k 平均法

$X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ を $\mathbb{R}^m$ の部分集合とし、k を自然数とする。以下のアルゴリズムを用いて、X を k 個のクラスターへと分類する。このアルゴリズムを k **平均法**という。

(1) k 個のデータ $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k} \in X$ を取る。
(2) 任意のデータ $\mathbf{x} \in X$ に対して、$\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k}$ の中で $\mathbf{x}$ に最も距離が近い $\mathbf{x}_{i_j}$（$1 \leq j \leq k$）を取る。
(3) $C(x_{i_j}) = \{\mathbf{x} \in X \mid \mathbf{x}_{i_j}$ は $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k}$ の中で $\mathbf{x}$ に最も距離が近い $\}$ とおき、x_{i_j} を $\frac{1}{|C(x_{i_j})|} \sum_{\mathbf{x} \in C(x_{i_j})} \mathbf{x}$ で置き換える（ただし、$|C(x_{i_j})|$ は $C(x_{i_j})$ の元の個数を表す）。
(4) (1) に戻る。

例 0.1. (コードは k 平均法.ipynb)

```
#データを生成
import numpy as np
X = np.array([[-2.7], [-1.3], [0.7], [3.5], [5.1]])

from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=2) #k=2 として、k 平均法を適用
y_pred = kmeans.fit(X).predict(X) #X をクラスタリングする
print(y_pred)
>[0 0 0 1 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] Aurëlien Gëron. (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.