excel interface with the 'ホーム' (Home) ribbon selected. The data table is visible with columns A, B, and C highlighted in green. A mouse cursor is pointing at the header of column C.

	A	B	C	D
1	Survived	Age	Fare	
2	0	22	7.25	
3	1	38	71.2833	
4	1	26	7.925	
5	1	35	53.1	
6	0	35	8.05	
7	0		8.4583	
8	0	54	51.8625	
9	0	2	21.075	
10	1	27	11.1333	
11	1	14	30.0708	
12	1	4	16.7	
13	1	58	26.55	
14	0	20	8.05	
15	0	39	31.275	
16	0	14	7.8542	
17	1	55	16	



The screenshot shows the Excel interface with the 'データ' (Data) ribbon selected. The '昇順で並べ替え' (Sort Ascending) button is highlighted with a red box. The data table is visible, showing the result of the sorting operation.

	A	B	C	D	E	F	G	H
1	Survived	Age	Fare					
2	0	22	7.25					
3	1	38	71.2833					
4	1	26	7.925					
5	1	35	53.1					
6	0	35	8.05					
7	0		8.4583					
8	0	54	51.8625					
9	0	2	21.075					
10	1	27	11.1333					
11	1	14	30.0708					
12	1	4	16.7					
13	1	58	26.55					

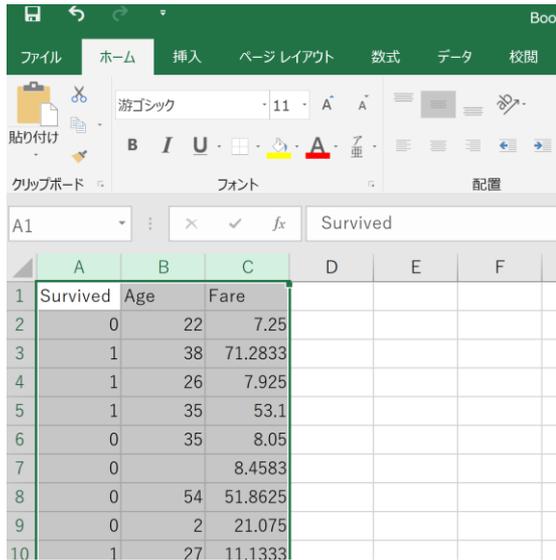
① 範囲選択

② リボンで「データ」 → 「昇順で並べ替え」

アプリ版での並べ替え

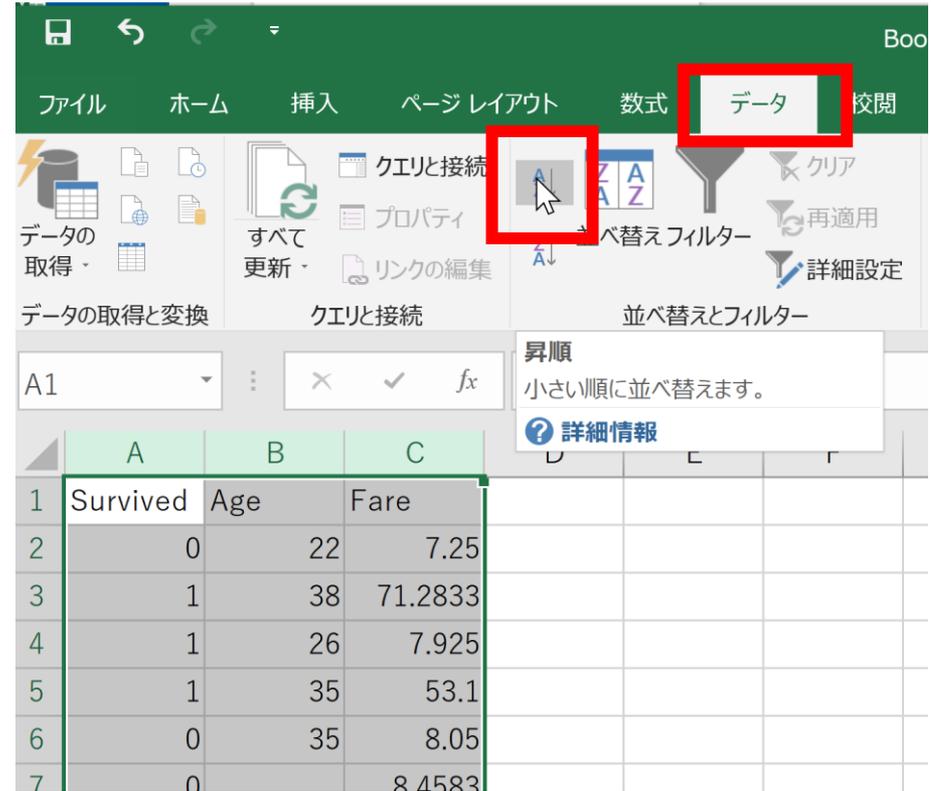
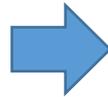
A, B, C 列を昇順で並べ替える操作

(このとき, A 列の値を基準として, 全体を並べ替え)



The screenshot shows the Excel app interface with the 'データ' (Data) ribbon selected. The data table is as follows:

	A	B	C	D	E	F
1	Survived	Age	Fare			
2	0	22	7.25			
3	1	38	71.2833			
4	1	26	7.925			
5	1	35	53.1			
6	0	35	8.05			
7	0		8.4583			
8	0	54	51.8625			
9	0	2	21.075			
10	1	27	11.1333			



The screenshot shows the Excel app interface with the 'データ' (Data) ribbon selected. The '昇順' (Ascending) option is highlighted in the '並べ替えとフィルター' (Sort & Filter) group. The data table is as follows:

	A	B	C	D	E	F
1	Survived	Age	Fare			
2	0	22	7.25			
3	1	38	71.2833			
4	1	26	7.925			
5	1	35	53.1			
6	0	35	8.05			
7	0		8.4583			

① 範囲選択

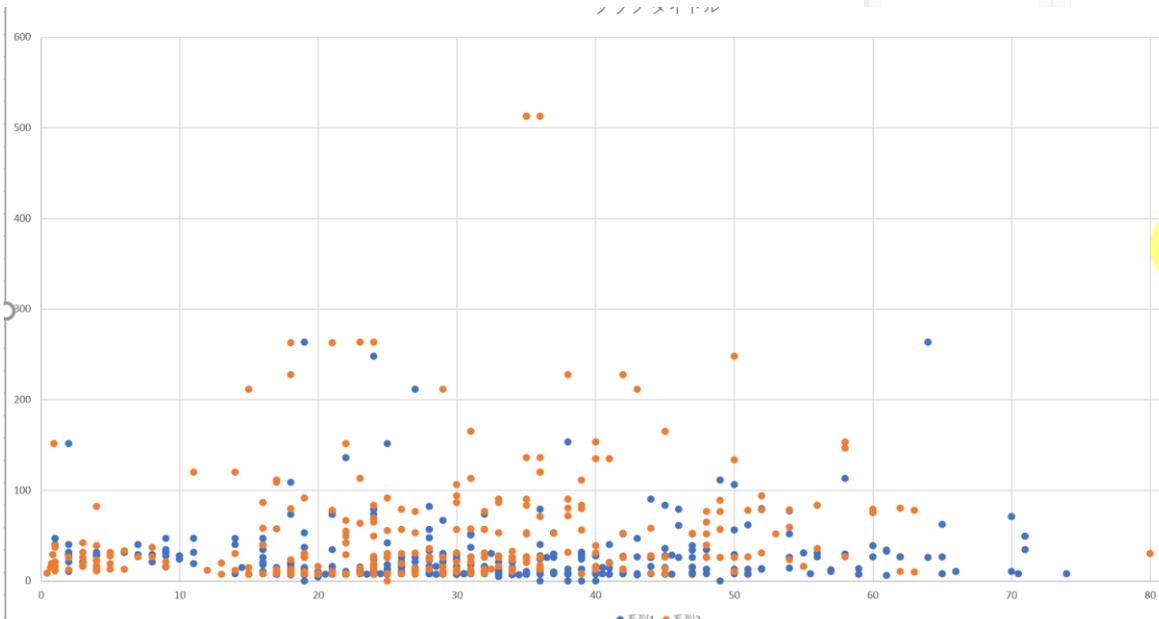
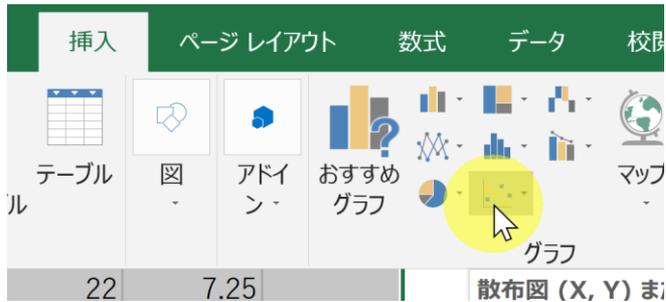
② リボンで「データ」→「昇順」

③ 救出が 1 の行についてのみ, C 列のデータを D 列に移す

(あとで, 散布図の色分けをしたい)

543	0	33	7.8958		
544	0	22	10.5167		
545	0	28	10.5		
546	0	25	7.05		
547	0	39	29.125		
548	0	27	13		
549	0		23.45		
550	0	32	7.75		
551	1	38		71.2833	
552	1	26		7.925	
553	1	35		53.1	
554	1	27		11.1333	
555	1	14		30.0708	
556	1	4		16.7	
557	1	58		26.55	
558	1	55		16	
559	1			13	

④ B, C, D 列から散布図を作成

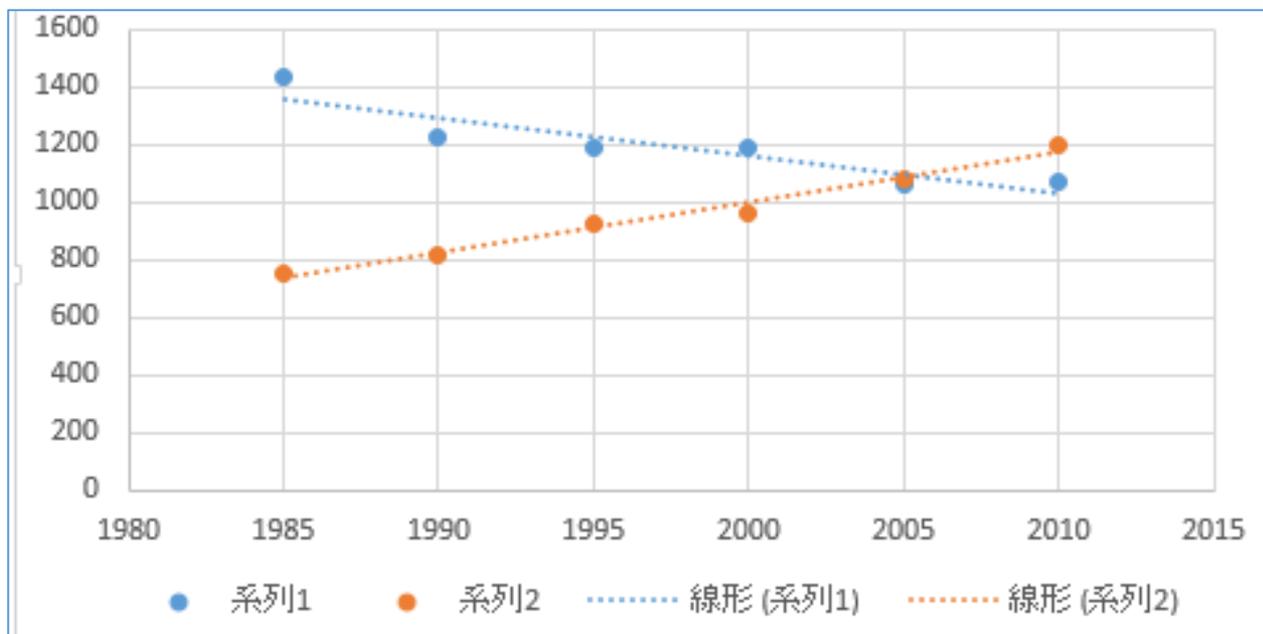


オレンジは上側に、
青は下側に偏る傾向
が分かる

4-5 線形近似 (教師あり学習の例)

線形近似

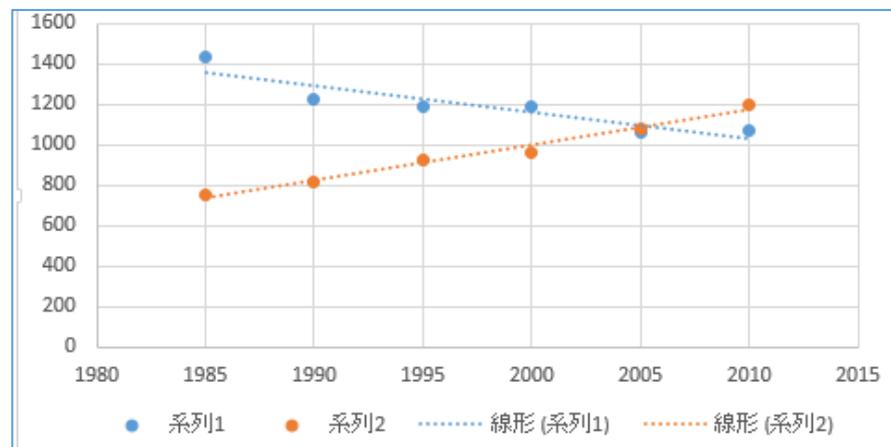
- 教師あり学習
- **学習のため，訓練データを使う**
- **未知の入力に対する出力を推論できる能力を獲得**



散布図 + 線形近似

Excel での線形近似

年次	出生数 (千人)	死亡数 (千人)
1985	1432	752
1990	1222	820
1995	1187	922
2000	1191	962
2005	1063	1084
2010	1071	1197



散布図 + 線形近似

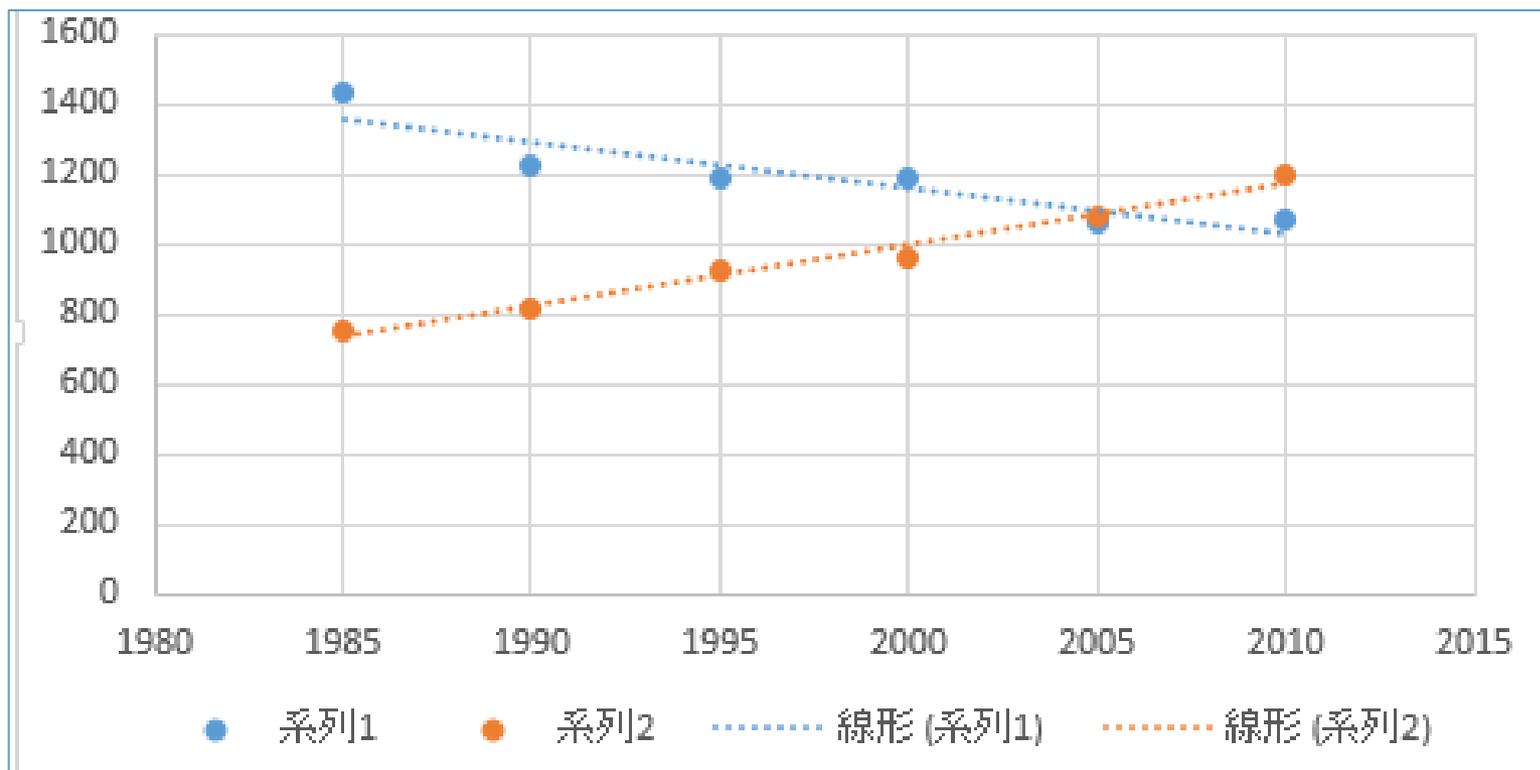
入力

出力

出生数, 死亡数の推移

出典: 総務省「第63回 日本統計年鑑 平成26年」

線形近似の例



線形近似とは、データの分布を近似した補助線

Excelでの散布図 + 線形近似の作成手順 (1 / 3)

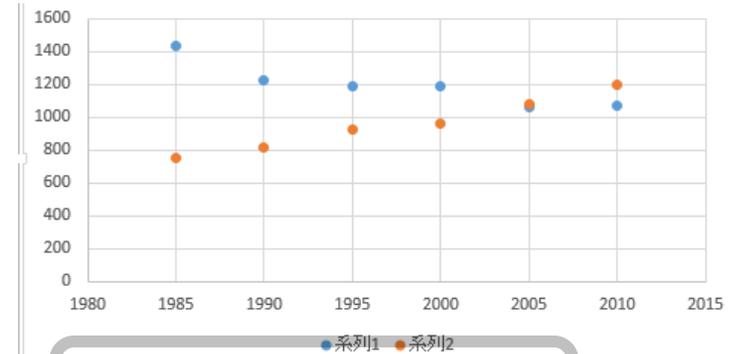
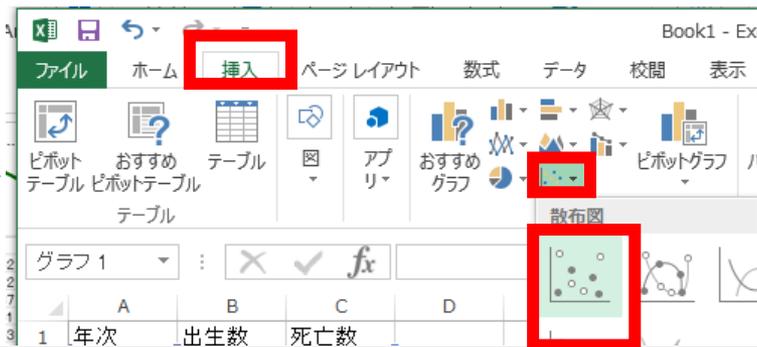
	A	B	C
1	年次	出生数	死亡数
2	1985	1432	752
3	1990	1222	820
4	1995	1187	922
5	2000	1191	962
6	2005	1063	1084
7	2010	1071	1197

元データ



	A	B	C
1	年次	出生数	死亡数
2	1985	1432	752
3	1990	1222	820
4	1995	1187	922
5	2000	1191	962
6	2005	1063	1084
7	2010	1071	1197

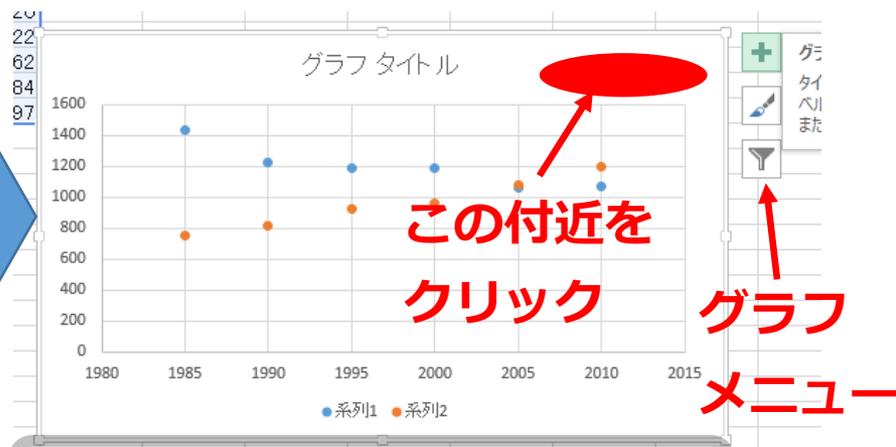
① データ部分を範囲選択



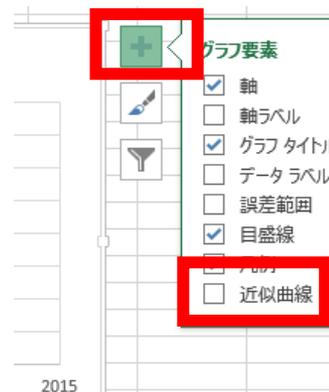
③ 散布図を確認

② リボンで「挿入」→散布図の選択

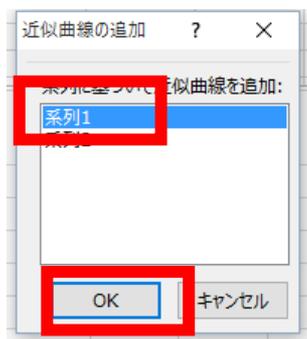
Excel での散布図 + 線形近似の作成手順 (2 / 3)



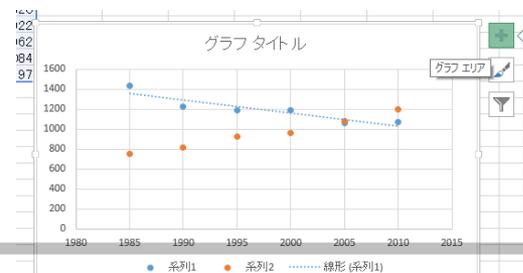
④ グラフメニューを出す



⑤ 「+」のボタンをクリックしたのちに、「近似曲線」をクリック

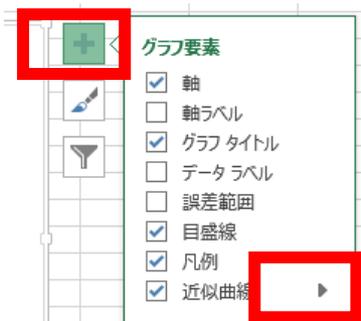


⑥ 「系列1」をクリックして「OK」

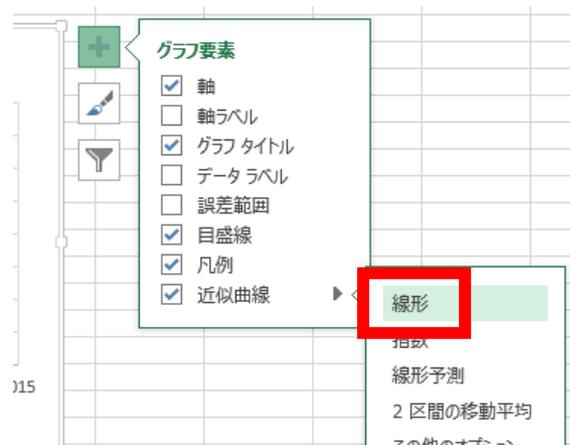


⑦ 補助線が1つ出るので確認

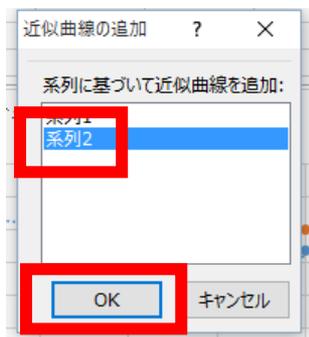
Excel での散布図 + 線形近似の作成手順 (3 / 3)



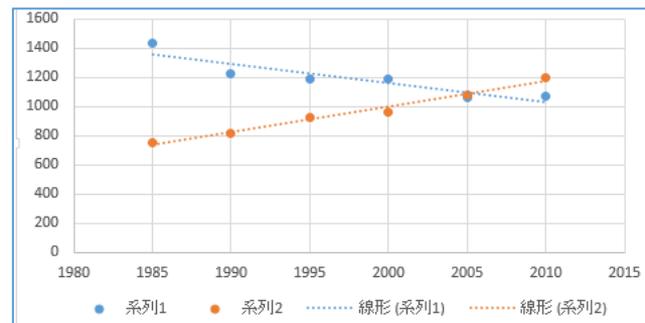
⑧ もう1度、**グラフメニュー**を出し、**近似曲線**の右横の「▶」を展開



⑨ 「**線形**」を選ぶ



⑩ 「**系列2**」をクリックして「**OK**」



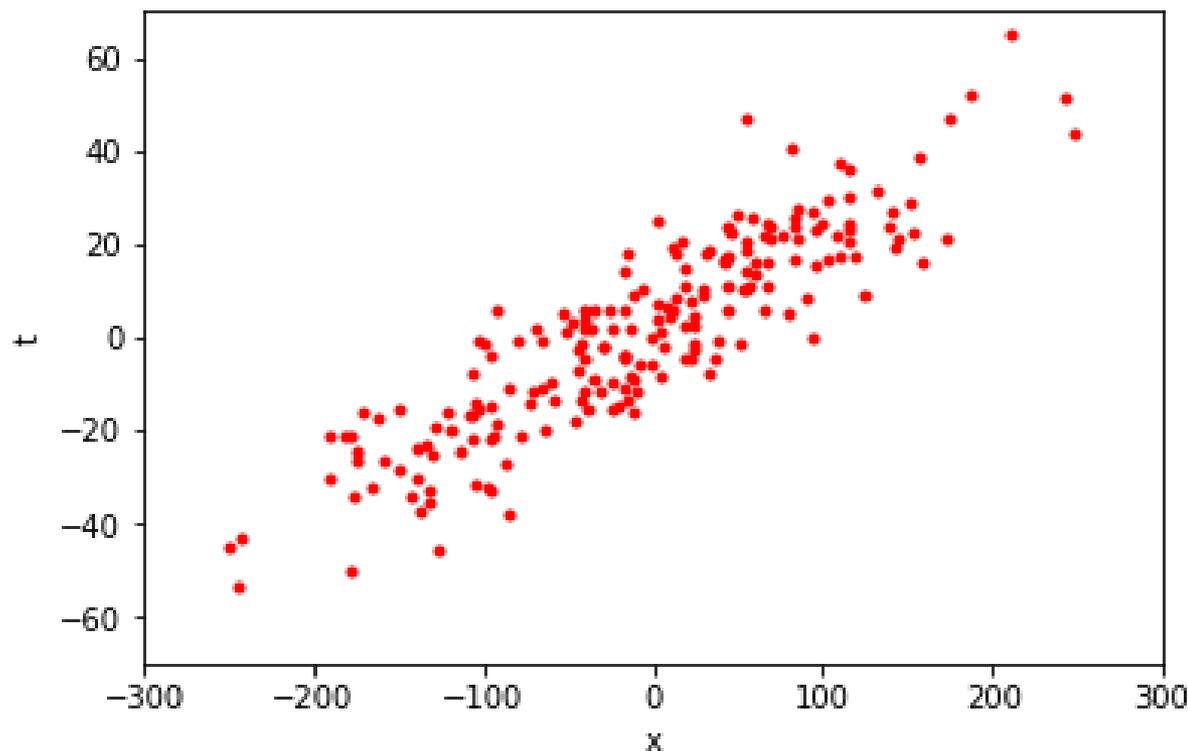
完成

4-6 最適化

最適化

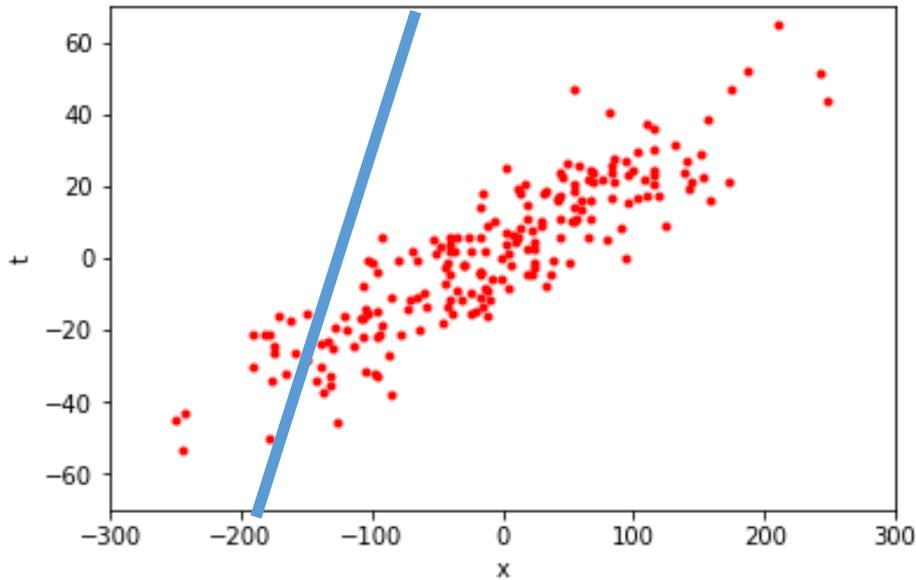
- **最適化**は、
パラメータを調整して、
ある尺度での**値**を**最適**にするように、
調整を行うこと
- **誤差を自動で最小化したい**ときに有効な技術

データの例

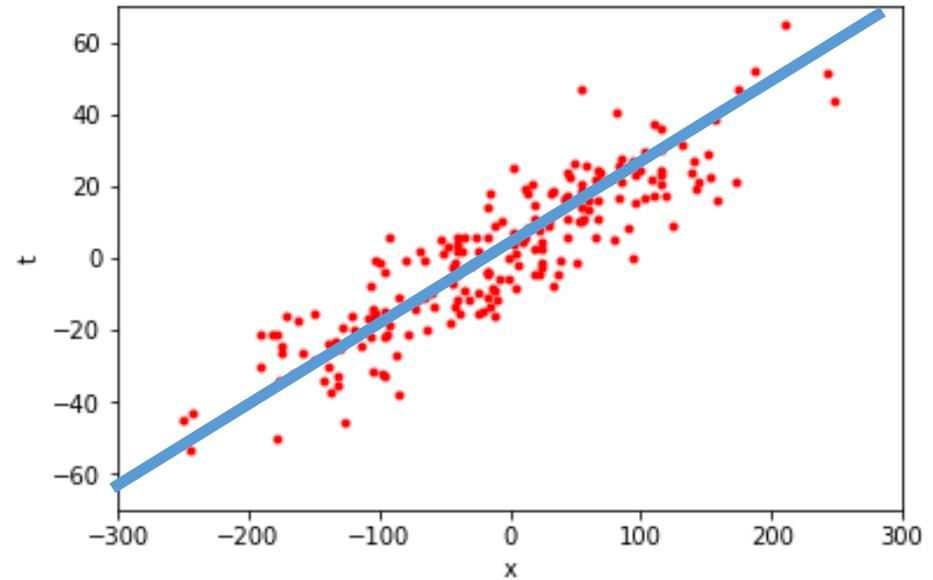


- 多数のデータの集まり
- 上の図では, 点1つで, 1つのデータ

直線による近似の例

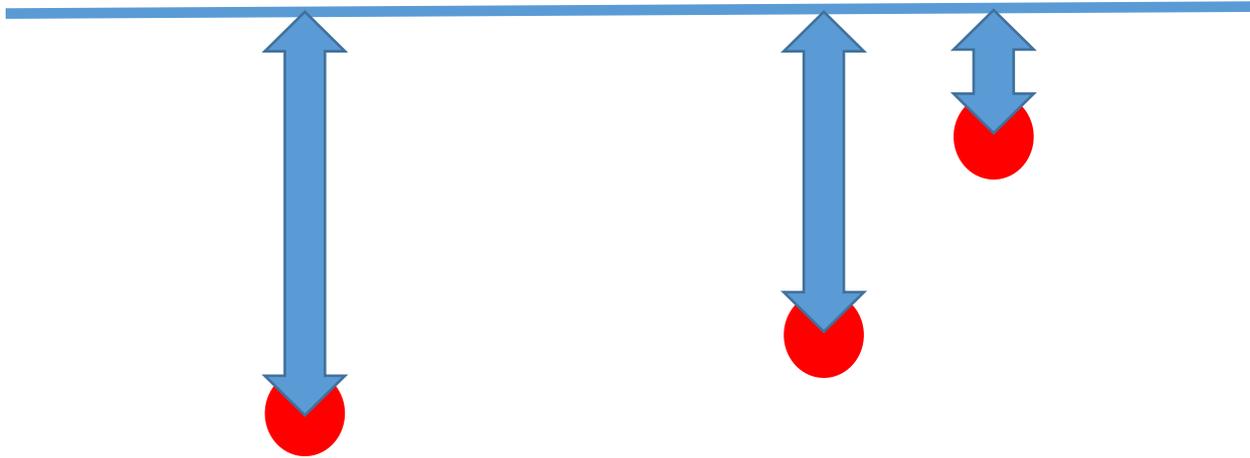


最適でない近似
(誤差大)



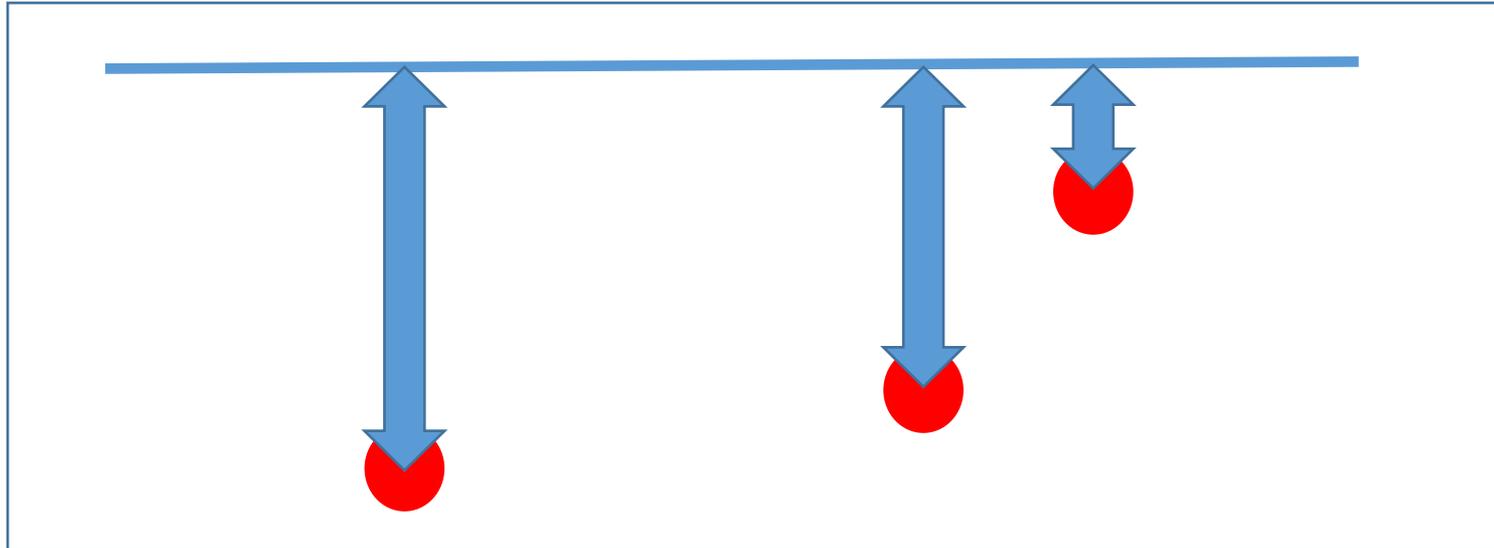
最適な近似
(誤差小)

誤差

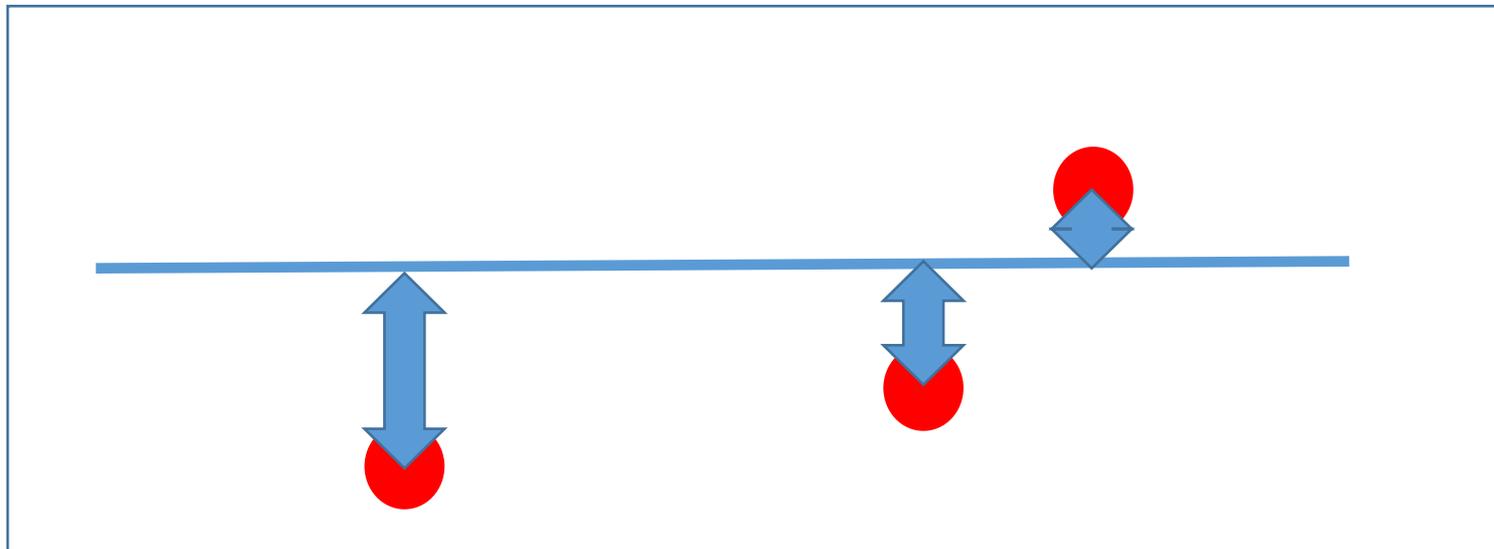


赤点：元データ

直線の上下移動による誤差の変化



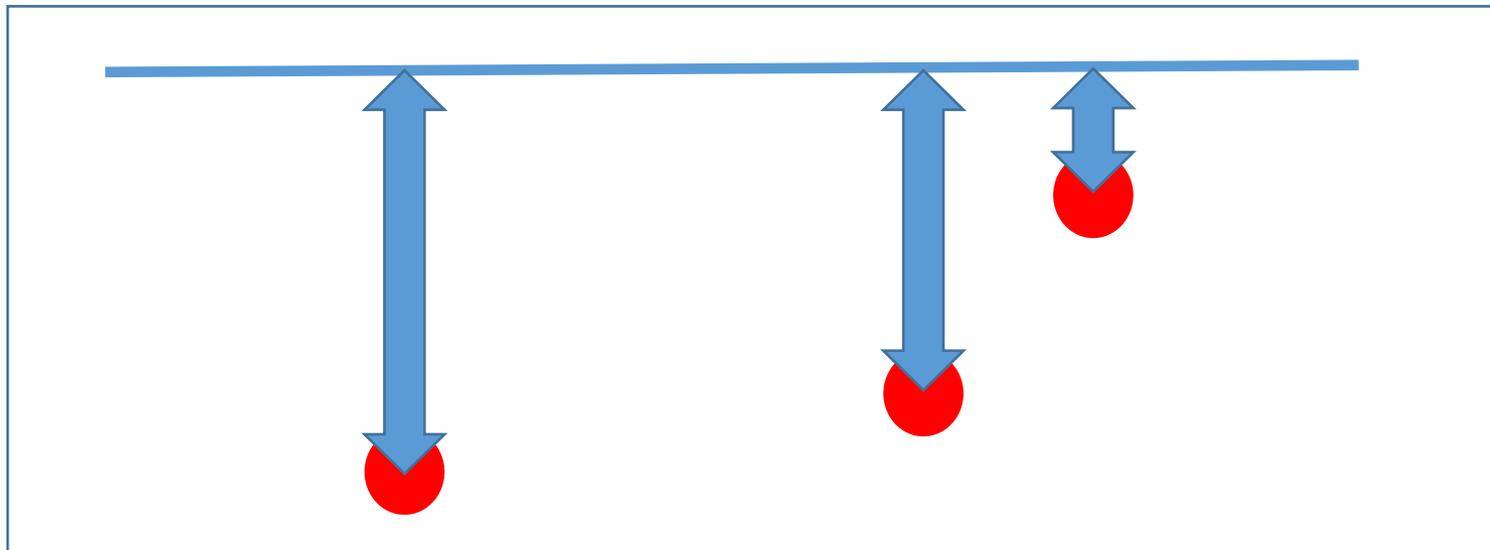
誤差大



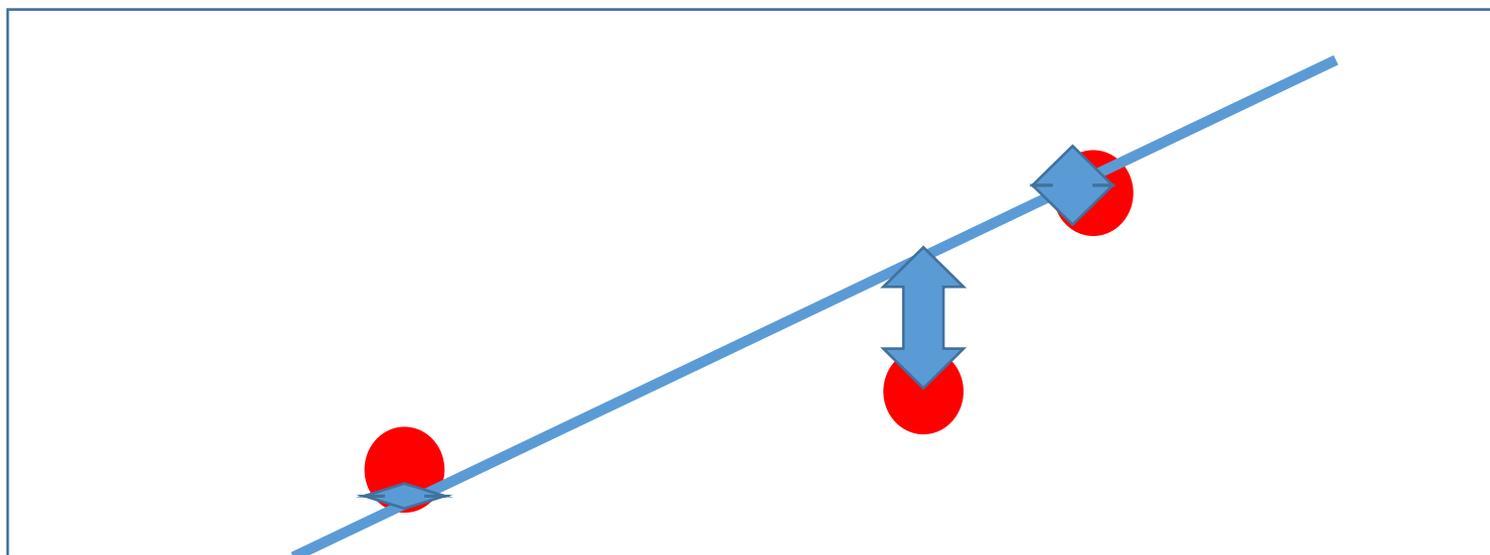
誤差小

赤点：元データ

直線の傾きの変化による誤差の変化



誤差大

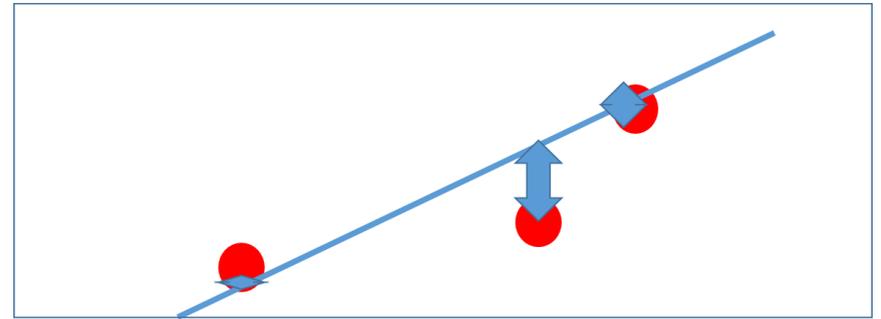
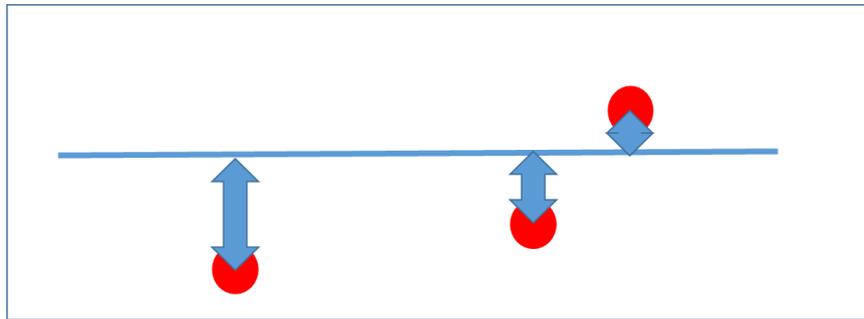
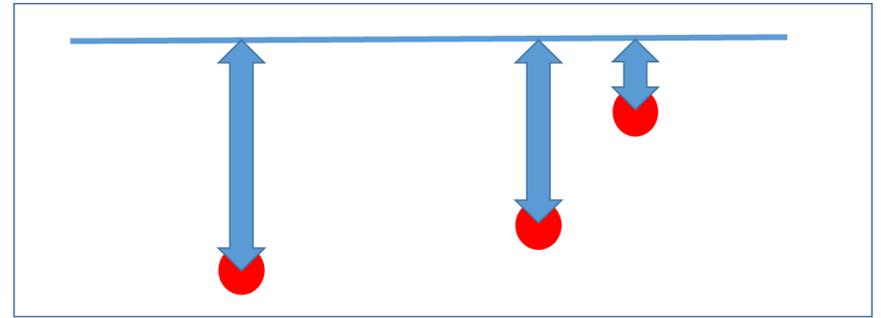
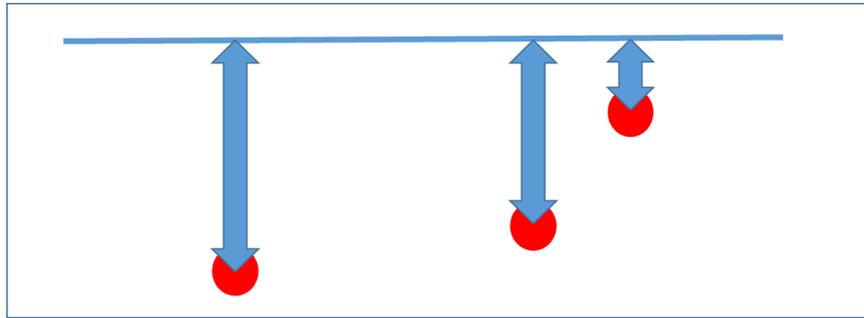


誤差小

赤点：元データ

誤差の変化

- 直線の上下移動や，傾きの変化により，**誤差**が変化



最適化

- **最適化**は、
パラメータを調整して、
ある尺度での**値**を**最適**にするように、
調整を行うこと

ゴール： 誤差の最小化

パラメータ： 直線の上下の位置と、
直線の傾き

誤差を自動で最小化したいときに有効な技術

4-7 最適化の用途

最適化の用途

- ① 方程式を解く
- ② 教師あり学習を行う

① 最適化により方程式を解く

- 次の式の値が最小になるように, x の値を定めたい. 但し, $N = 5$ とする.

$$f(\mathbf{x}) = \sum_{i=2}^N 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2.$$

- <https://docs.scipy.org/doc/scipy/reference/tutorial/optimize.html>

① 最適化により方程式を解く Python プログラム

```
import numpy as np
from scipy.optimize import minimize
def rosen(x):
    """The Rosenbrock function"""
    return sum(100.0*(x[1:]-x[:-1]**2.0)**2.0 + (1-x[:-1])**2.0)

x0 = np.array([1.3, 0.7, 0.8, 1.9, 1.2])
res = minimize(rosen, x0, method='nelder-mead',
               options={'xtol': 1e-8, 'disp': True})
print(res.x)
```

① 最適化により方程式を解く Python プログラム

```
▶ import numpy as np
from scipy.optimize import minimize
def rosen(x):
    """The Rosenbrock function"""
    return sum(100.0*(x[1:]-x[:-1])**2.0)**2.0 + (1-x[:-1])**2.0

x0 = np.array([1.3, 0.7, 0.8, 1.9, 1.2])
res = minimize(rosen, x0, method='nelder-mead',
               options={'xtol': 1e-8, 'disp': True})
print(res.x)
```

```
↳ Optimization terminated successfully.
      Current function value: 0.000000
      Iterations: 339
      Function evaluations: 571
[1. 1. 1. 1. 1.]
```

$x = [1\ 1\ 1\ 1\ 1]$ のとき（すべての値が 1 のとき）
最適であると求まった。

② 線形近似での最適化

線形近似は「誤差が最小になるような直線」と考えることもできる

