

k 近傍法

$\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ をラベル付けされたデータとし、 N をデータの数、 \mathbf{x}_i を D 次元特徴ベクトル、 y_i を \mathbf{x}_i のラベルとする。ただし y_i の値は $1, \dots, n$ のいずれかであるとする。 k を自然数とする。以下のアルゴリズムを用いて、未知の D 次元特徴ベクトル \mathbf{x} に対して、 \mathbf{x} のラベル y を予測する。このアルゴリズムを k 近傍法という。

- (1) \mathbf{x} と最も距離の近い k 個のデータ $\mathbf{x}_{m_1}, \dots, \mathbf{x}_{m_k}$ を取る。
- (2) y_{m_1}, \dots, y_{m_k} の中で最も多い値を \mathbf{x} に割り当てる。

例 0.1. (コードは k 近傍法.ipynb)

#データを生成

```
import numpy as np
```

```
X = np.array([[ -0.1], [ 0.7], [ 1.0], [ 1.6], [ 2.0], [ 2.5], [ 3.3]])
```

```
y = np.array([0, 0, 1, 1, 1, 0, 0])
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
neigh = KNeighborsClassifier(n_neighbors=3)
```

```
neigh.fit(X,y) #k=3 として、k 近傍法を適用
```

```
neigh.predict([[0.8]]) #値を予測
```

```
>array([1])
```

参考文献

- [1] Andriy Burkov. (2019). The hundred-page machine learning book.
- [2] Marc Peter Deisenroth., A. Aldo Faisal., Cheng Soon Ong. (2020). Mathematics for machine learning. Cambridge University Press.
- [3] Aurélien Geron. (2019). Hands-on machine learning with Scikit-Learn, Keras & TensorFlow. 2nd Edition. Oreilly.
- [4] 小縣信也., 斎藤翔汰., 溝口聡., 若杉一幸. (2021). ディープラーニング E 資格エンジニア問題集 インプレス.
- [5] Sebastian Raschka., Vahid Mirjalili. (2019). Python machine learning. Third Edition. Packt.