構造データで
簡単な機械学習

ぽんて

目次

- 構造データとは
- 機械学習とは
- 機械学習の流れのイメージ
- 今回機械学習でやろうとしていること
- 実際に動かしてみよう+ソース解説
- まとめ

構造データとは

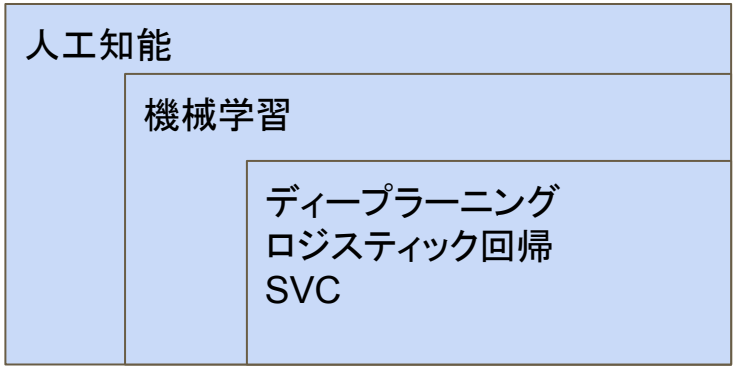
- 構造データ
  - 「行」と「列」の概念をもつデータ
  - EXCEL、CSV

- 非構造データ
  - 動画や画像のような「行」と「列」の概念でもてないデータ

機械学習とは

- 人工知能（AI）とは
  - 『計算』と『コンピュータ』を用いて『知能』を実現させようとする研究分野

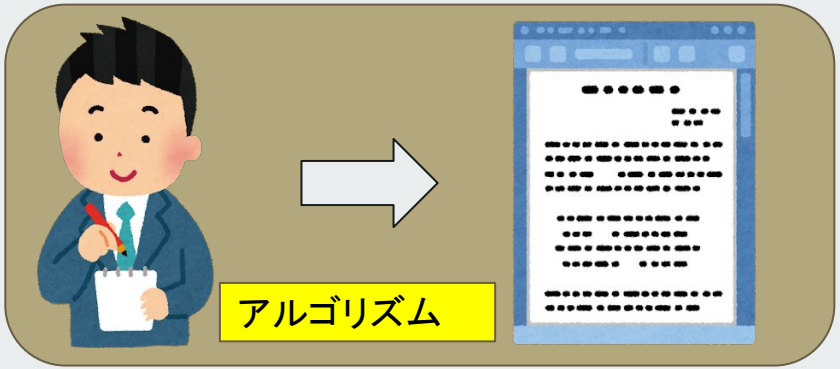
- 機械学習とは
  - 人工知能を実現させるための技術



機械学習の流れのイメージ

学習

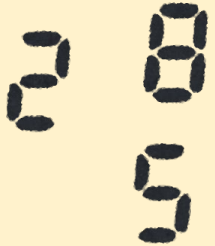
入力データ



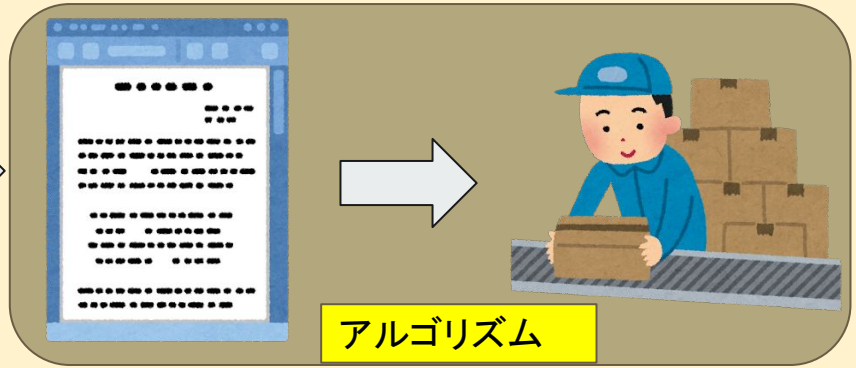
アルゴリズム

学ぶ

推論



入力データ



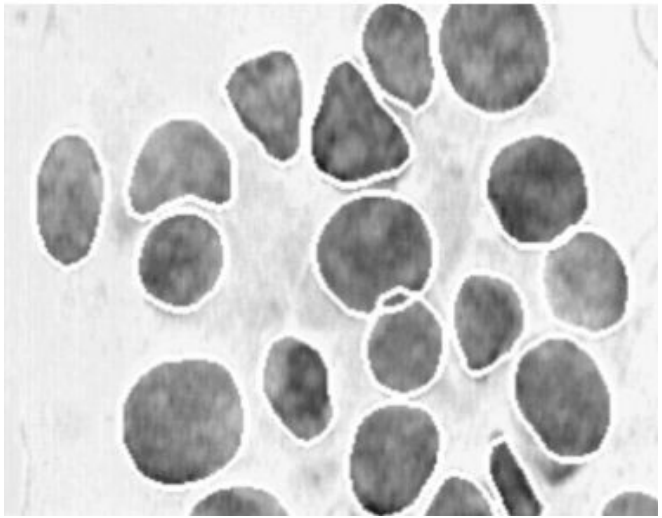
アルゴリズム

分類

今回機械学習でやろうとしていること

- 腫瘍細胞を**悪性か良性かを分類**する機械学習

	半径_平均	きめ_平均	周長_平均	面積_平均	正解
20	13.0800	15.7100	85.6300	520.0000	1
21	9.5040	12.4400	60.3400	273.9000	1
22	15.3400	14.2600	102.5000	704.4000	0
23	21.1600	23.0400	137.2000	1404.0000	0
24	16.6500	21.3800	110.0000	904.6000	0



悪性：0
良性：1

実際に動かしてみよう+ソース解説①

```
# # # # # # # # # # # # # # # # # # # #     前準備     # # # # # # # # # # # # # # # # # # # #
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from IPython.display import display
from sklearn.datasets import load_breast_cancer
from google.colab import drive
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
np.set_printoptions(suppress=True, precision=4)
pd.options.display.float_format = '{:.4f}'.format
pd.set_option("display.max_columns",None)
plt.rcParams["font.size"] = 14
random_seed = 100
```

※ソース解説①と②を
googlecolaboratoryに貼り付けて実行すると機械学習の結果が表示されます

実際に動かしてみよう+ソース解説②

```
# # # # # # # # # # # # # # # # # # # #      データセット      # # # # # # # # # # # # # # # # # # # #
cancer = load_breast_cancer()#構造データのロード
df = pd.DataFrame(cancer.data,columns=cancer.feature_names)
y = pd.Series(cancer.target)
x_train, x_test, y_train, y_test = train_test_split(df, y,
    train_size=0.7, test_size=0.3, random_state=random_seed)#学習と推論用にそれぞれデータを分ける

# # # # # # # # # # # # # # # # # # # #      学習      # # # # # # # # # # # # # # # # # # # #
algorithm = LogisticRegression(random_state=random_seed)#アルゴリズムにロジスティック回帰を選択
algorithm.fit(x_train, y_train) #学習実施

# # # # # # # ## # # # # # # # # # # # #      推論      # # # # # # # # # # # # # # # # # # # #
x_pred = algorithm.predict(x_test[0:5])#推論用データの上から5つに対し推論を実施
display(x_test[0:5])#推論したデータを表示
display(x_pred)#推論した結果を表示
```

まとめ

- 適切な構造データを準備できれば簡単に機械学習は可能（2値分類）
- 実用に耐えうる程度の正解率を出すには苦労するかも
  - 今回は用意されているデータセットを利用したため良い数値が出せたが実際は難しい。
- 製品のデータによる売れる売れない予想など