

k - Najbliższych Sąsiadów

(1)

- nieparametryczna metoda klasyfikacji
- nowy obiekt przyporządkowany jest do tej klasy, do której należy k najbliższych sąsiadów



Elementy alg. k - Najbliższych Sąsiadów

Ranking / Wskaźnik jakości /

$$\text{ranking}(z_{\text{nowy}}, C_i) = \sum_{z(k) \in KNN} \text{dist}(z_{\text{nowy}}, z(k)) \cdot \delta(z(k), C_i)$$

gdzie:

- KNN to zbiór najbliższych sąsiadów nowego elementu z_{nowy}
- $z(k)$ pochodzi ze zbioru $\overset{\text{danych}}{z(k)}$ $z_k = (z(1), z(2), \dots)$
- C_i pochodzi ze zbioru klas $C = \{c_1, c_2, \dots, m\}$

• $\delta()$ to funkcja

(2)

$$\delta(z(k), c_i) = \begin{cases} 1 & \text{jeżeli } z(k) \in c_i \\ 0 & \text{jeżeli } z(k) \notin c_i \end{cases}$$

• $\text{dist}()$ to funkcja

$$\text{dist}(z, z(k)) = \sqrt{\sum (z^{(i)} - z^{(i)}(k))^2}$$

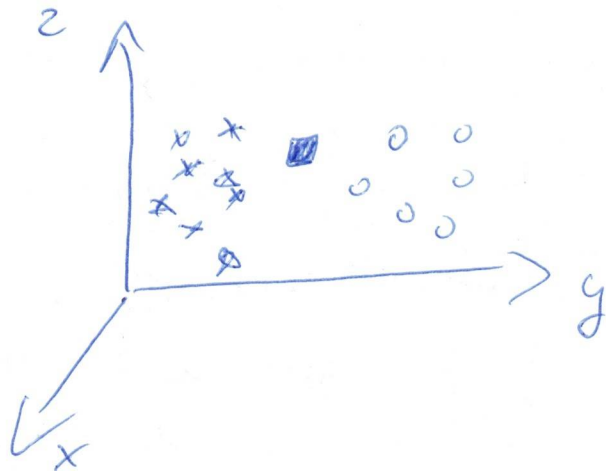
W metodzie k -NN klasa ze zbioru C wybierana jest na podstawie rozwiązania następującego zadania optymalizacyjnego

$$C = \arg \min_{c \in C} \text{ranking}(z_{\text{nowy}}, c_i) =$$

$$= \arg \min_{c \in C} \left[\sum_{z(k) \in k\text{-NN}} \text{dist}(z_{\text{nowy}}, z(k)) \delta(z(k), c_i) \right]$$

Zadanie do wykonania

1. Samodzielnie uzupełnić przedstawione metody k -NN w języku Python
2. Weryfikację metody na syntetycznym zbiorze danych.



3

3. Porównanie samodzielnie opracowanej metody k -NN z metodą dostępną w bibliotece sklearn