

機械学習とその応用

ー従来手法から深層学習までー

Safety Network Japan 第92回定例会
2019.02.15

高橋 聖 | 日本大学理工学部 応用情報工学科

本日の講演

- 機械学習とは？
 - 実現したいタスク
 - 学習の3つの種類
- 機械学習の手法
 - 従来手法から深層学習まで
- 機械学習の応用
 - 深層学習（ディープラーニング）の応用
- 機械学習のツール
 - プログラミング言語 Python（パイソン）

2

機械学習とは？

3

機械学習で何をしたいのか？

機械学習で実現したい三つのタスク

- **回帰分析** (regression analysis)
 - 数値を予測したい
- **分類** (classification)
 - クラス判定したい
- **クラスタリング** (clustering)
 - データをグループ化したい

機械学習の3つの種類

- **教師あり学習**（主に分類問題）
 - 学習データに望ましい出力（ラベル）が判明している
- **教師なし学習**（主にクラスタリング問題）
 - 学習データに望ましい出力（ラベル）がない
- **強化学習**（主にプランニング問題）
 - エージェントが環境（報酬を含む）とのやりとりをする

機械学習で何をしたいのか？

機械学習で実現したい三つのタスク

- **回帰分析** (regression analysis)
 - 数値を予測したい
 - **分類** (classification)
 - クラス判定したい
 - **クラスタリング** (clustering)
 - データをグループ化したい
- 教師あり学習
- 教師なし学習

機械学習の手法

機械学習のいろいろな手法

- **回帰分析**
 - **最小二乗法**
- **分類**
 - パーセプトロン
 - SVM（サポートベクトルマシン）
 - 決定木学習
 - k-近傍法（k-NN）
 - ディープラーニング
- **クラスタリング**
 - k-平均法（k-means）

多項式近似と誤差関数

次の多項式を想定

$$f(x) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M$$
$$= \sum_{m=0}^M w_mx^m$$

誤差： $f(x)$ と観測値 t の差の2乗

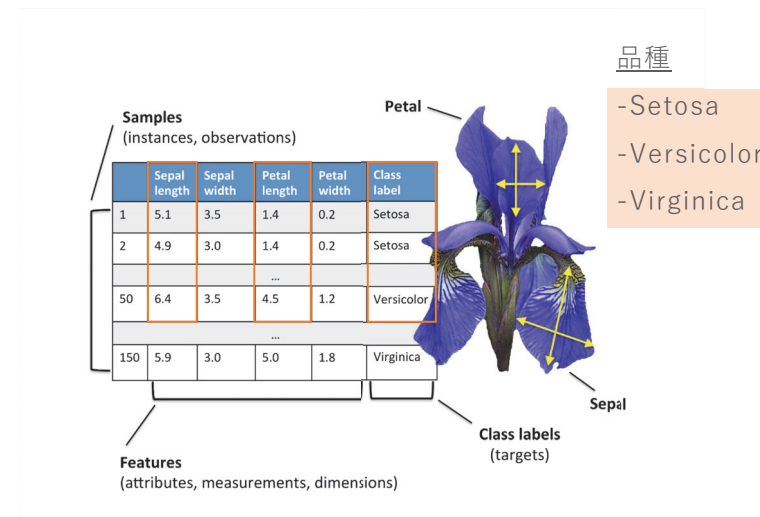
$$\{f(x_1) - t_1\}^2 + \{f(x_2) - t_2\}^2 + \dots + \{f(x_{10}) - t_{10}\}^2$$

誤差関数

$$E_D = \frac{1}{2} \sum_{n=1}^N \{f(x_n) - t_n\}^2 \rightarrow \text{誤差関数を最小にしたい}$$

分類

花びらとガクの長さでアヤメの品種を分類

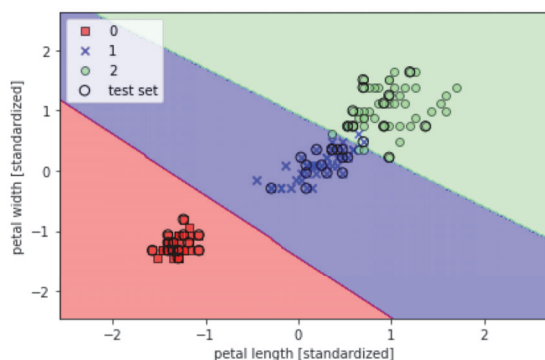


出典 | <http://github.com/rasbt/python-machine-learning-book.git>

分類

機械学習による分類結果の一例

0: Setosa
1: Versicolor
2: Virginica



機械学習のいろいろな手法

回帰分析

- 最小二乗法

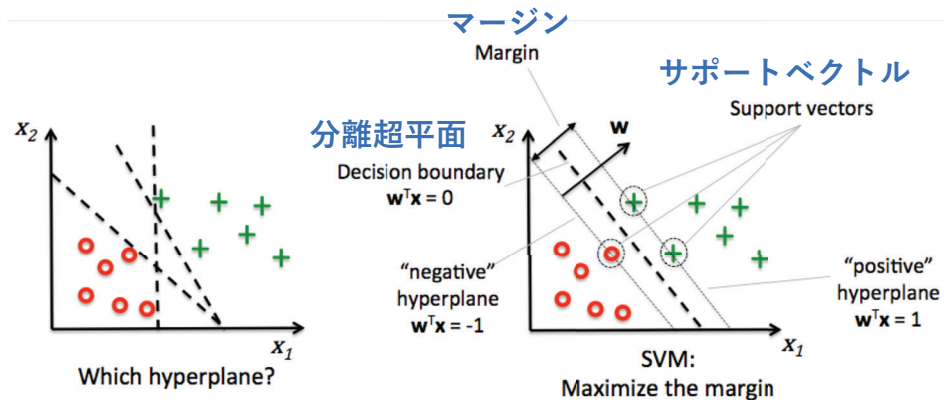
分類

- パーセプトロン
- SVM (サポートベクトルマシン)
- 決定木学習
- k-近傍法 (k-NN)
- ディープラーニング

クラスタリング

- k-平均法 (k-means)

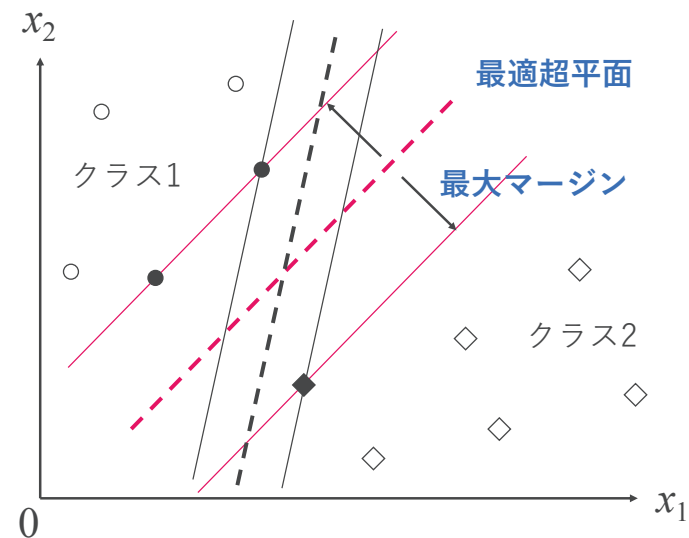
SVMによる最大マージン分類



出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

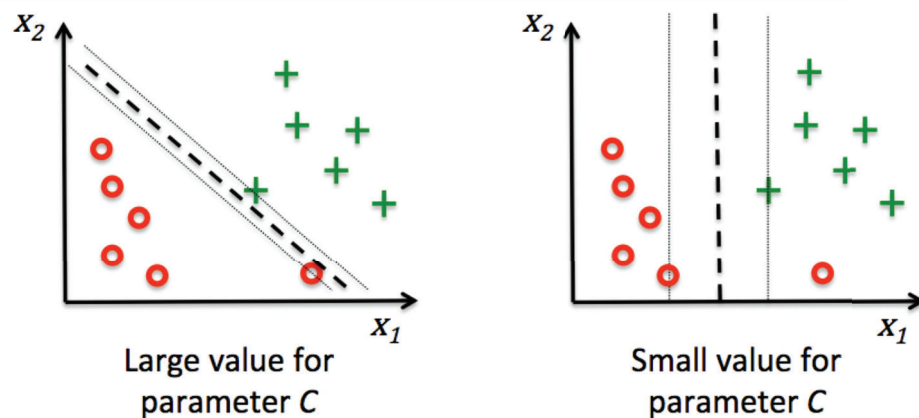
13

2次元空間における最適超平面



14

マージンパラメータ C でマージンの幅を制御



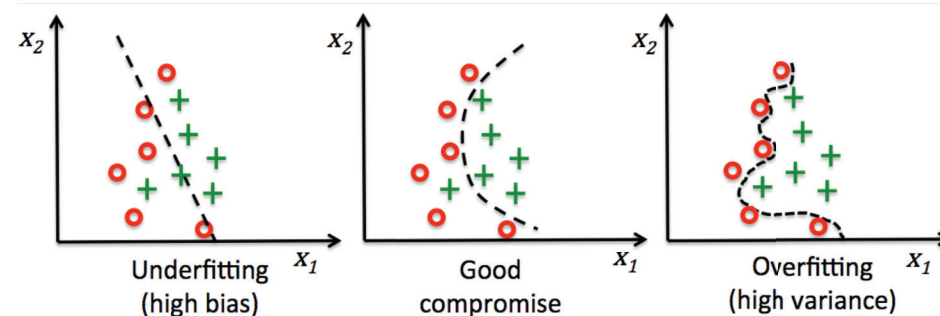
C が大きければ教師データの認識率上がる
(大きすぎると過学習)

C が小さいければ教師データの誤分類に寛大
(汎化能力が上がる)

出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

15

過学習と学習不足



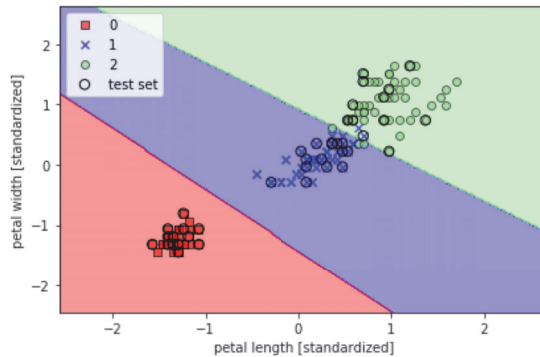
出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

16

SVMによる分類結果

機械学習による分類結果の一例

0: Setosa
1: Versicolor
2: Virginica



出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

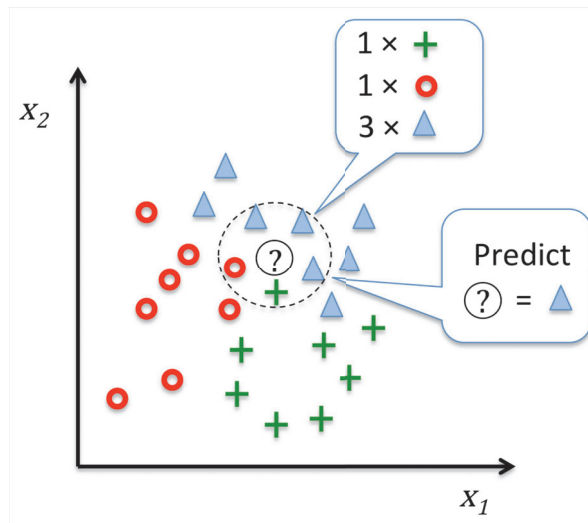
17

機械学習のいろいろな手法

- 回帰分析
 - 最小二乗法
- 分類
 - パーセプトロン
 - SVM (サポートベクトルマシン)
 - 決定木学習
 - k-近傍法 (k-NN)
 - ディープラーニング
- クラスタリング
 - k-平均法 (k-means)

18

最近傍のデータから多数決でクラスを決定



出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

19

k-NNアルゴリズム

1. kの値と距離指標を選択する
2. 分類したいサンプルからk個の最近傍のデータ点を見つけ出す
3. 多数決によりクラスラベルを割り当てる

20

距離指標の選択も重要

Minkowski距離

$$d(x^{(i)}, x^{(j)}) = \sqrt[p]{\sum_k |x_k^{(i)} - x_k^{(j)}|^p}$$

$p=1$: マンハッタン距離

$$d(x^{(i)}, x^{(j)}) = \sum_k |x_k^{(i)} - x_k^{(j)}|$$

$p=2$: ユークリッド距離

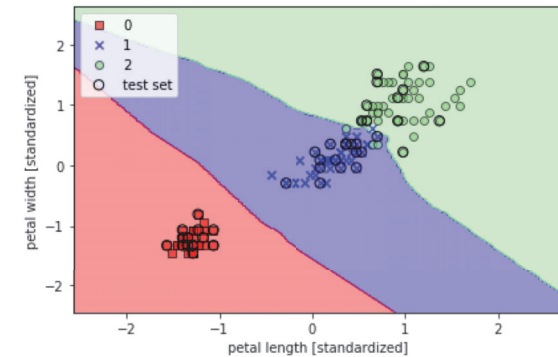
$$d(x^{(i)}, x^{(j)}) = \sqrt{\sum_k |x_k^{(i)} - x_k^{(j)}|^2}$$

21

k-近傍法による分類結果

$k=5$, ユークリッド距離

0: Setosa
1: Versicolor
2: Virginica



出典 | <https://github.com/rasbt/python-machine-learning-book/tree/master/code>

22

機械学習のいろいろな手法

回帰分析

- 最小二乗法

分類

- パーセプトロン
- SVM (サポートベクトルマシン)
- 決定木学習
- k-近傍法 (k-NN)
- ディープラーニング

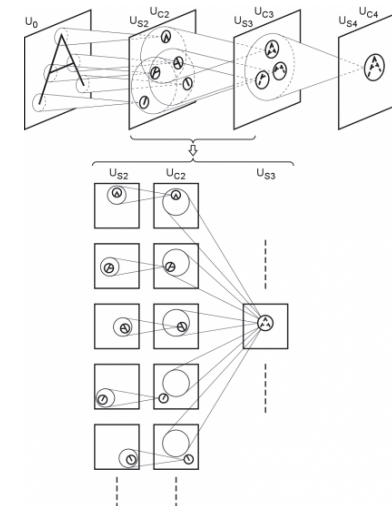
クラスタリング

- k-平均法 (k-means)

23

ニューラルネット「ネオコグニトロン」

福島邦彦によるディープラーニングの先駆的研究



出典 | <http://dbnst.nii.ac.jp/pro/detail/498>

機械学習のいろいろな手法

- 回帰分析
 - 最小二乗法
- 分類
 - パーセプトロン
 - SVM (サポートベクトルマシン)
 - 決定木学習
 - k-近傍法 (k-NN)
 - ディープラーニング
- クラスタリング
 - k-平均法 (k-means)

25

機械学習の応用

26

機械学習のツール

27

Pythonで機械学習



- Numpy : ベクトルや行列を扱う数値計算ライブラリー
- SciPy : 科学計算用ライブラリー
- matplotlib : グラフ描画ライブラリー
- pandas : Rに類似のデータフレームを提供するライブラリー
- scikit-learn : 機械学習用ライブラリー
- lpython : 対話的操作環境を提供するツール
- Jupyter notebook : Pythonの対話型実行環境をより便利に拡張した環境

ANACONDA

Pythonとパッケージをインストール

クラウドで実行される Jupyter notebook環境

- <https://colab.research.google.com/notebooks/welcome.ipynb>