

# Pythonで行う機械学習プログラム

Lesson\_0 Python の基礎: インストール、主要ライブラリ操作 (pandas, numpy, matplotlib)

Lesson\_1 データセット取り扱いと機械学習基礎

Lesson\_2 Python: 機械学習 (分類) 評価と応用

Lesson\_3 Python: 機械学習 (回帰) 評価と応用

Lesson\_4 Python: 機械学習 前処理、最適化

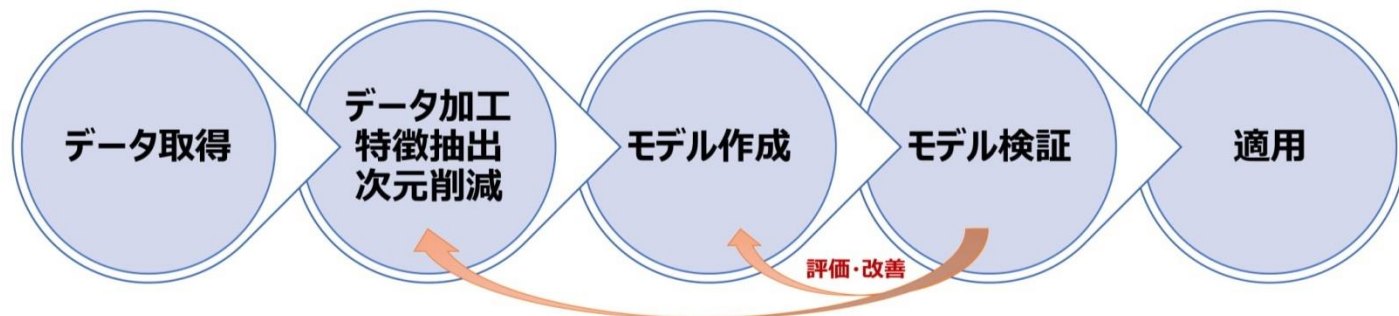
Lesson\_5 Python: OpenCV による画像処理

Lesson\_6 Python: DL 画像判別 (1) MLP と CNN

Lesson\_7 Python: DL 画像判別 (2) 転移学習、学習データの再利用、結果の保存

# 機械学習概要

## データ分析の進め方と留意点



1. やりたいこと（目的変数）が明確であること  
正常と異常の定義、判別したいものの定義、求めたい数値
2. データの質と量（説明変数）が十分であること  
関係ないデータや間違い、欠損が少なく、統計処理に量も十分
3. 予測や判別ができたとして、対策が具体化できること  
対策の内容、タイミング、体制、現場の希望を考慮

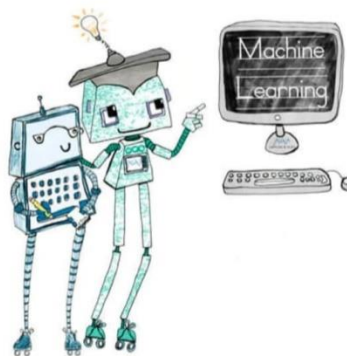
# 機械学習プロセス



# 機械学習概要

## What Machine Learning ?

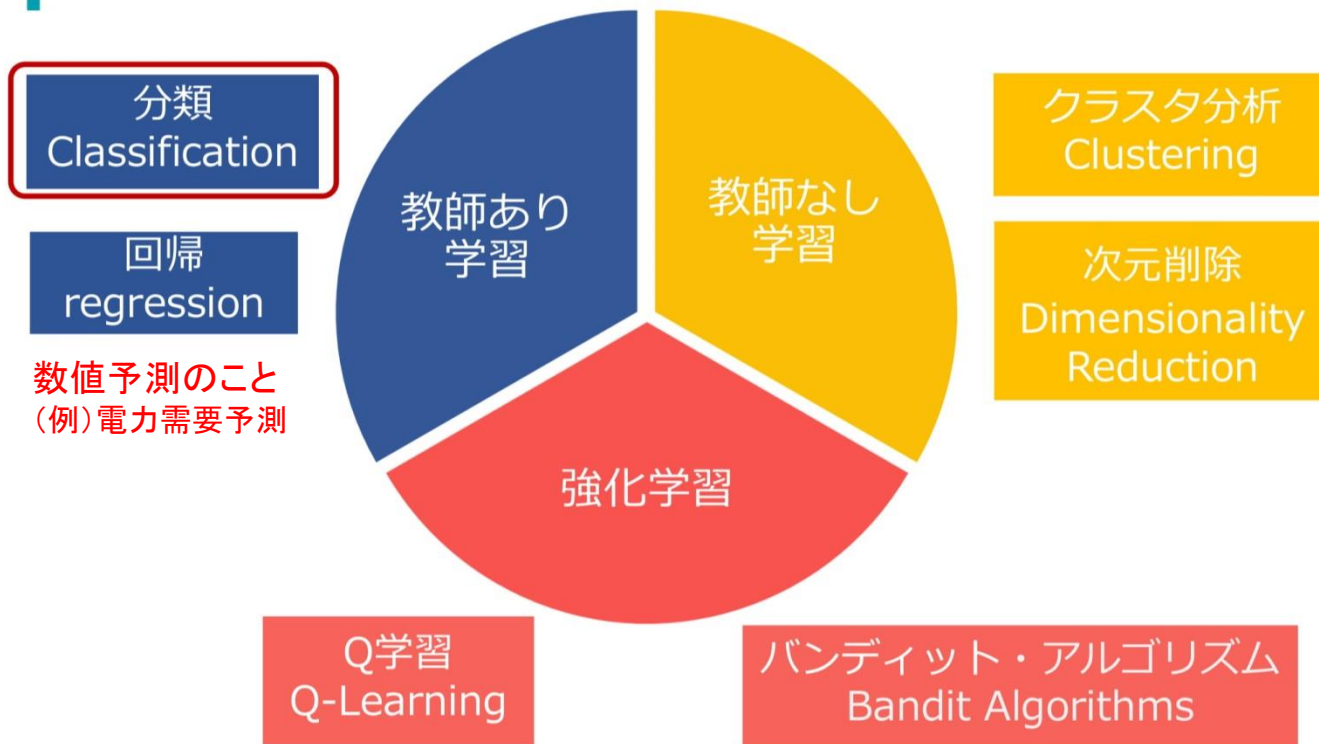
- AI: 人工知能の中核技術であり、コンピュータに人間のような問題解決能力を獲得させるための技術の総称
- 大量のデータ（ビッグデータ）から  
有用な知識を数式として掘り出す  
データマイニングを目的として使用される
- 扱うデータの種類によって、  
「教師つき学習」「教師なし学習」「強化学習」  
の3種類に分類できる



# 機械学習概要

What Machine Learning ?

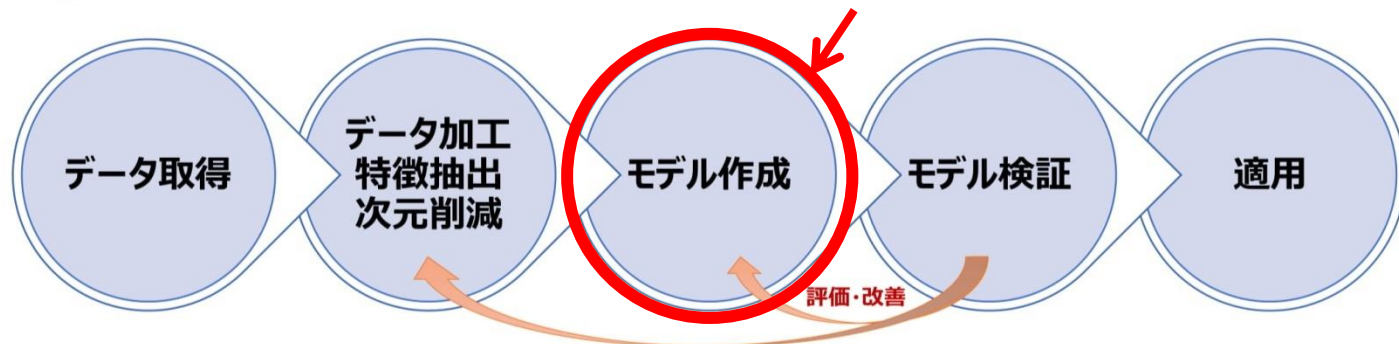
勉強会ではまず  
分類をしっかり  
やります



# 機械学習概要

データ分析の進め方と留意点

ここで各種の機械学習アルゴリズムを使う



1. やりたいこと（目的変数）が明確であること  
正常と異常の定義、判別したいものの定義、求めたい数値
2. データの質と量（説明変数）が十分であること  
関係ないデータや間違い、欠損が少なく、統計処理に量も十分
3. 予測や判別ができたとして、対策が具体化できること  
対策の内容、タイミング、体制、現場の希望を考慮

# 機械学習のアルゴリズム

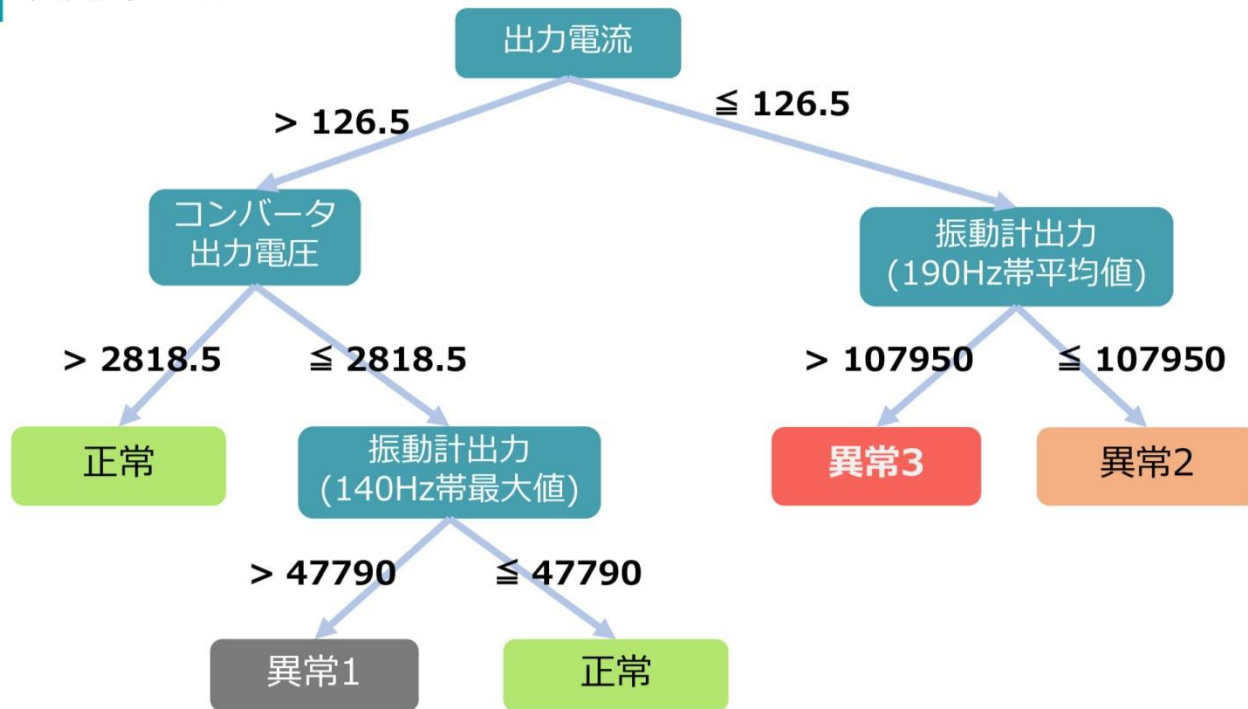
- 線形回帰
- K近傍法
- 決定木
- Gradient Boosting Machine
- ロジスティック回帰
- Support Vector Machine
- Neural Network
- Deep Learning      etc...

本日はこのアルゴリズム  
だけ使います。



# 決定木について

決定木とは





# 決定木について

## 決定木とは

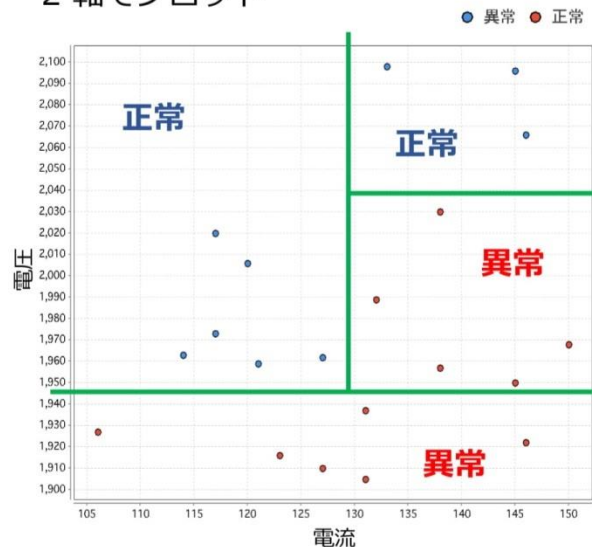
説明変数が 2 つだとすると、

電流	電圧	異常
117	2020	0
138	2030	1
146	2066	0
150	1968	1
117	1973	0
132	1989	1
133	2098	0
106	1927	1
131	1937	1
120	2006	0
127	1962	0
138	1957	1
114	1963	0
145	1950	1
:	:	:

説明変数

目的変数  
0: 正常  
1: 異常

2 軸でプロット



グラフに水平もしくは垂直の  
直線を引く作業に等しい

# 機械学習デモ＋実習

## 今回使用するデータ

- ・統計学者ロナルド・フィッシャーが測定したアヤメの花のデータセット。3種の花にどのような違いがあるだろうか？その違いから、3種の花を分類することはできるだろうか？



**Iris-Versicolor**  
(バージカラー)



**Iris-Setosa**  
(サトサ)



**Iris-Virginica**  
(バージニカ)

変数A1 : sepal length    がく片の長さ(cm)

変数A2 : sepal width    がく片の幅(cm)

変数A3 : petal length    花弁の長さ(cm)

変数A4 : petal width    花弁の幅(cm)

設問： 変数A1～A4(説明変数)を用いて、アヤメの種類(目的変数)を分類する

# 機械学習デモ＋実習

## 今回準備したデータファイル

Iris\_original.csv → オリジナルデータセット  
iris\_train.csv → モデル学習用データ  
Iris\_test.csv → 予測用データ