

Scikit-learn

1.4. Nearest Neighbors

1.4.1. Unsupervised Nearest Neighbors

1.4.1.1. Finding the Nearest Neighbors

1.4.1.2. KDTree and BallTree Classes

永田 純平

1.4. Nearest Neighbors

最近傍法

- ノンパラメトリックな手法
- 学習データとの距離が一番近いものからクラスを識別する
- 距離の測定はユークリッド距離を用いるのが一般的
- 学習データが多いほど精度が高いが実行に時間がかかる

1.4. Nearest Neighbors

K近傍法

- 最近傍から順にK個の近傍事例を選び、そのクラスの多数決でクラスを決定する
- 学習データを全探索するため実行に時間がかかる
- 適切なkを与える必要があり、クロスバリデーションなどを利用して適切な値を設定する。

※k=1だと最近傍法となる

1.4.1. Unsupervised Nearest Neighbors

- 教師なしの最近傍学習では
['auto', 'ball_tree', 'kd_tree', 'brute']の中から
最近傍探索のアルゴリズムを選択する。

```
>>> from sklearn.neighbors import NearestNeighbors
>>> import numpy as np
>>> X = np.array([[ -1, -1], [ -2, -1], [ -3, -2], [ 1, 1], [ 2, 1], [ 3, 2]])
>>> nbrs = NearestNeighbors(n_neighbors=2, algorithm='ball_tree').fit(X)
>>> distances, indices = nbrs.kneighbors(X)
>>> indices
array([[0, 1],
       [1, 0],
       [2, 1],
       [3, 4],
       [4, 3],
       [5, 4]]...)
>>> distances
array([[ 0.          ,  1.          ],
       [ 0.          ,  1.          ],
       [ 0.          ,  1.41421356],
       [ 0.          ,  1.          ],
       [ 0.          ,  1.          ],
       [ 0.          ,  1.41421356]])
```