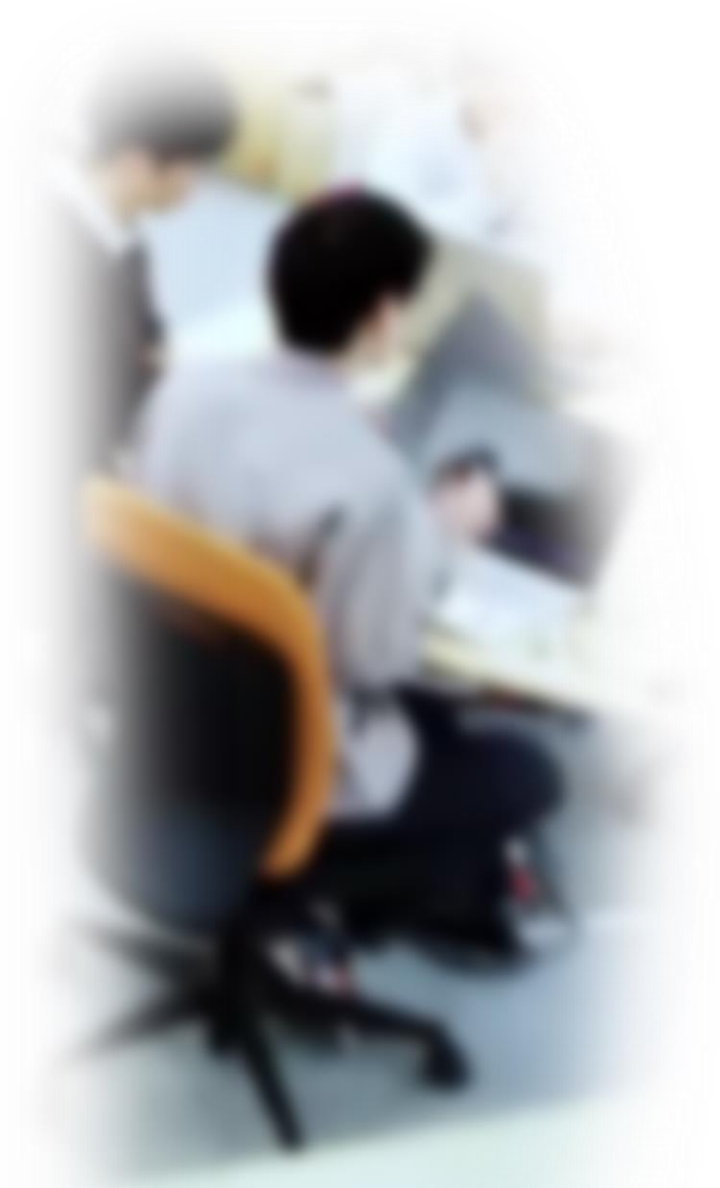
4. 機械学習

（人工知能）

URL: https://www.kkaneko.jp/ai/mi/index.html

金子邦彦

CC BY NC SA



アウトライン

1. 機械学習
2. 機械学習の現状
3. ニューラルネットワークの概要
4. ニューラルネットワークの種類，応用分野
5. 線形近似（学習の例）
6. 最適化

今日の授業では，**Office 365 のインストール**を済ませておくことが便利です．

4-1 機械学習

機械学習

機械学習は，**コンピュータ**が**データ**を使用して**学習**することより**知的能力を向上**させる技術

- **情報の抽出**：パターンや関係性を見つけ出す能力
- **簡潔さ**：人間によるルールの作成や煩雑な計算を省くことができる
- **限界の超越**：他の方法では解決が難しかった問題が解決できる可能性

人工知能の種類

人工知能
（知的なITシステム）

機械学習

コンピュータが**データ**を使用して**学習**することより**知的能力を向上**させる技術

ルール，知識による人工知能

人間が**ルール，知識**を**コンピュータに与える**ことによる人工知能

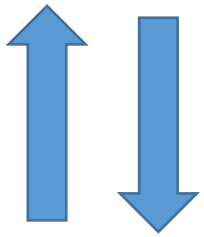
機械学習のメリット，デメリット

機械学習

コンピュータが**データ**を使用して**学習**することより**知的能力を向上**させる技術

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

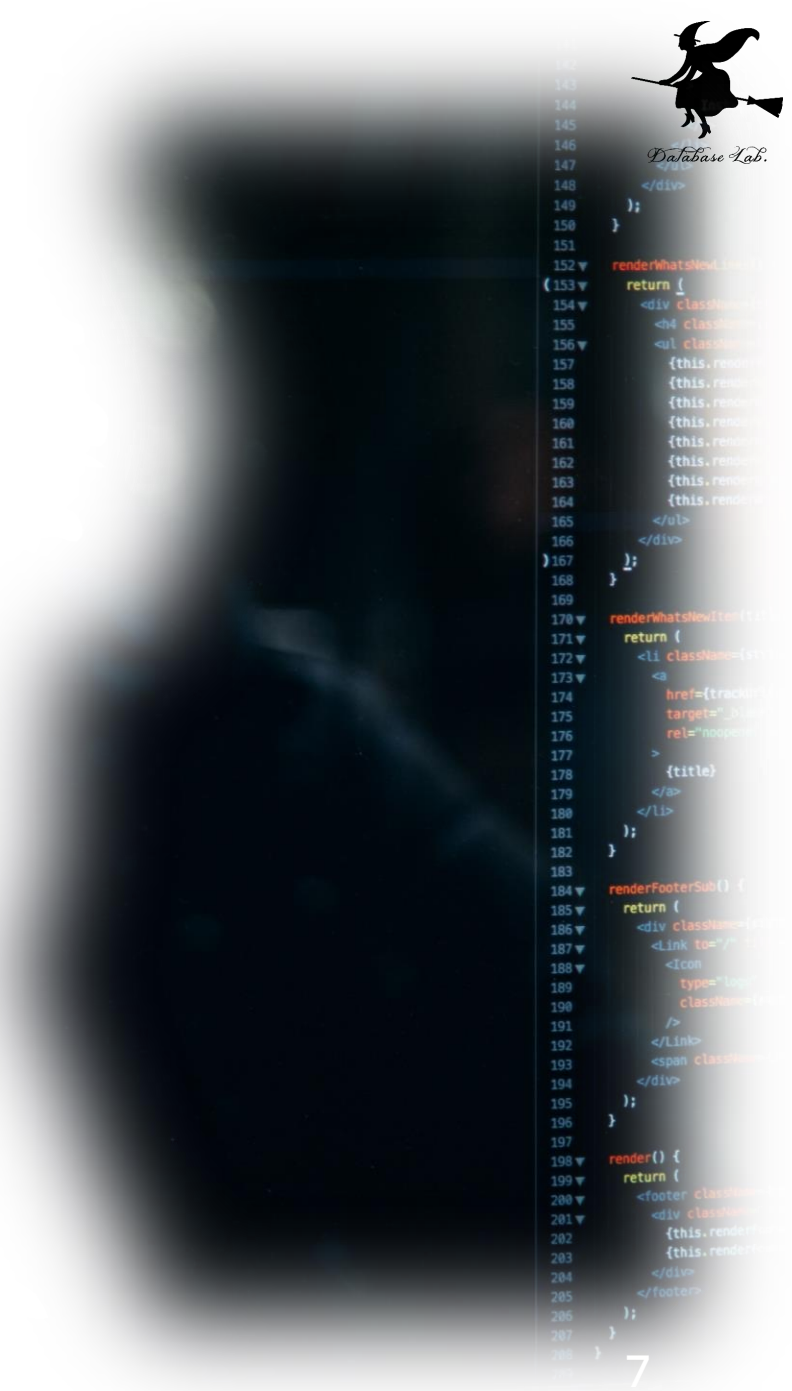
学習不足，過学習，などの注意点がある．完璧に学習が成功するわけではない．

ルール，知識による人工知能

人間が**ルール，知識**を**コンピュータに与える**ことによる人工知能

機械学習のメリット

- **ルールや知識のプログラム化不要**
- **データの追加で，知的能力が向上**する
- 人間が手作業でルールや知識を作成するよりも効率的で，人間の主観を排除できるという考え方も



機械学習の特徴まとめ

- **学習による上達：データを使用し，学習を通じて知的能力を向上.**
- **ルールや知識のプログラム化不要**：機械学習では**自動でパターンや関連性を見つけ出す**.
- **限界の超越**：機械学習は**他の方法では解決が難しい問題にも取り組む**ことができる.

4-2 機械学習の現状

機械学習の特徴と応用

- **能力の多様性:**

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、情報推薦など、**さまざまな能力を持つ**

- **応用範囲の広さ:**

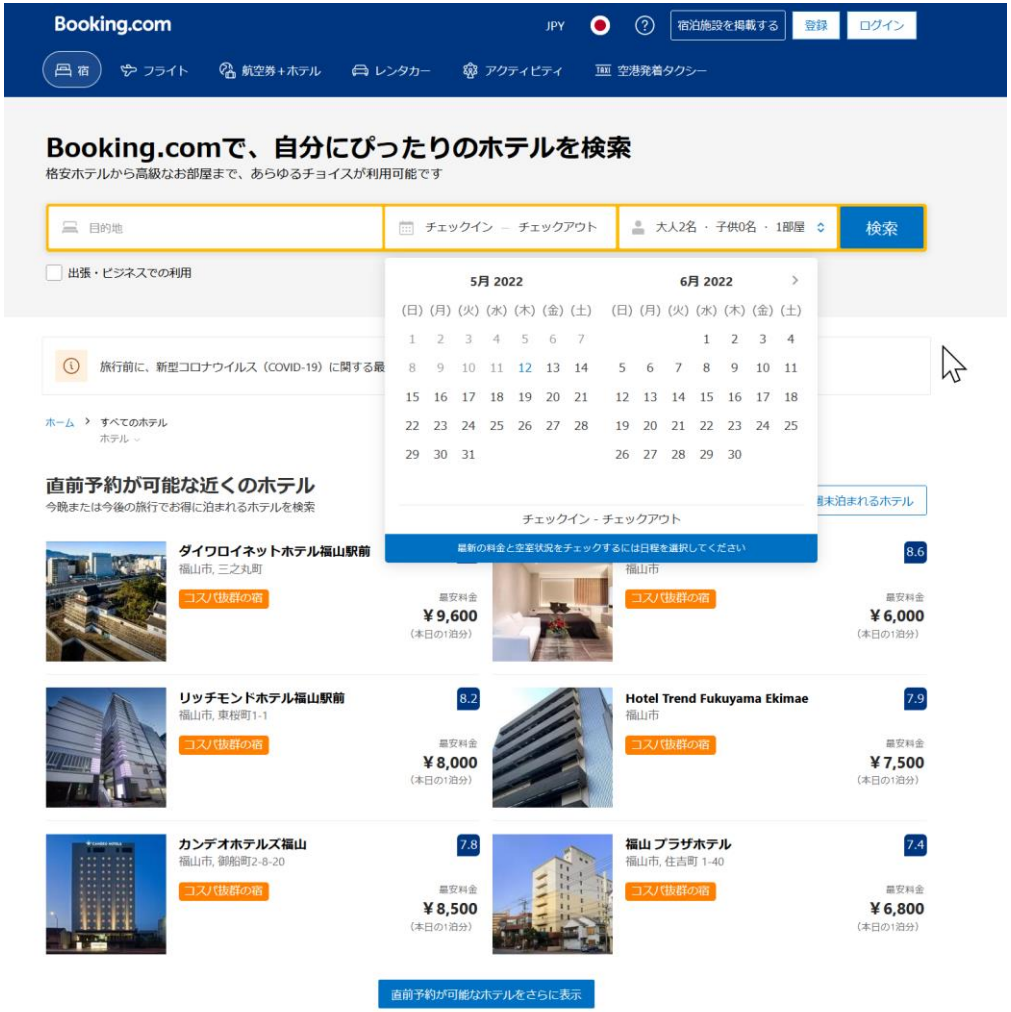
経済、金融、ロボット、医療、医学研究（遺伝子解析）など、様々な分野で問題解決に活用

機械学習は多様な能力を持ち、経済や金融から医療やロボットまで広範な応用が可能

人工知能の宿泊サイトでの応用

- 宿泊予約サイト Booking.com

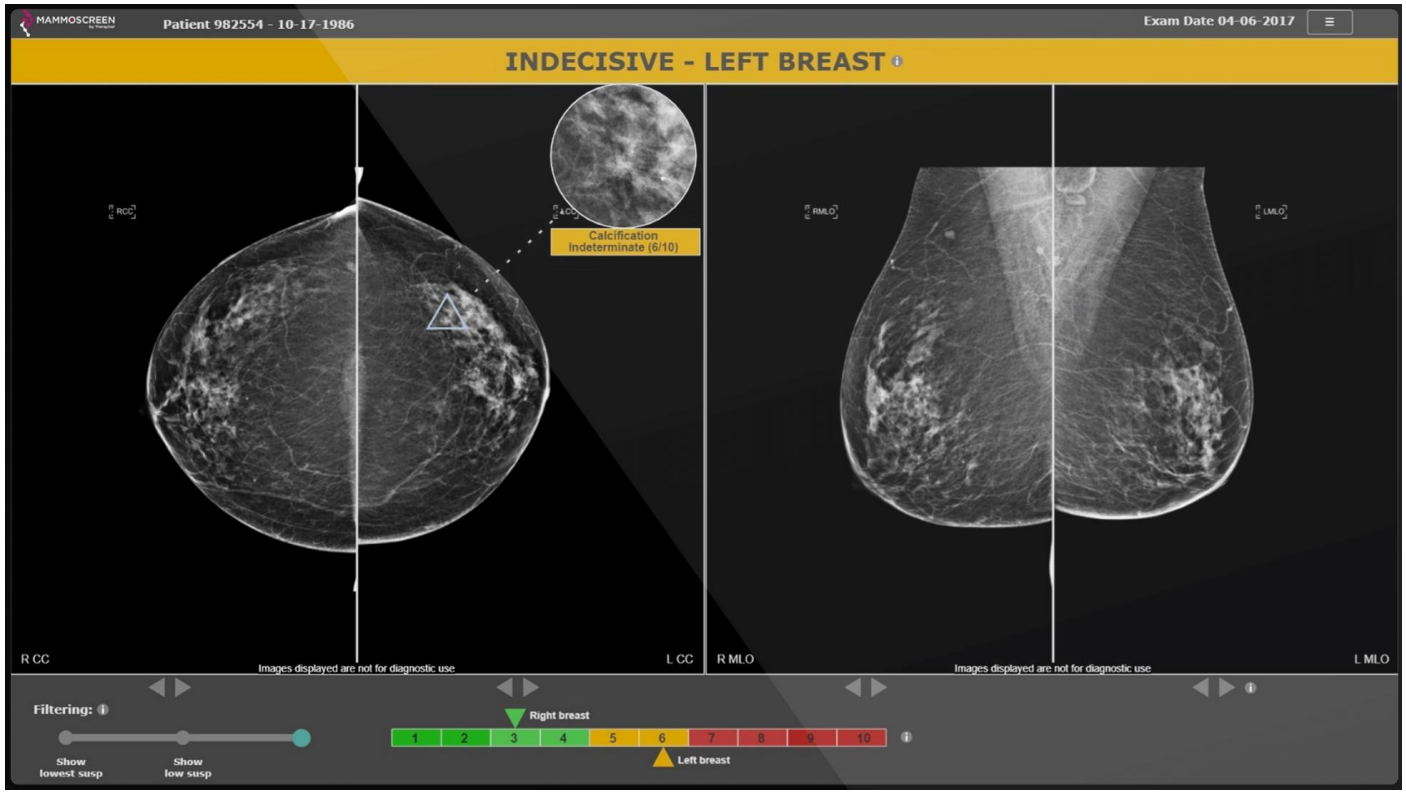
情報推薦，表示の改良，好みの予測



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

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

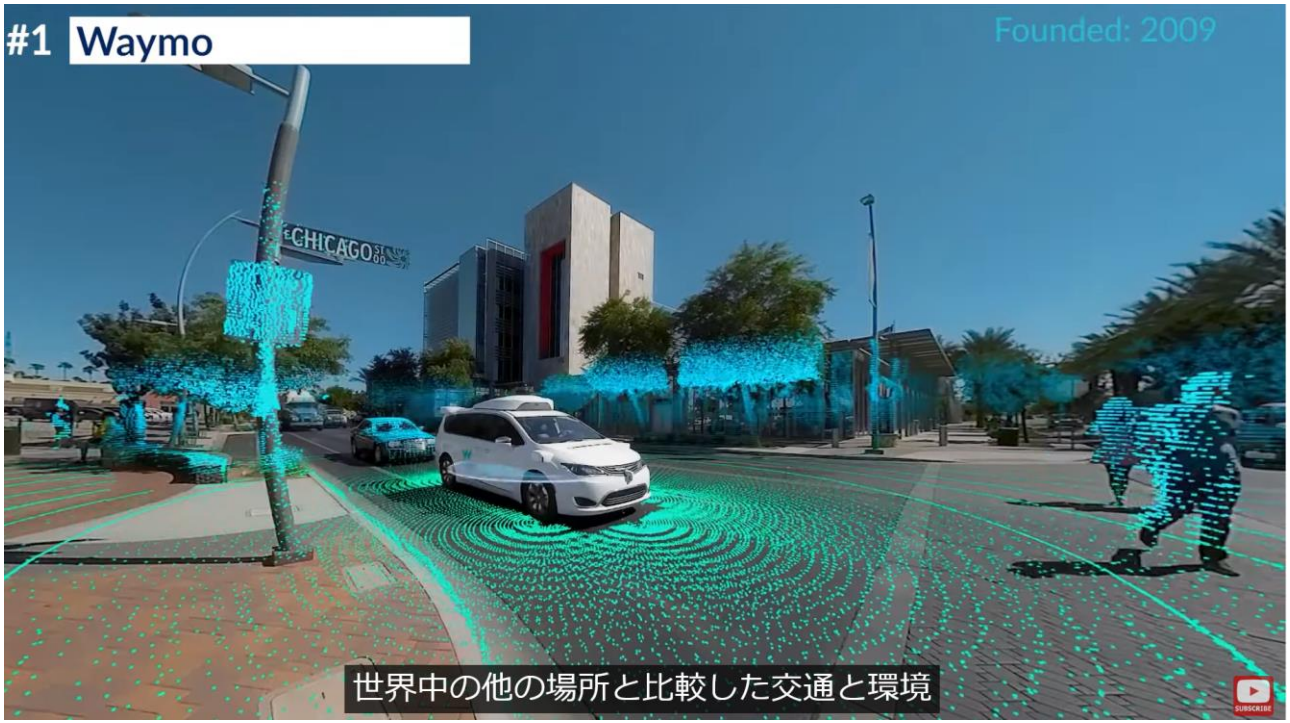
能力の向上が続いている.

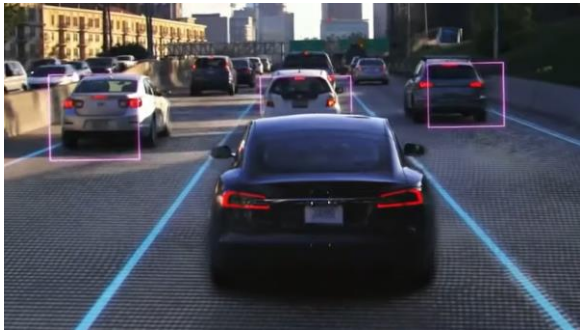
- 1997年：ディープ・ブルーがチェスの世界チャンピオンに勝利
- 2017年：自己対戦による学習能力を持つAlphaZeroが登場
- 2017年5月：AlphaZeroが囲碁で世界レベルの選手に対して3戦全勝

機械学習が、複雑なゲームでも、知的な戦略や判断を行う能力を持っている

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

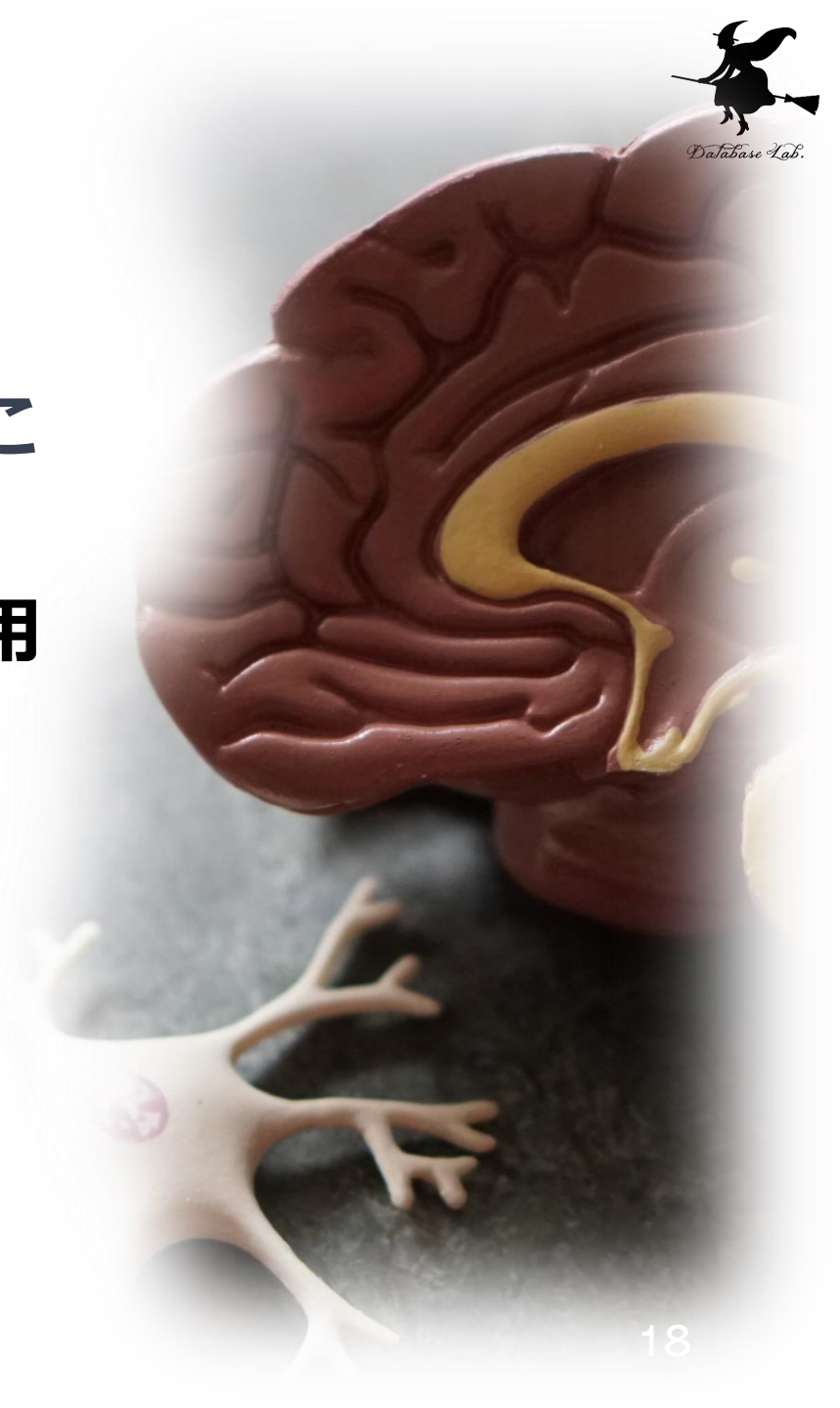
さまざまな種類の画像を，安定して，高精細に生成することが可能に

Large Scale GAN Training for High Fidelity Natural Image Synthesis
Andrew Brock, Jeff Donahue, Karen Simonyan, https://arxiv.org/abs/1809.11096v2

4-3. ニューラルネットワークの概要

ニューラルネットワーク

- **ニューラルネットワーク**は、**人工知能**の一種
- **生物の脳の働きを模倣すること**を目指している
- **機械学習**に，**すでに広く活用が進んでいる**

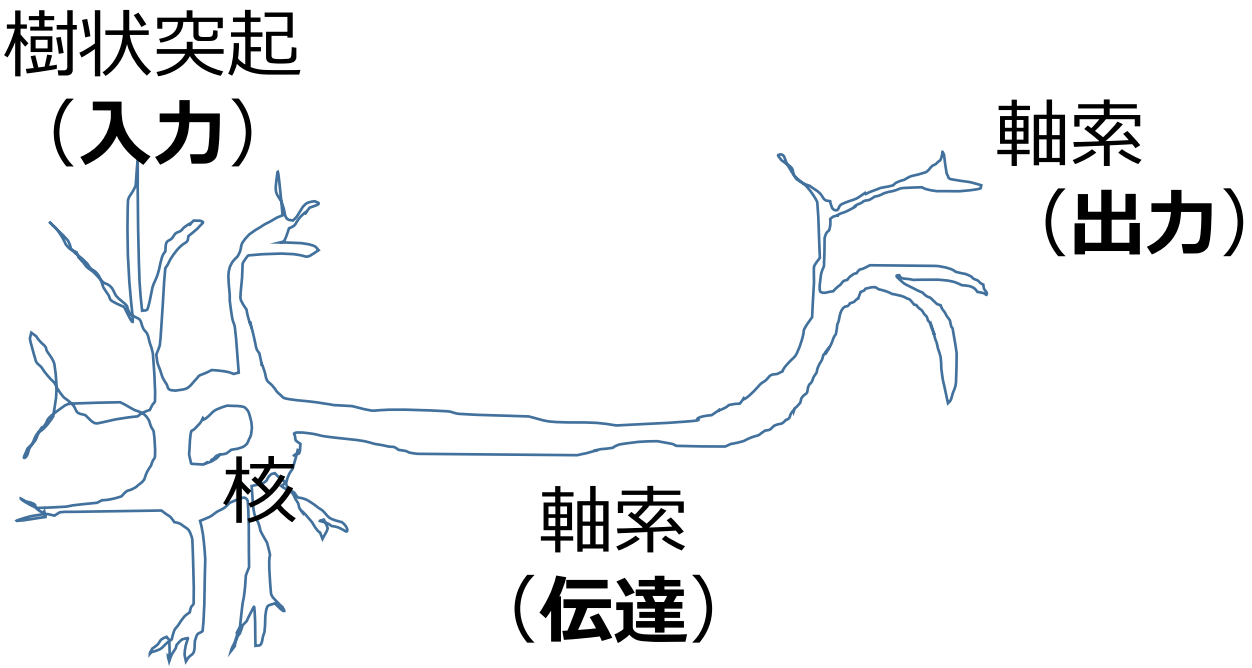


ニューラルネットワークとニューロン

- **ニューラルネットワーク**は，**脳や神経系のニューロン（神経細胞）**を参考に作られている
- **ニューロン**が**信号を受け取り，処理**し，**伝達**する**仕組みを模倣**している

ニューロン

脳や神経系の**ニューロン**は，生物の脳や神経系で、**情報伝達，記憶，情報処理**を行う．脳や神経系の大切な要素.



脳や神経系の研究の進展

1900年頃: **脳や神経系**は，**ニューロンの集まり**と考えられるようになった．ニューロンが情報伝達を行い，脳や神経系の基本要素と分かってきた．

1980年頃: **脳機能マッピングによる脳の解明が進んだ**．どの部分がどの機能を持つかの解明のために，生きた脳を対象とした測定が行われるように．

2010年頃：**異なる脳領域の連携**が観察されるようになった．脳は異なる領域が密接に連携して情報処理を行っていることが観察できるように．

脳の仕組みと機能について**深い洞察**が得られるように

脳のニューロン

生物のニューロンの数（推定値）

- カタツムリ　11,000
- ロブスター　100,000
- アリ　250,000
- カエル　16,000,000
- ハツカネズミ 71,000,000
- タコ　500,000,000
- ネコ　760,000,000
- **ヒト　86,000,000,000**
- アフリカゾウ 257,000,000,000

Wikipeida の記事:
https://ja.wikipedia.org/wiki/%E5%8B%95%E7%89%A9%E3%81%AE%E3%83%8B%E3%83%A5%E3%83%BC%E3%83%AD%E3%83%B3%E3%81%AE%E6%95%B0%E3%81%AE%E4%B8%80%E8%A6%A7 より

- 生物の脳は**非常に多くのニューロン**から成り立つ
- 人間の脳には**約860億個**のニューロンが存在
- 複雑なネットワークを形成

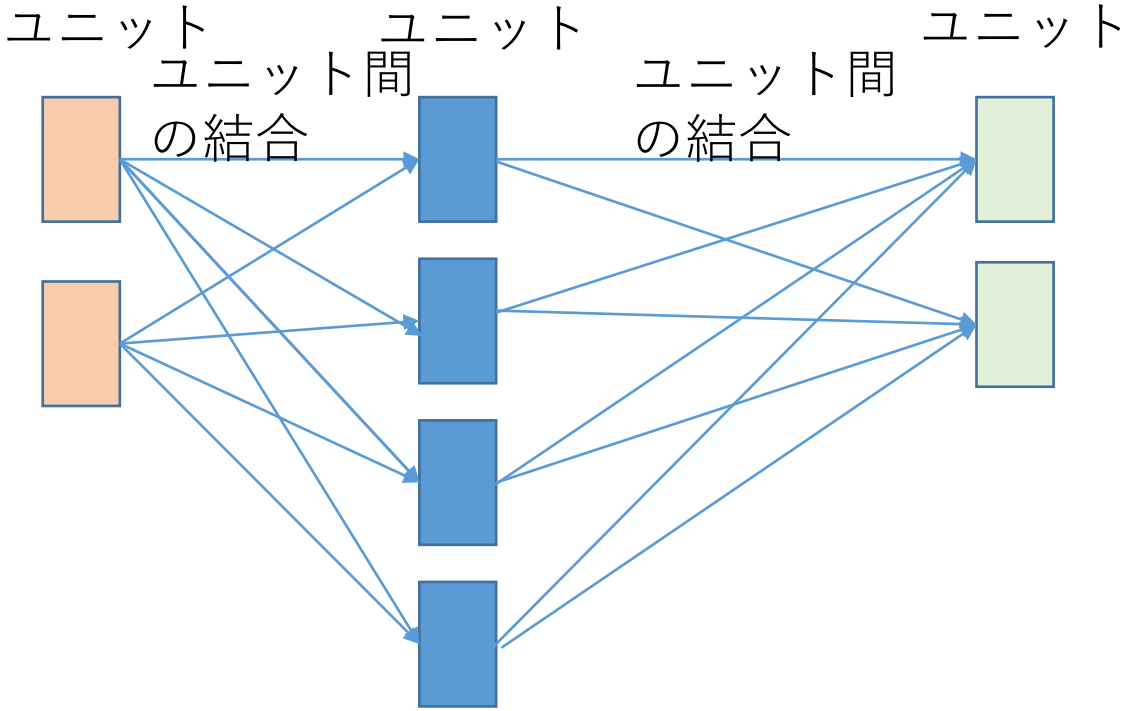
ニューラルネットワークの基礎

- 単純な仕組み．明確な原理に基づいて処理を行う

ユニットは，基本，**入力の重みづけ**，**合計とバイアス**，**活性化関数の適用**を行う

- ユニットのネットワークである

多数の**ユニット**とそれらの**ユニット間の結合**から構成



入力の重みづけ，合計とバイアス，活性化関数の適用

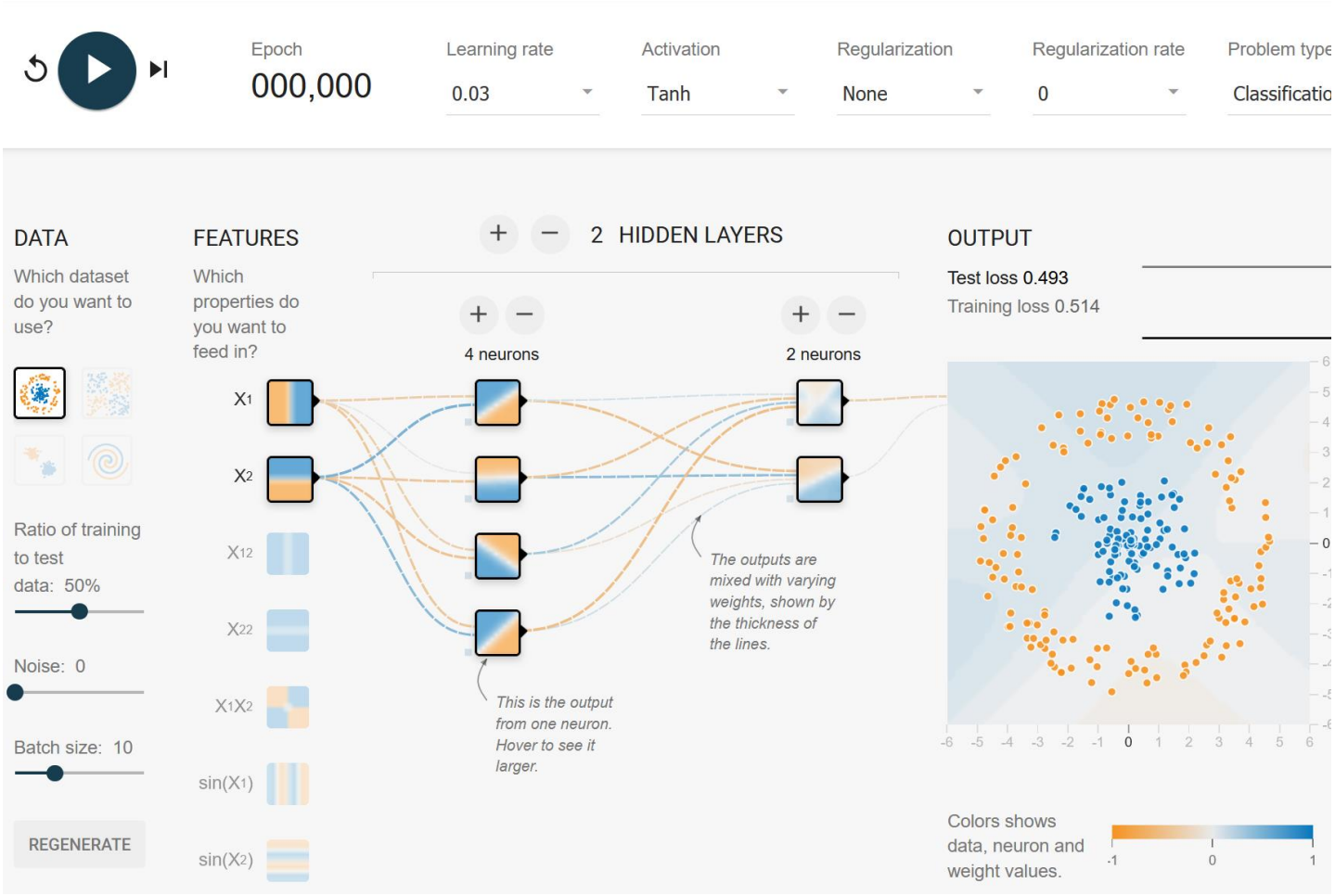
1. **入力の重みづけ**: **ユニット**は**複数の入力を受け取り**、それぞれの入力に対して重みを掛ける（掛け算）
2. **合計とバイアス**: **1. の合計を求め，バイアスと呼ばれる値を足す（足し算）**
3. **活性化関数の適用**: 2. の値に，活性化関数を適用して出力値を得る）

入力		重み		
0.3	×	0.1	⇒	0.03
-0.5	×	0.8	⇒	-0.4
0.2	×	-0.5	⇒	-0.1

合計
-0.47

合計に応じた出力値
0.3846

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

コンピュータの進展とニューラルネットワーク進展

年	コンピュータで扱えるユニット数の規模
2010年	100,000個
2020年	2,000,000個
2030年	50,000,000個
2040年	1,000,000,000個
2050年	20,000,000,000個

所説あります

2055年ごろには，860億を超えるかも？

【過去】

人間の脳の**約860億個**のニューロンを模倣することは不可能

【現状】

GPU，クラウドコンピューティングで，**コンピュータの性能向上が続く**

【将来】

生物の脳に近いニューラルネットワークの実現へ進む

ニューラルネットワークの歴史

- **1940年代**:
  ニューラルネットワークの誕生。
- **1980年代**:
  バックプロパゲーションやボルツマンマシンなどの**学習アルゴリズムが登場。**
  ニューラルネットワークの学習技術が進展した。
- **2010年代**:
  新技術の誕生。多層化、正規化、ReLU（Rectified Linear Unit）、ドロップアウトなど。
  ニューラルネットワークの性能が向上し、勾配消失や過学習といった**機械学習の難問に対処できるように**なった。

ニューラルネットワークの性能や学習の効率が向上し、広く活用されている。

ニューラルネットワークの現状と将来まとめ

- **ニューラルネットワーク**は**人工知能の一種**
- **生物の脳の働きを模倣**することを目指す.
- すでに，**機械学習**において、**広く活用されている**.

【将来】

- **コンピュータの処理能力の向上**により、生物の脳に近いニューラルネットワークの実現が期待されている
- ニューラルネットワーク用いた**機械学習の進化**，**普及**が進む

4-4. ニューラルネットワークの種類、応用分野

ニューラルネットワークの種類

① 自己符号化（オートエンコーダ）

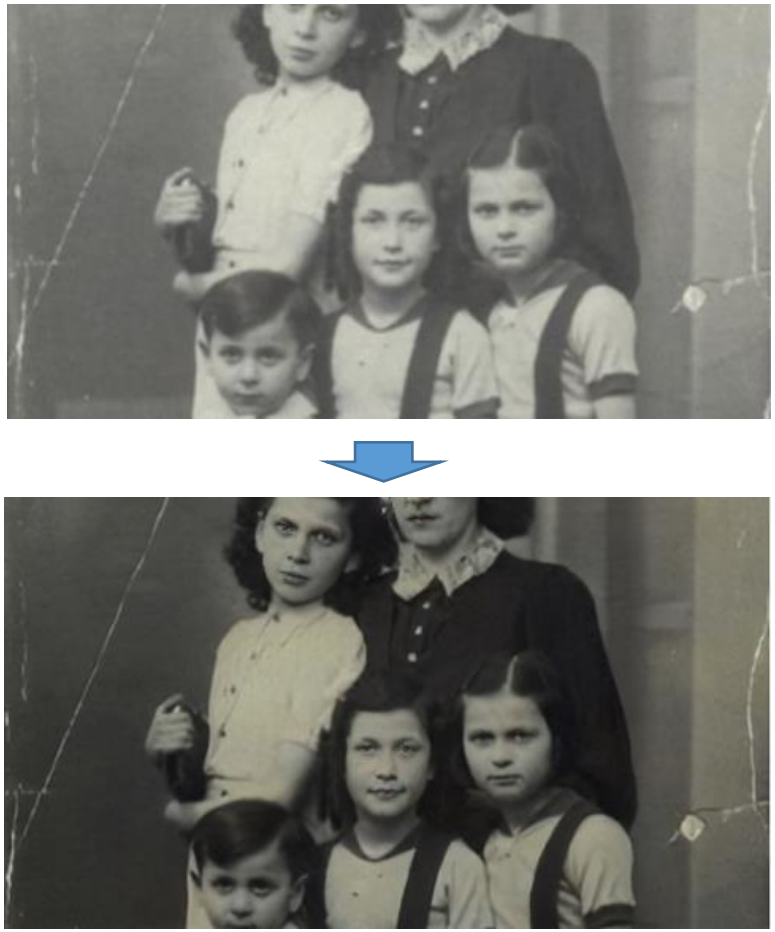
データを低次元の符号にマッピング

② 分類や予測

データから分類結果，予測結果を導く

データを用いて学習を行うことは共通している

自己符号化の応用例



画像復元

写真からの顔の３次元化

自己符号化の応用分野

- **特徴抽出**

自己符号化は、データを低次元の符号にマッピングするので，特徴抽出に活用できる

- **データ圧縮**

低次元の符号化表現は元のデータよりも効率的な表現となる

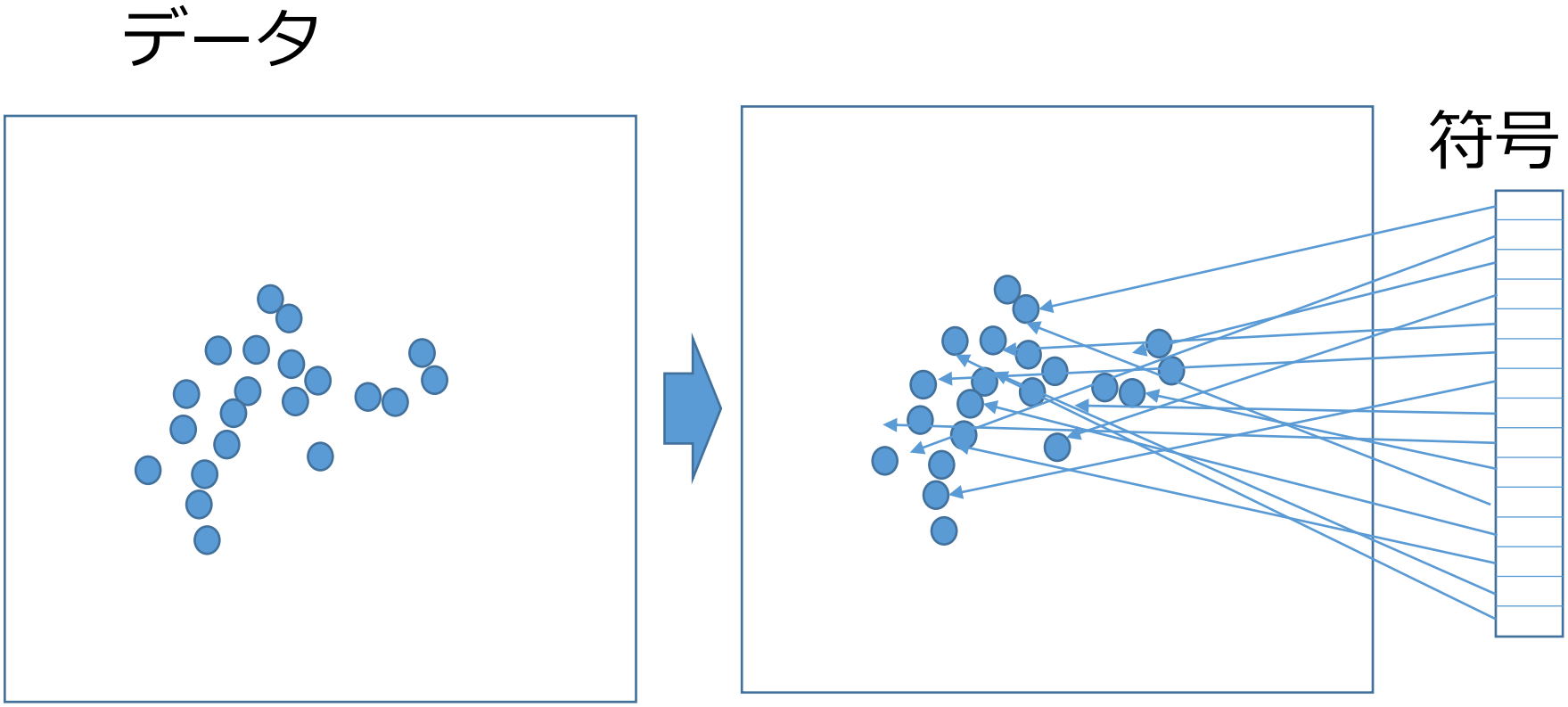
- **ノイズ除去**

ノイズのあるデータを与え、ノイズのないデータを復元できる

- **異常検出**

正常なデータのパターンを学習し、異常なデータを検出するために使用できる

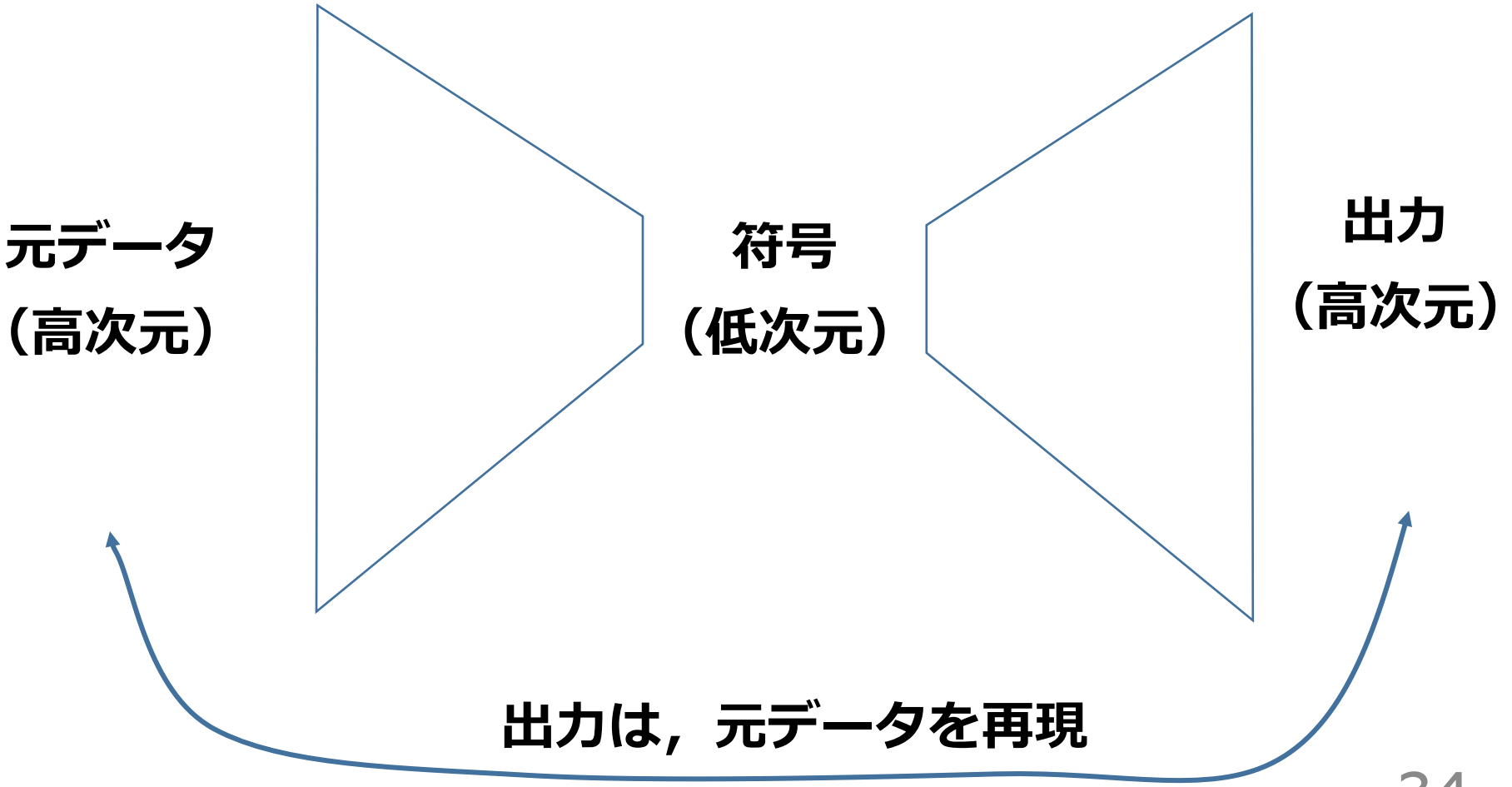
自己符号化でのデータと符号



データを自動で**符号化**

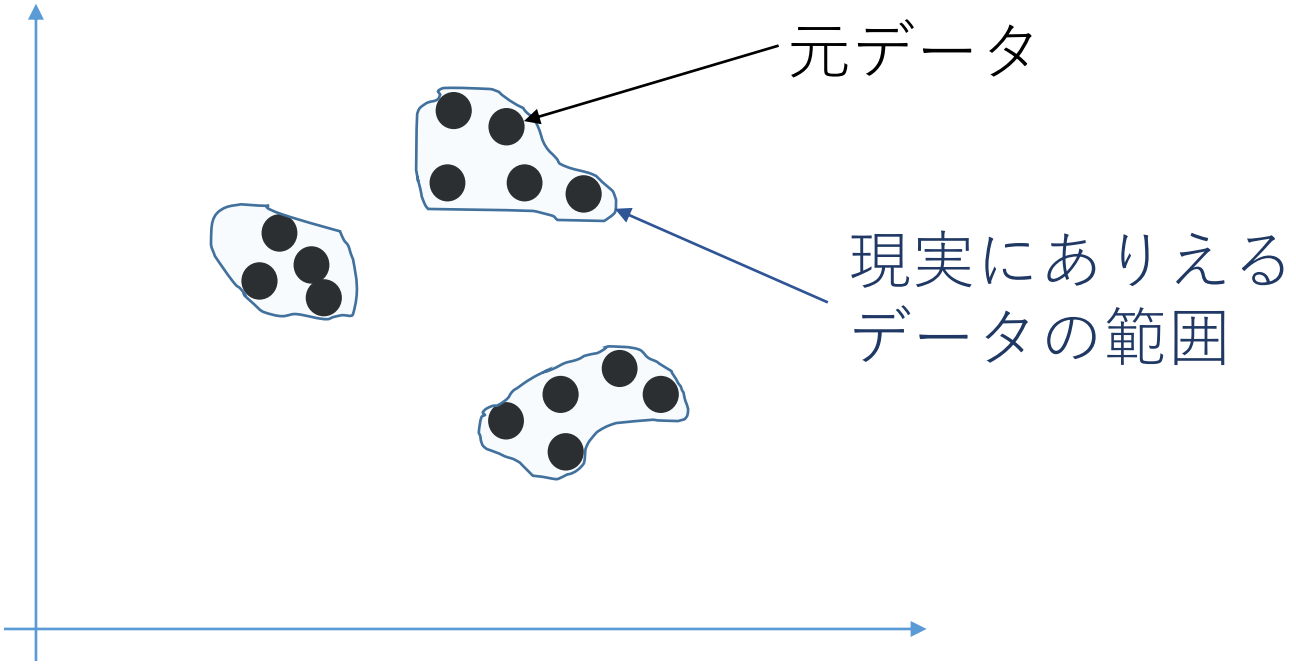
自己符号化での符号化と再構築

元データを**符号化**して**低次元の符号**に**変換**、
それを**再構築する復号化**の過程を通じて、
データの特徴を抽出することが目的



自己符号化でできること

自己符号化の結果として，**復号化（符号から元データの再構築）ができる**ことから，「**現実にあり得るデータを生成できる能力を獲得**」と考えることもできる



自己符号化まとめ

- 機械学習の一手法

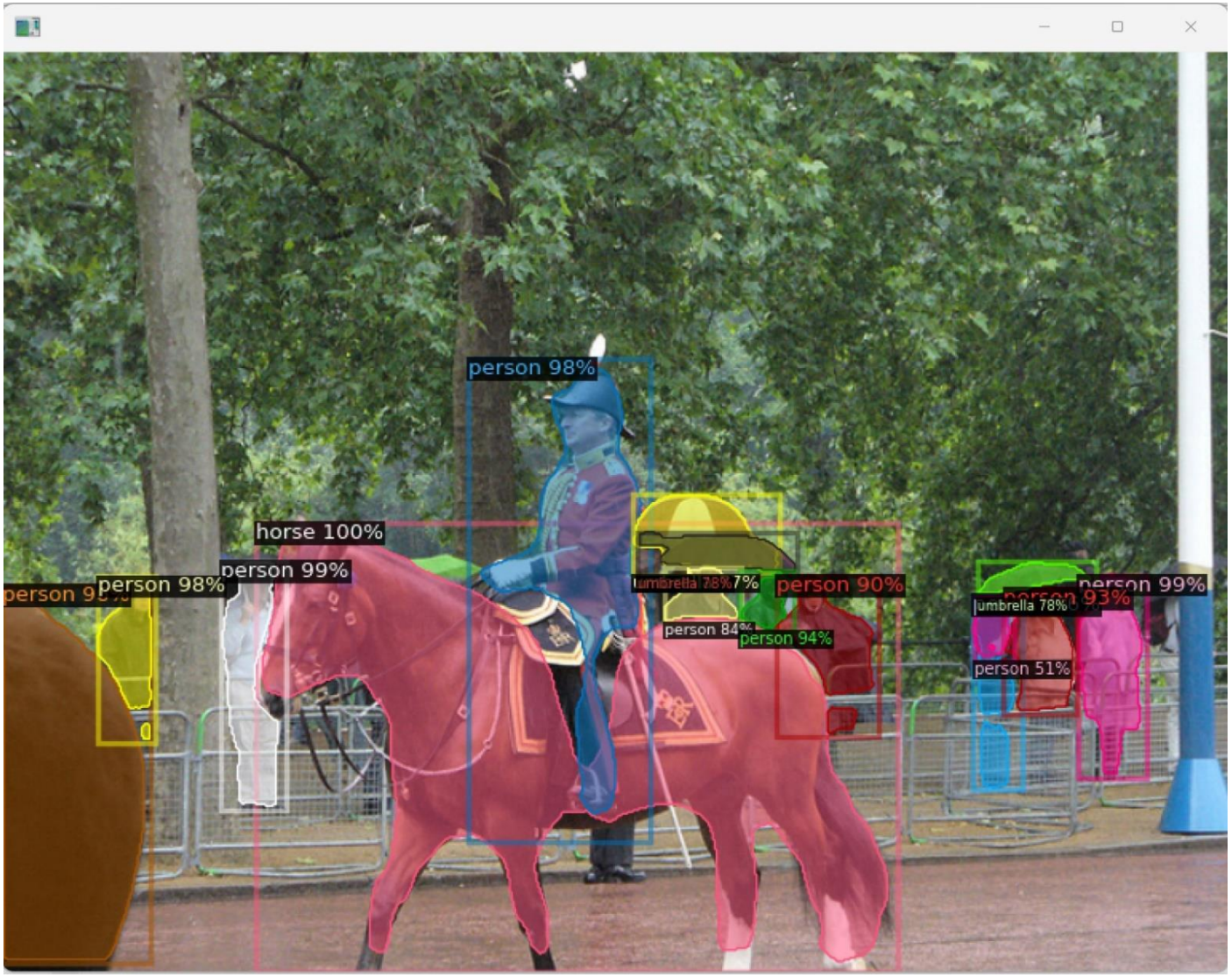
- データを**低次元の符号**に**マッピング**

- **データの特徴を抽出**することが**目的**

入力データを**符号化**して**低次元の符号**に変換し、それを**再構築**する**復号化**.

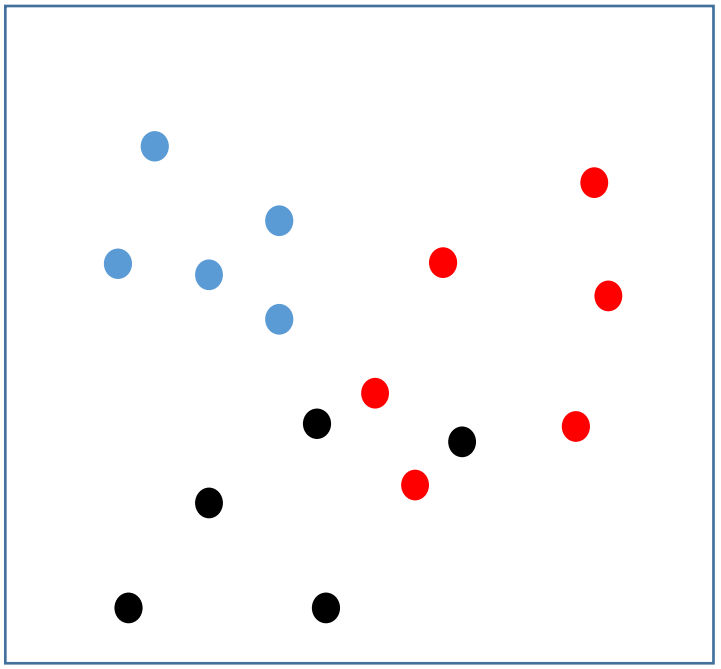
- **データ**を用いることで，**符号化と復号化の精度向上**.

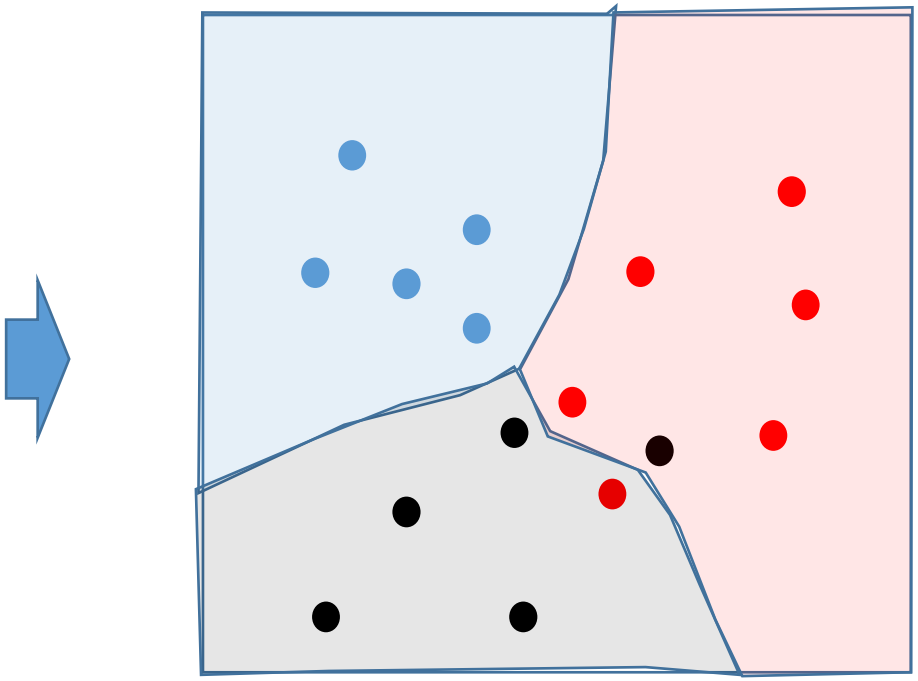
分類の応用例



物体検知、セグメンテーションなどの画像認識

分類





分類や予測の応用分野

人工知能による**分類や予測の応用分野**は**多岐にわたる**

- **画像認識と画像理解**

道路標識の認識や顔認識、物体検出、画像セグメンテーションなど。自動運転技術や監視カメラの映像解析などに応用。

- **自然言語処理**

要約、校正、質問応答、翻訳、音声認識、音声合成、感情分析、文書分類、アイデア出しなど。

- **バイオインフォマティクス**

遺伝子解析やタンパク質構造の予測、薬剤の設計など。

- **金融と経済予測**

市場のトレンド分析やリスク評価、予測モデルの構築、意思決定のサポートなど。

- **医療診断と予測**

疾患の診断や予測、画像診断（MRIやCTスキャン）、病気の予後予測、薬物副作用の予測など。

ニューラルネットワークで分類や予測を行う仕組み

過去のデータや**入力**に基づいて，**分類や予測**を行う

- **過去のデータによる学習**

データからパターンや関連性を抽出できるようになる

- **入力に対して，分類や予測を行う**

データから，複雑なパターンや関係性が学習され、予測や分類が行われる

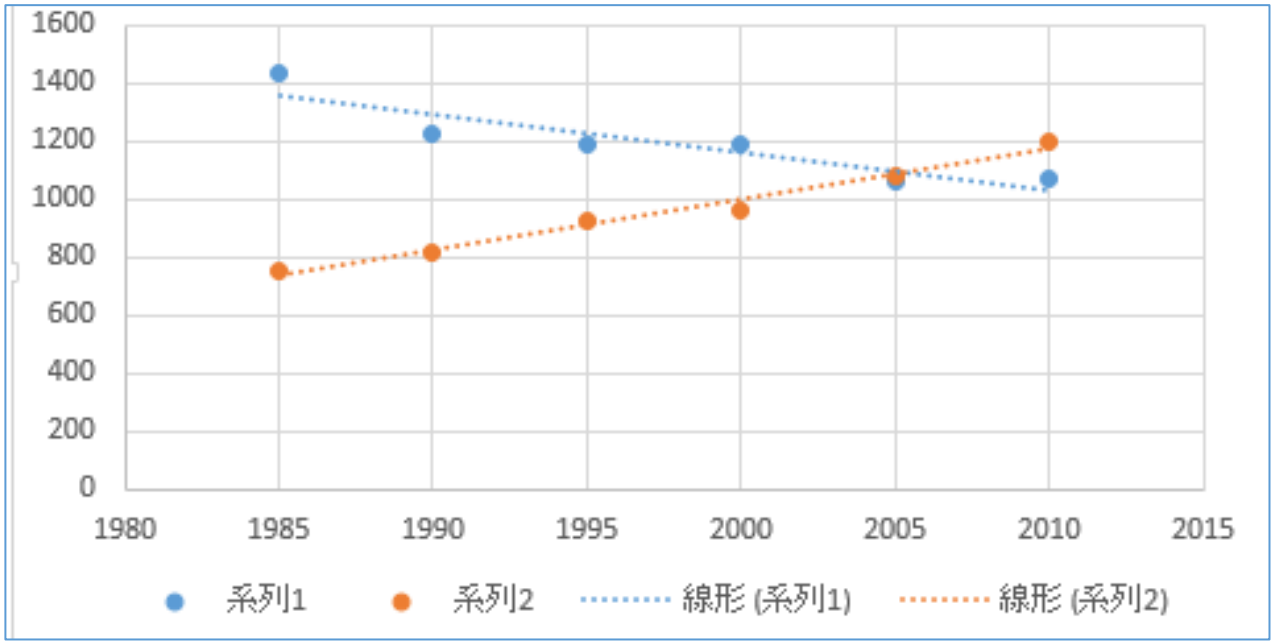
ニューラルネットワークによる分類や予測のまとめ

- **分類や予測**は、データから**特定の結果（分類結果や予測結果）**を**導き出す**ことを指す。
- **ニューラルネットワーク**は，過去のデータに基づいて**学習**し、**データからパターンや関連性を抽出**する能力を持つ
- **ニューラルネットワーク**は人工知能による**分類や予測**に**広く活用**されている。
- **応用**は，画像認識と画像理解，自然言語処理，バイオインフォマティクス，金融と経済予測，医療診断と予測など**幅広い**

4-5 線形近似（学習の例）

線形近似

- **線形近似**は、**データを直線で近似**する手法
- **データ点が多い**ほど、**より正確な近似結果**
- 線形近似はデータの特徴を捉える一種の学習手法としても考えることができる

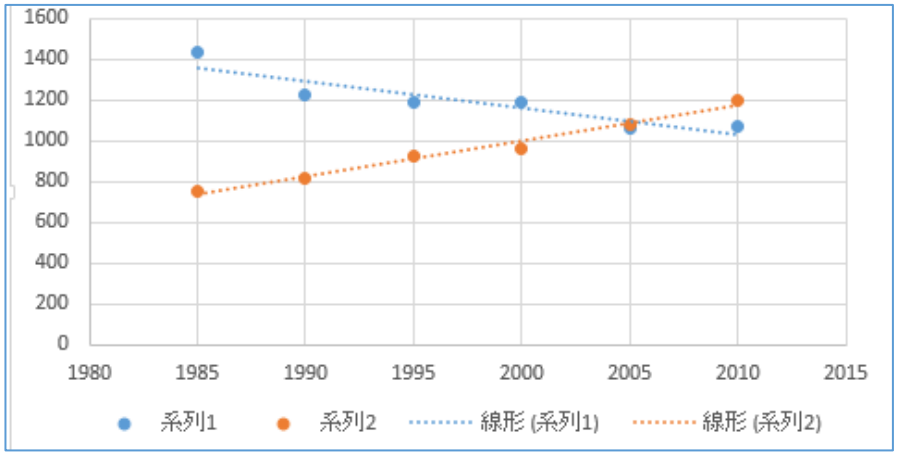


散布図＋線形近似

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

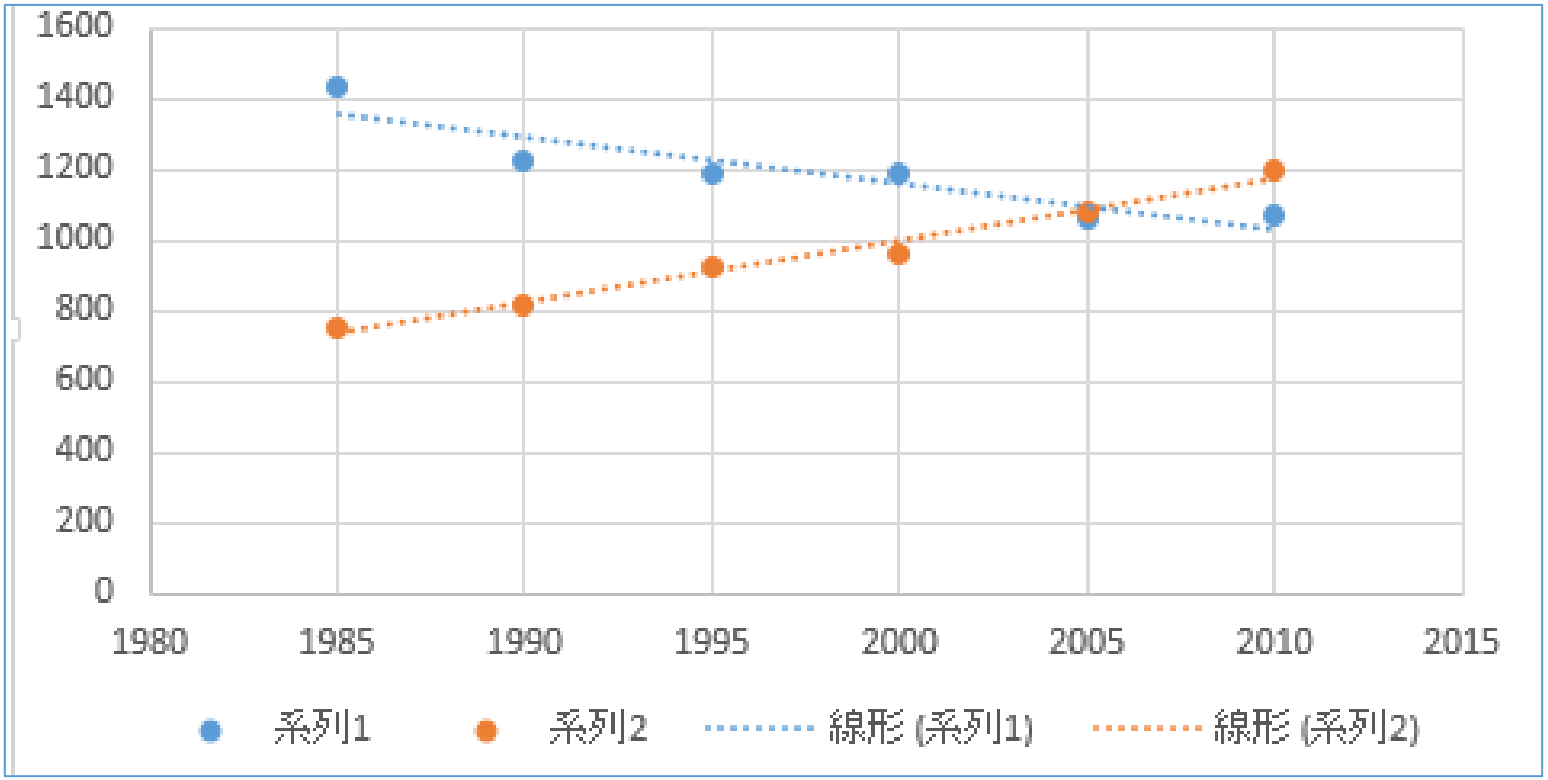
入力　　出力



散布図＋線形近似

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

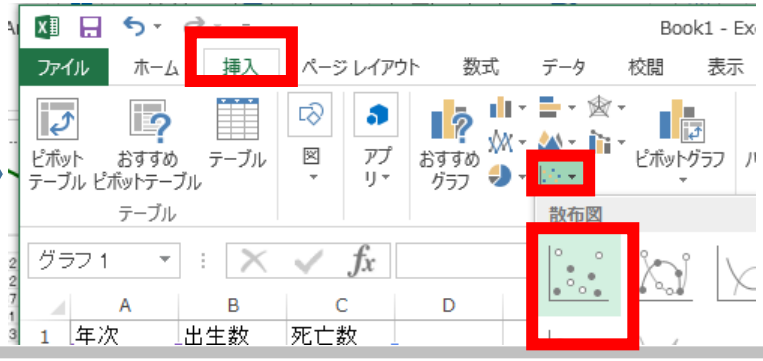
Excel での散布図＋線形近似の作成手順（１／３）

	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

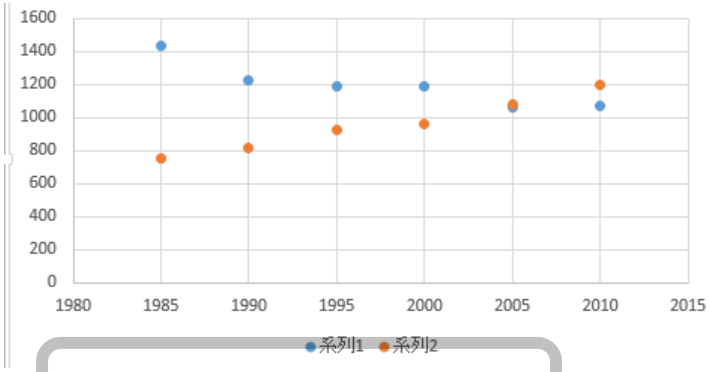
元データ

	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**

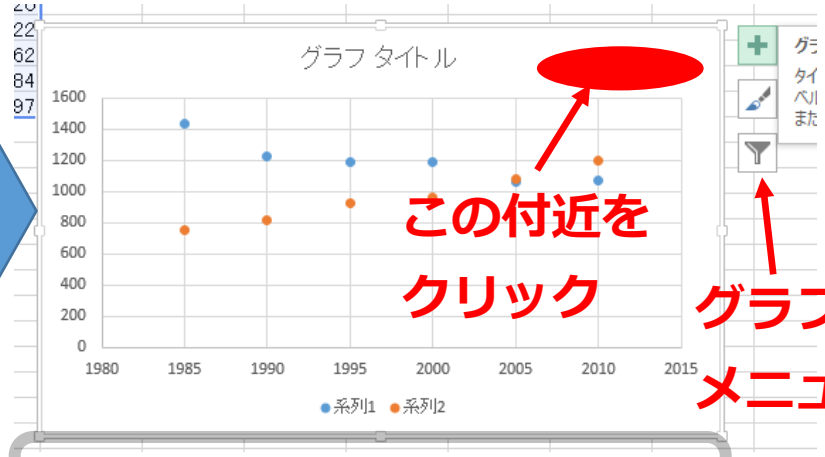


② リボンで「**挿入**」→散布図の選択

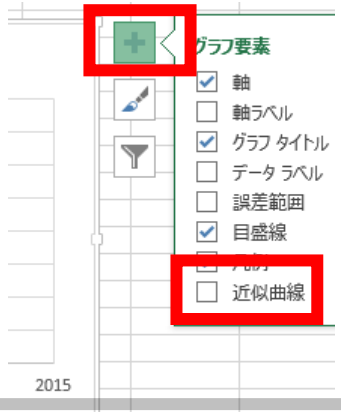


③ 散布図を確認

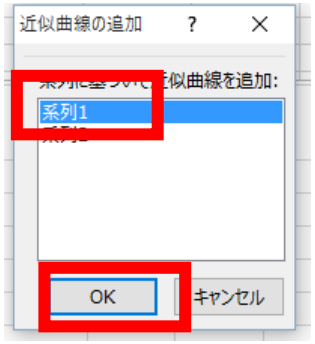
Excel での散布図＋線形近似の作成手順（２／３）



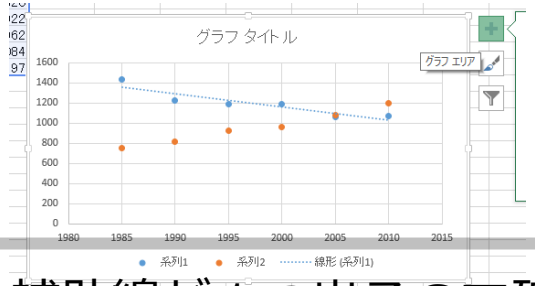
④　グラフメニューを出す



⑤　「＋」のボタンをクリックしたのちに，「近似曲線」をクリック

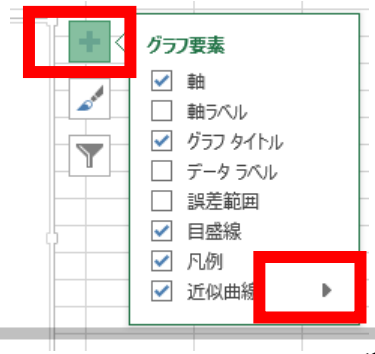


⑥「**系列１**」をクリックして「**ＯＫ**」

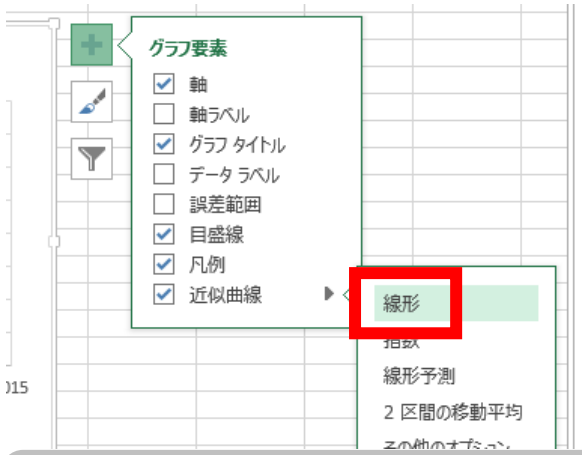


⑦　補助線が１つ出るので確認

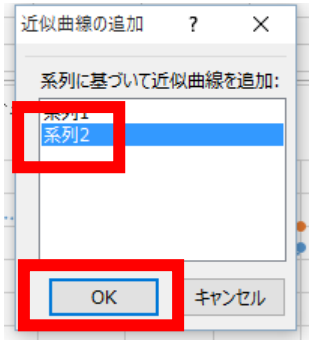
Excel での散布図＋線形近似の作成手順（３／３）



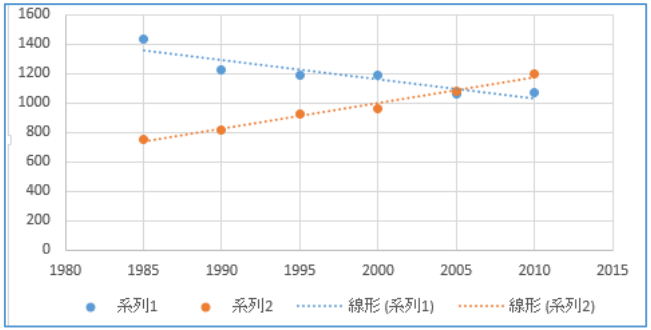
⑧　もう１度，**グラフメニュー**を出し，**近似曲線の右横の「▶」を展開**



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、与えられた問題に対して**最適な解を求めるプロセス**
- 最適化は、**パラメータや変数の値などを調整**して、目的関数（ゴールともいう）と呼ばれるものを**最小化または最大化**する

ゴールの例：　誤差の最小化

パラメータの例：　直線の上下の位置と，直線の傾き

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．

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

- 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.]
```

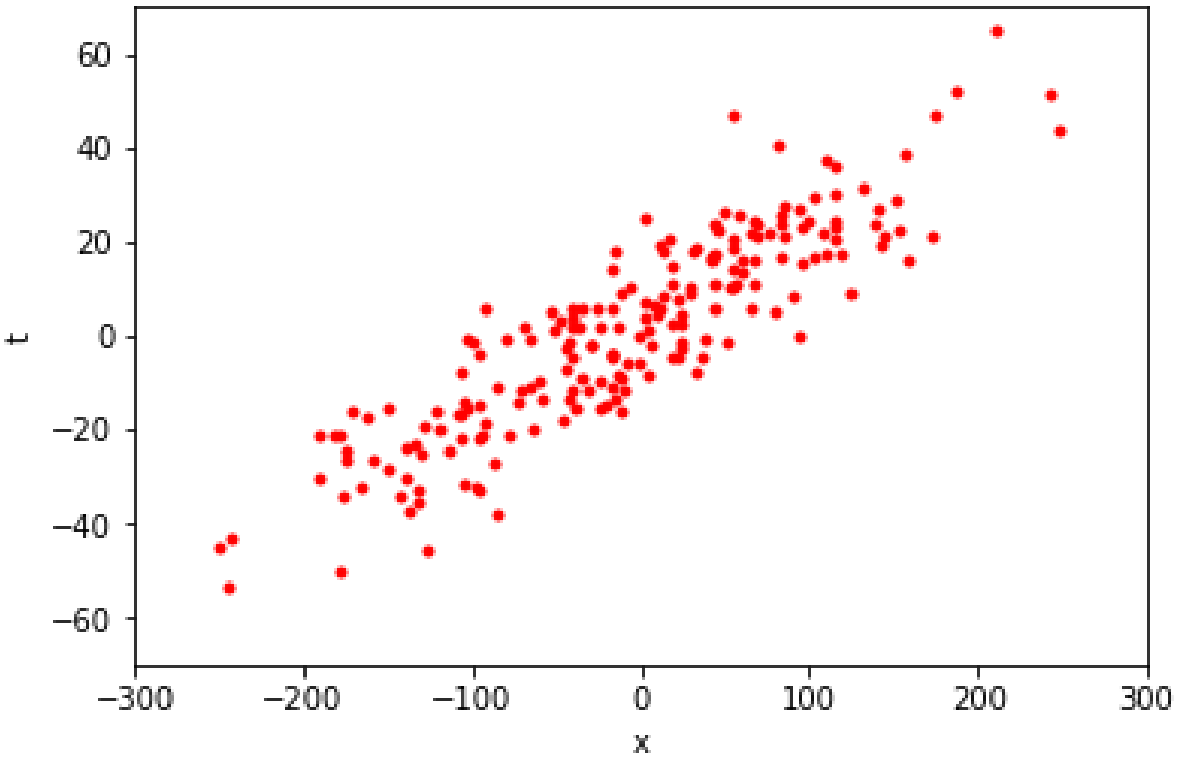
x = [1 1 1 1 1] のとき（すべての値が 1のとき）
最適であると求まった.

線形近似での最適化

線形近似では**最適化**が行われ、**最適な直線の探索**が行われる

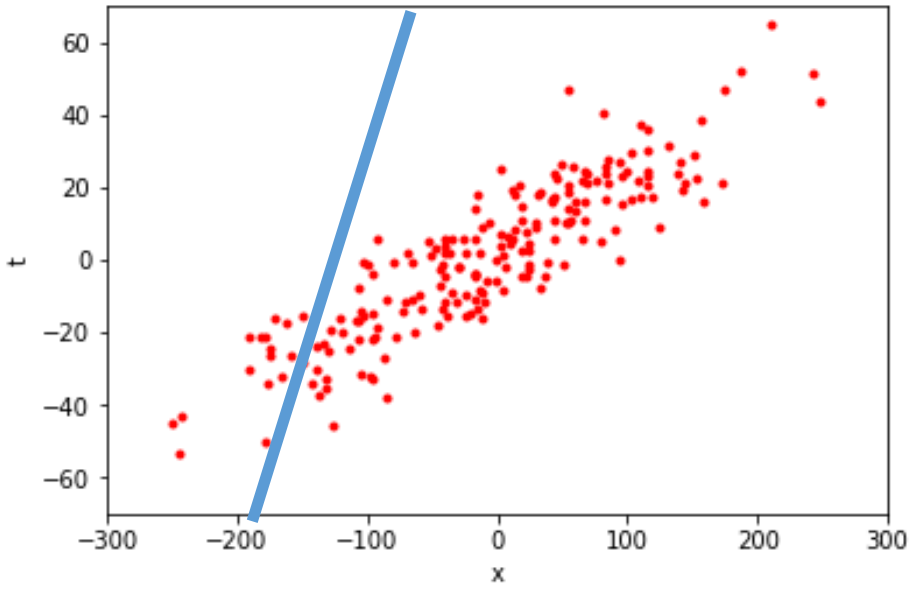
- 線形近似は、**データを直線で近似**することを目指す
- **最適な直線**を探索するために、**最適化**の手法を使用
- 直線を適切に調整することで、**データに最もよくフィットする直線を見つける**

データの例

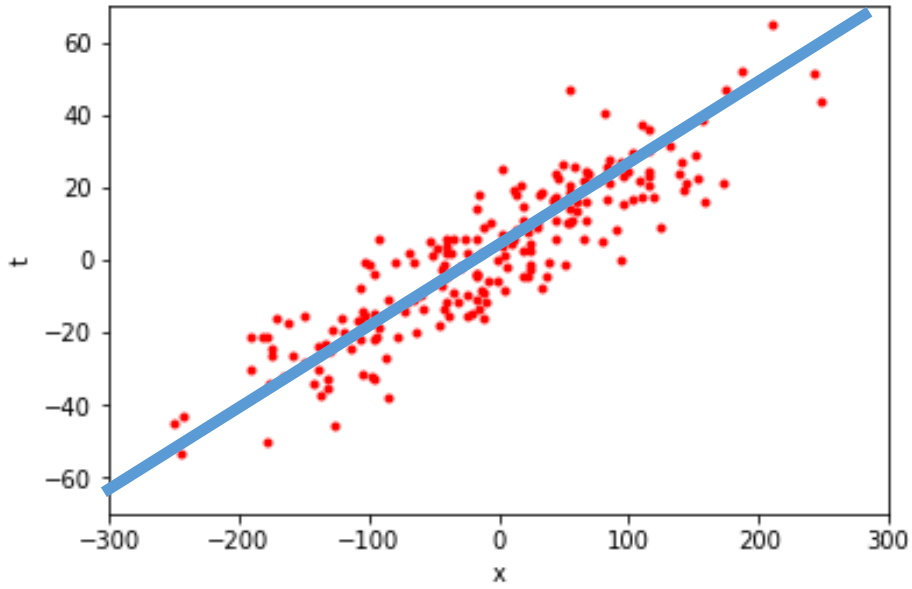


- **多数のデータの集まり**
- 上の図では，点１つで，１つのデータ

直線による近似の例

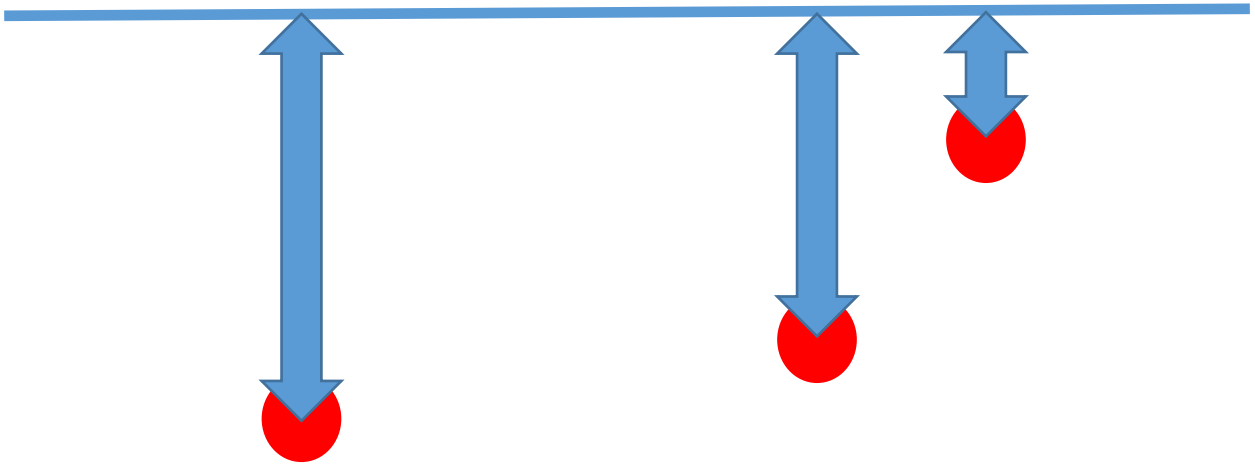


最適でない近似
（誤差大）



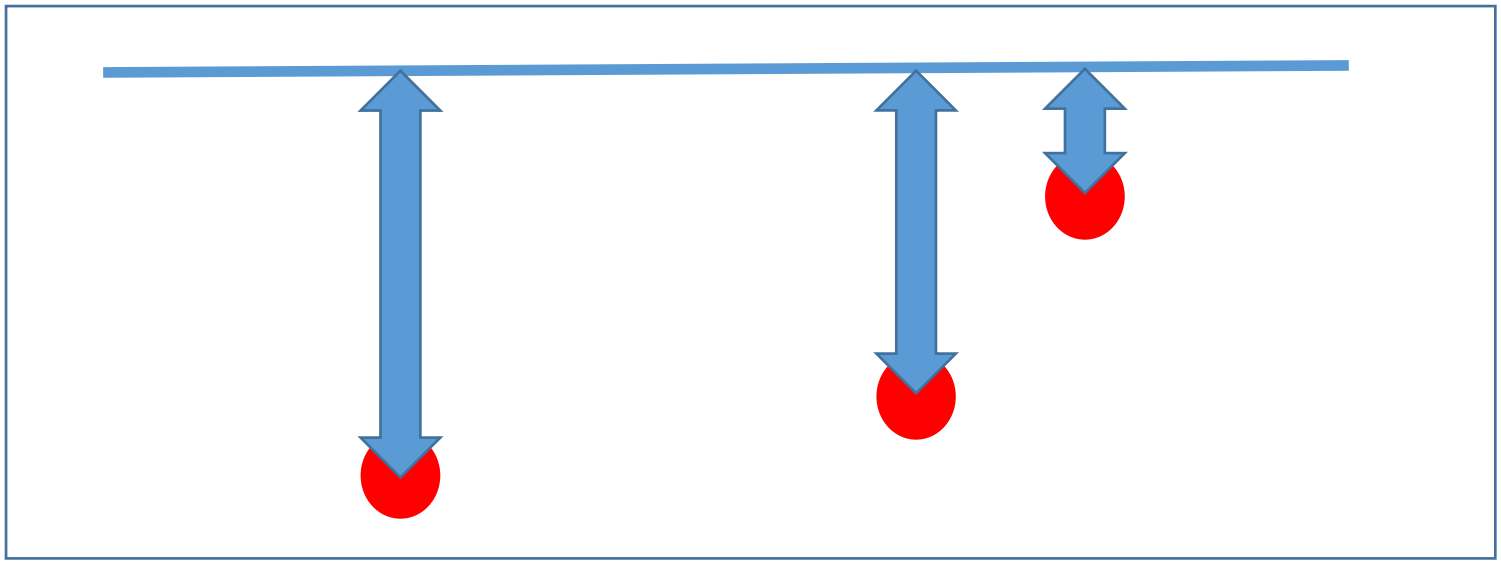
最適な近似
（誤差小）

誤差

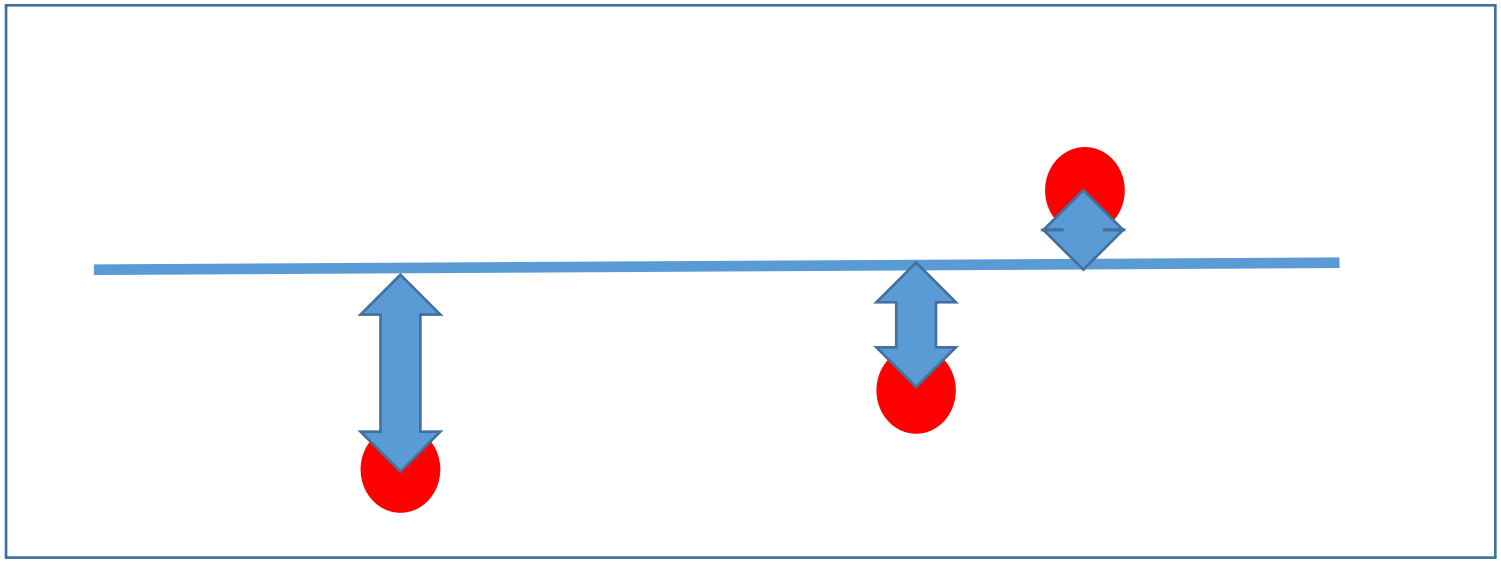


赤点：元データ

直線の上下移動による誤差の変化



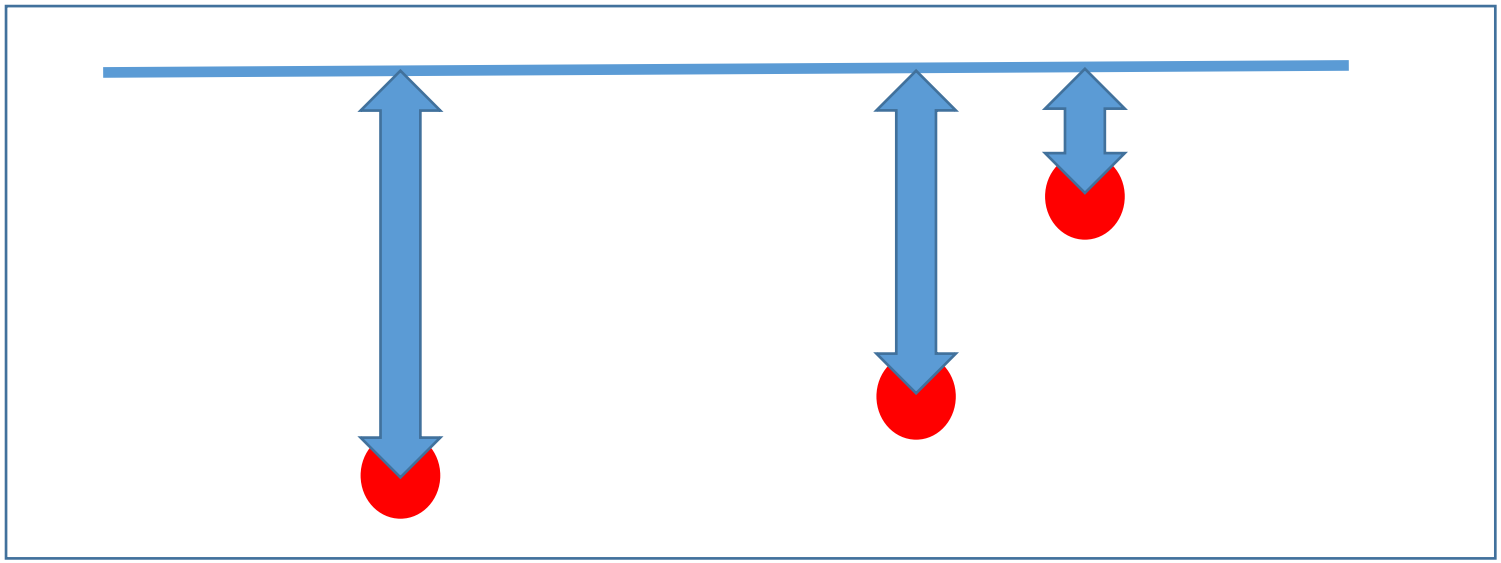
誤差大



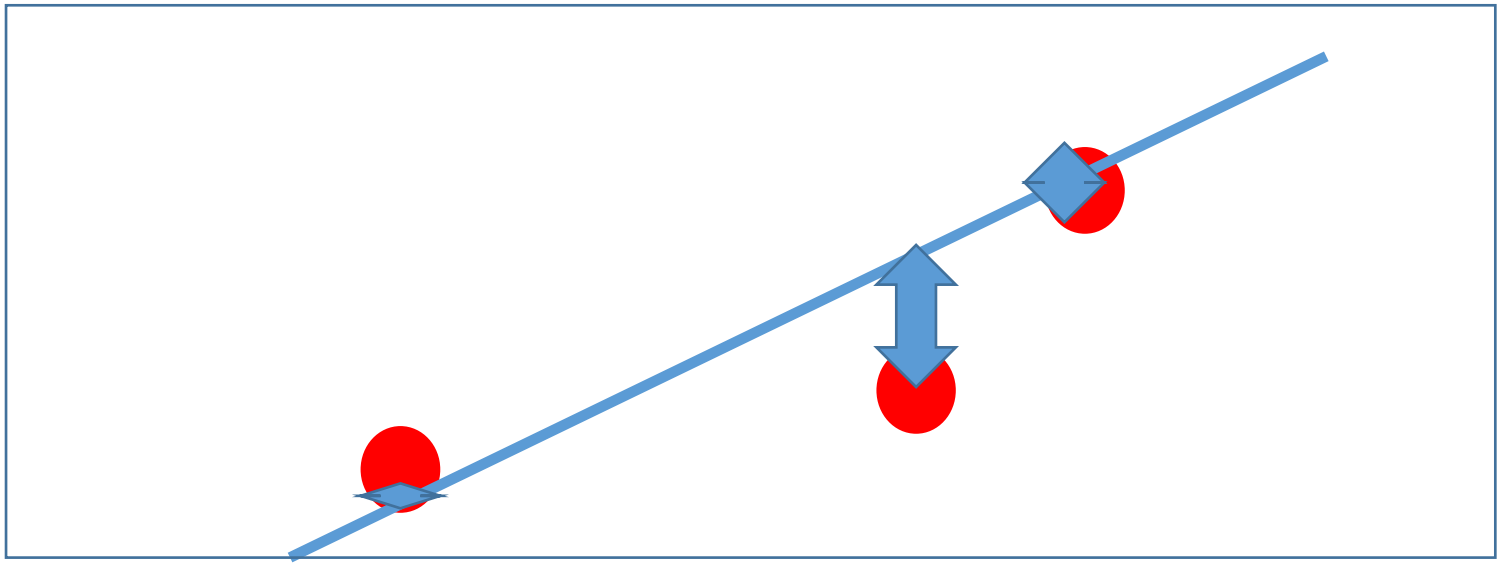
誤差小

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

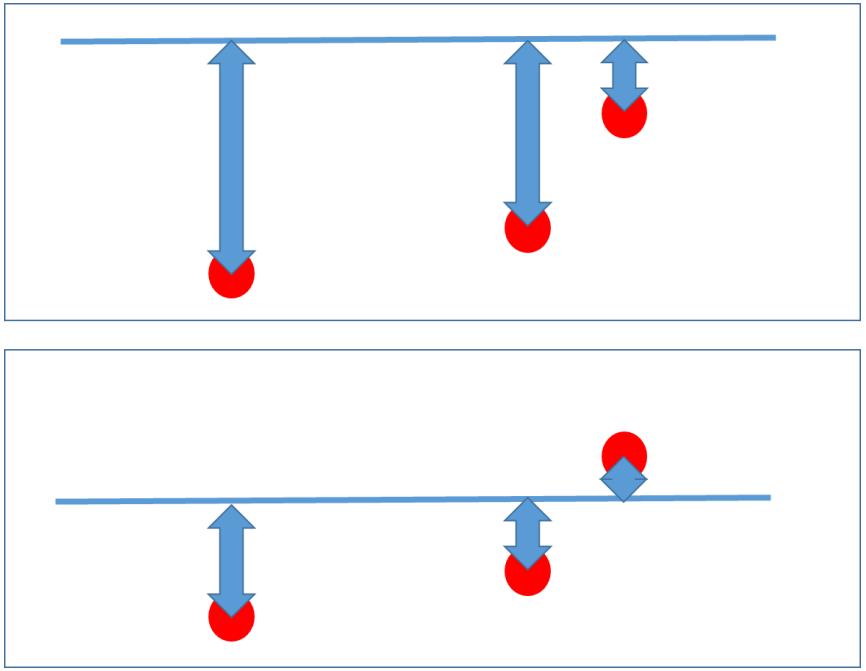


誤差小

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

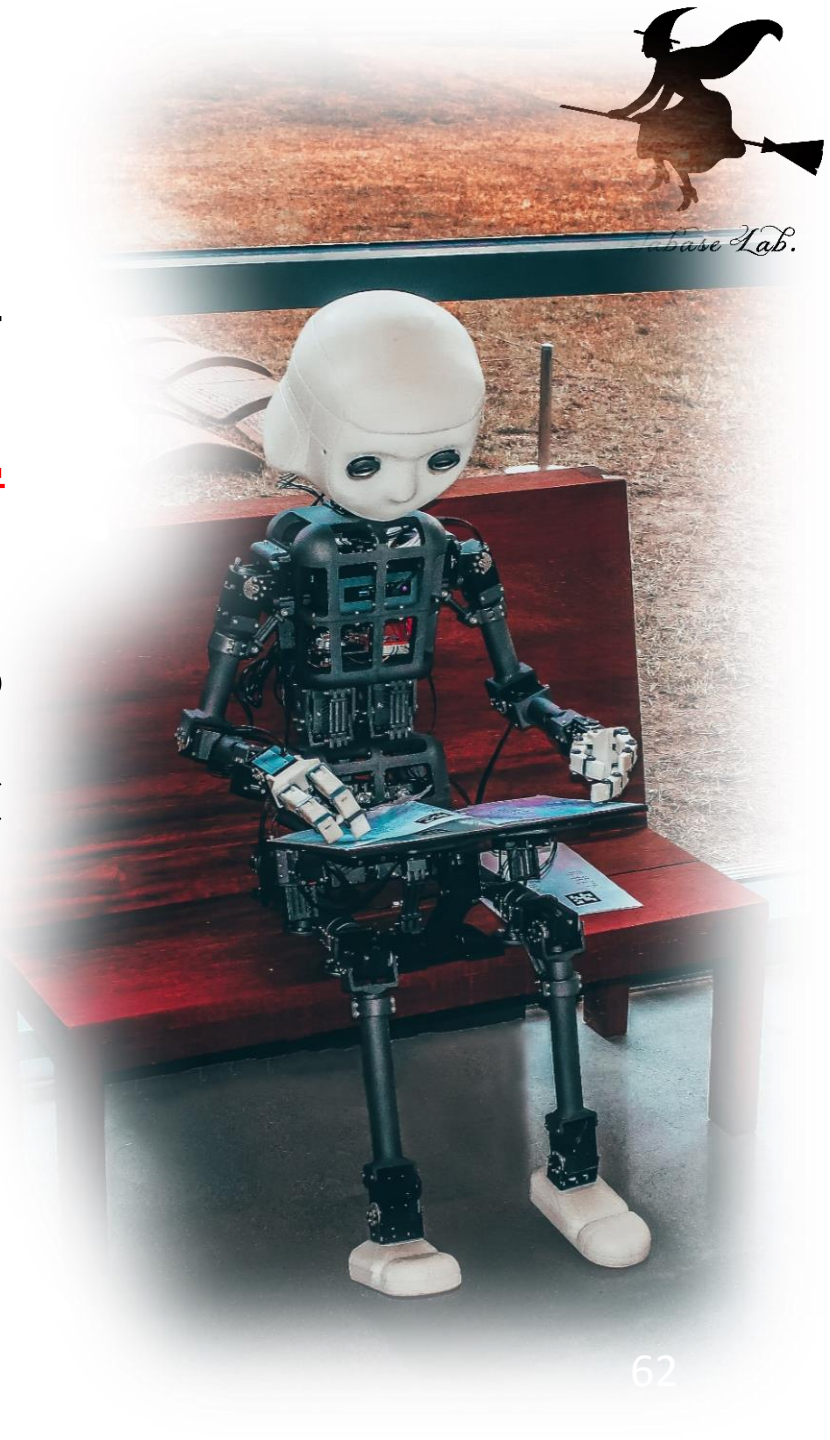


最適化は機械学習でも重要

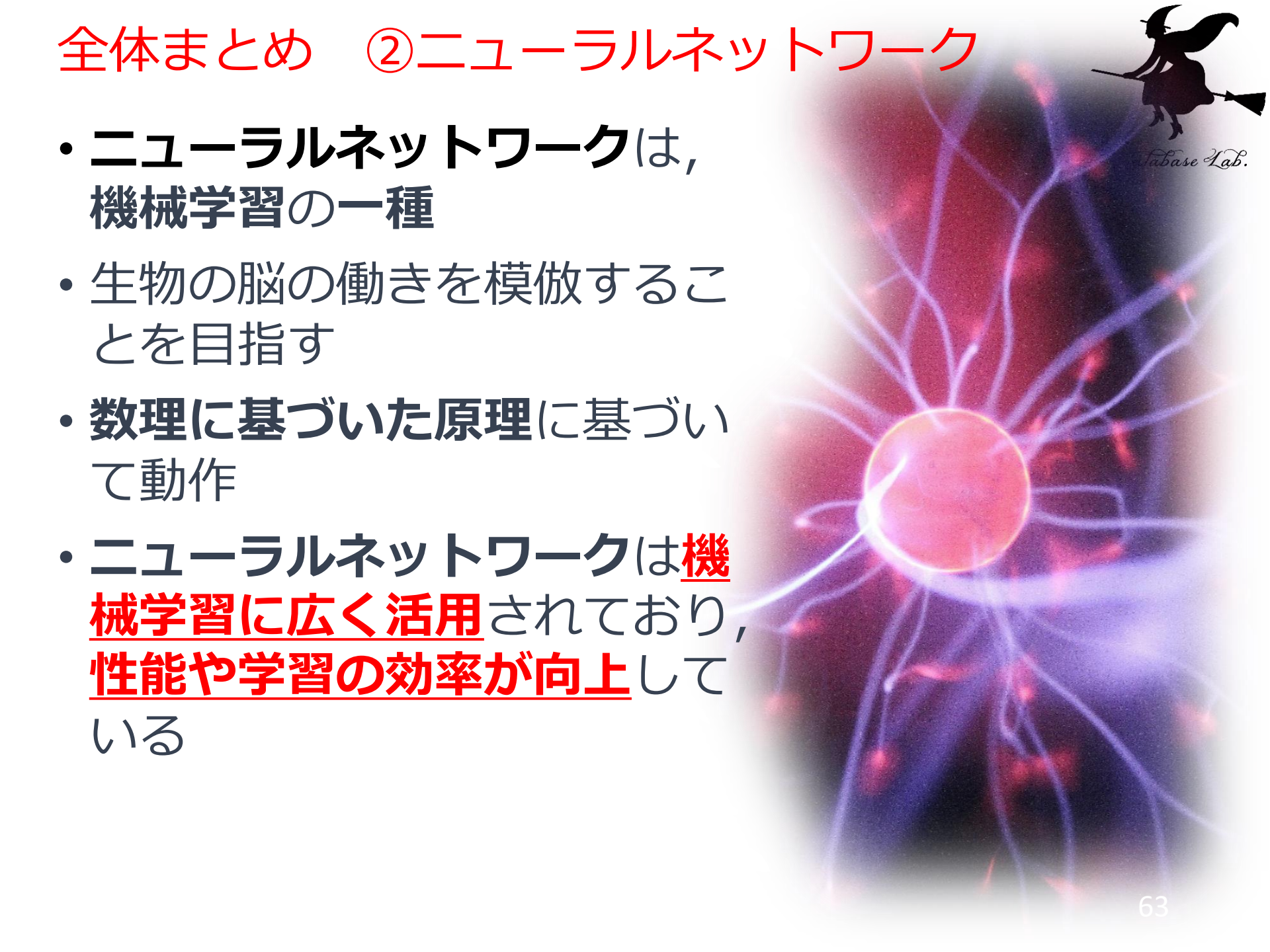
- **ニューラルネットワーク**は，**学習**の過程で，**最適な状態になるように調整**される
- **最適化**は，**ニューラルネットワーク**の**学習**で，重要な役割を果たす
- **ニューラルネットワーク**のパラメータ（重みやバイアス）を**最適化**することで、**性能を向上**させる

全体まとめ　①機械学習

- **機械学習**は，**コンピュータ**が**データ**を使用して**学習**することより**知的能力を向上**させる技術
- 情報の抽出能力，ルールや知識のプログラム化不要などの**特徴**を持つ
- 他の方法では**解決が難しかった問題に取り組むことが可能**に



全体まとめ　②ニューラルネットワーク



- **ニューラルネットワーク**は，**機械学習の一種**
- 生物の脳の働きを模倣することを目指す
- **数理に基づいた原理**に基づいて動作
- **ニューラルネットワーク**は**機械学習に広く活用**されており，**性能や学習の効率が向上**している