

# Report on the Experiment

No. 1

Subject Python による機械学習実験 1

Date 2021. 07. 13

Weather 腫れ Temp °C Wet %

Class E5  
Group 4  
Chief  
Partner

No 14  
Name 小畠 一泰

Kure National College of Technology

## 1 目的

機械学習とは、データから学習した結果をもとに、新たなデータに対して判定や予測を行う。すでにさまざまな機械学習のアルゴリズムが開発されている。本実験では、そうした機械学習アルゴリズムについて背景にある理論や特徴を解説した上で、Python プログラミングによる実装を行う。初期の機械学習アルゴリズムから初めて、さまざまな機械学習アルゴリズムを取り上げていく。今後、AI プログラミングの第一歩を踏み出すために、解説を読み進めながら、実際にソースコードを実行して結果を確認しつつ、機械学習アルゴリズムについて理解を深めることを本実験の目的とする。

## 2 考察

### 2.1 各アルゴリズムの決定境界の特徴についてまとめよ

- パーセプトロン
  - 線形モデルであることから決定境界は直線
  - 分類対象を線形分離できない場合は収束しない
- ロジスティック回帰
  - 二値分類のための線形モデルであり決定境界は直線
  - 産業界において最も広く使用されている分類アルゴリズムの 1 つ
- サポートベクトルマシン (SVM)
  - 広く利用されている強力な学習アルゴリズムの 1 つで、パーセプトロンの拡張と見なすことができる
  - マージンを最大化するような直線で分類するため決定境界は直線
- カーネル SVM
  - SVM では非線形分類ができないため「カーネル法」を使うことで非線形分類を可能にしたもの。
    - \* カーネル法: カーネル関数を用いて入力空間を高次元特徴空間へ拡張 (写像) し、線形分離可能なデータに変化させること。
  - カーネル法にて得た分離超平面を逆写像することで決定境界を得ることができる。
- 決定木
  - 意味解釈可能性に優れたモデルであり、一連の質問に基づいて決断を下すことで、データを分類する
  - 特徴空間を矩形に分割するため軸に平行な決定境界が得られる
- ランダムフォレスト
  - 決定木を利用しより複雑な矩形分類が可能となる
  - 決定木同様に軸に平行な決定境界が得られる
- k 近傍法のアルゴリズム
  - 選択された距離指標に基づき、訓練データセットのサンプルの中から分類したいデータ点に最も近い (最も類似する) k 個のサンプルを見つけ出し、新しいデータ点のクラスラベルは、k 個の最近傍での多数決によって決まる
  - k の値を変更すると直線的な決定境界が得られる
  - ユークリッド距離の指標を使用すると、滑らかな決定境界が得られる

## 2.2 パーセプトロンについて調べ動作原理をまとめよ

- パーセプトロンはニューラルネットワークの起源となるアルゴリズム
- 複数の信号を入力として受け取り, 1 つの信号を出力する
- 信号は全て 0 か 1 の 2 値で表される
- 動作原理
  1. 入力信号がニューロンに送られる際に, それぞれに固有の重みが乗算される
  2. 各ニューロンでは, 送られてきた信号の総和が計算される
  3. 総和がしきい値を超えた場合 1 を出力する

## 2.3 カーネルトリックについて説明せよ

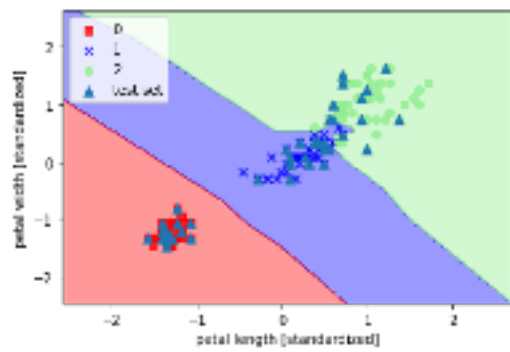
カーネル関数を用いて入力空間を高次元特徴空間へ拡張 (写像) し, 線形分離可能なデータに変化させ, 拡張空間で線形の識別を行う手法のこと

## 2.4 決定木で用いた 3 つの情報利得についてまとめよ

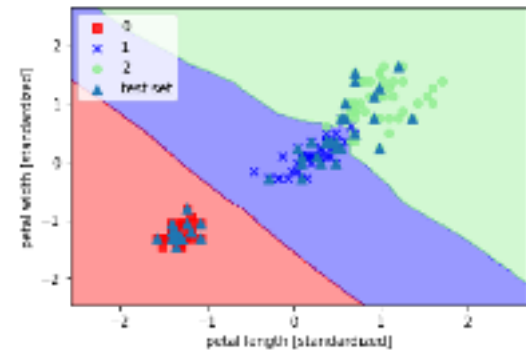
- ジニ不純度
  - ノードごとに, ターゲットがどのくらい分類できていないかを測る指標のこと
  - ノードにおいて完全にサンプルが分類されている場合は, ジニ不純度は 0 になる
- エントロピー
  - 情報の乱雑さを測る指標のこと
  - 情報がランダムだと値が大きくなり, 情報が偏っているときは値が小さくなる
- 分類誤差
  - ノードに含まれるサンプルのある特定のクラスに分類される確率を計算して, それを全体の確率から引いて出た誤差のこと
  - シンプルな方法だが, ノードのクラスの確率の変化にはあまり敏感に反応できないため, 決定木を成長させることには向かない

## 2.5 k 近傍法にて k の値を変化させるとどのような変化があるかを調べよ

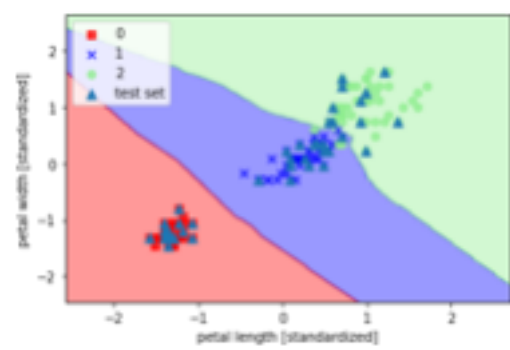
k 近傍法にて, k の値 (n\_neighbors) を 1,3,5,7,9,11,13,15,25 と変化させてグラフを比較して, k の値の変化に対して決定境界にどのような変化があるかを調べた. n\_neighbors を順に変化させ, それぞれ出力された図を次に示す. k の値が小さいときはクラス 1,2 の境目がくっきりと分かれているが, 大きくなるにつれて, 境界が直線に近くなり, 境界が明確になっていない.



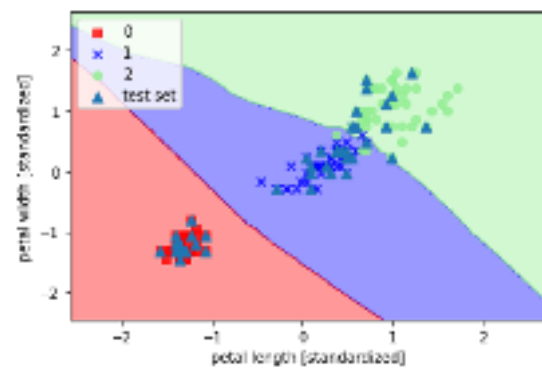
$k = 1$



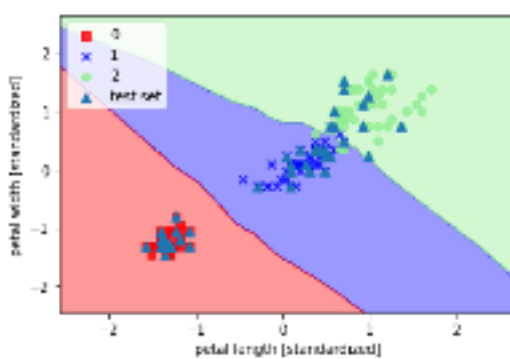
$k = 3$



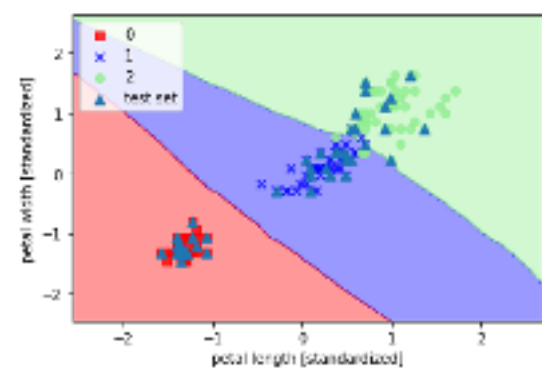
$k = 5$



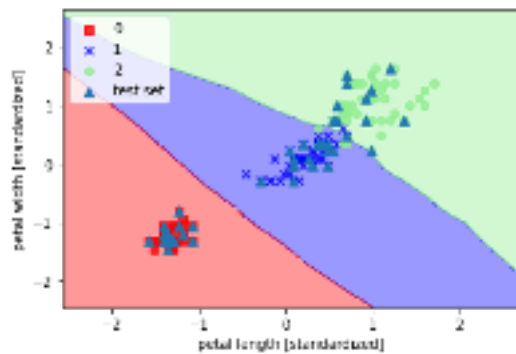
$k = 7$



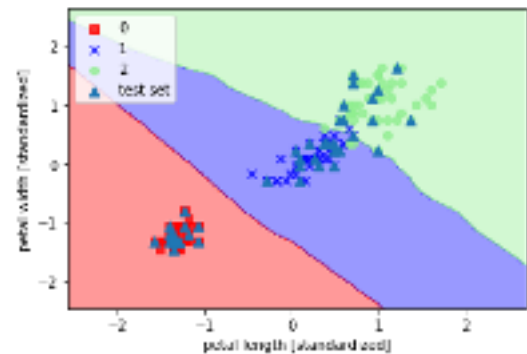
$k = 9$



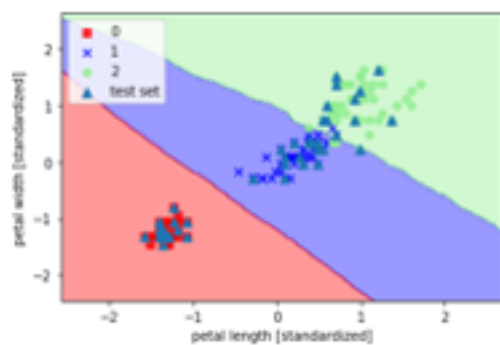
$k = 11$



k = 13



k = 15



k = 25

## 2.6 本実験を通して、機械学習アルゴリズムについて特に印象に残ったことをまとめよ

単に分類といってもさまざまな分類方法があることがわかった。特にカーネル法が興味深く、線形に分類できない際に写像を用いて空間を拡張し線形分類可能にする発想には驚いた。

## 3 参考文献

1. 斎藤康毅著, ゼロから作る Deep Learning —Python で学ぶディープラーニングの理論と実装, オライリージャパン (2018)
2. <https://yolo-kiyoshi.com/2019/09/16/post-1226/>
3. <https://qiita.com/renesisu727/items/1c6014c909202ff089f3>
4. <https://ja.wikipedia.org/wiki/カーネル法>
5. <https://spjai.com/regression-tree/>