

Języki i Biblioteki Analizy Danych

Laboratorium 9.: Algorytm kNN

mgr inż. Zbigniew Kaleta

Algorytm k najbliższych sąsiadów (ang. k-nearest neighbors; kNN)

- jeden z najprostszych i najbardziej intuicyjnych algorytmów uczenia maszynowego
- etap "treningu" polega po prostu na zapamiętaniu wszystkich przypadków uczących

Wymaga:

- ustalenia parametru k
- ustalenia metryki w przestrzeni obserwacji

Etap predykcji polega na obliczeniu odległości przypadku badanego do wszystkich przypadków uczących.

Następnie spośród wszystkich przypadków uczących wybieramy k najbliższych w sensie wybranej metryki.

Badany przypadek otrzymuje taką etykietę, jaka dominuje wśród wybranych w poprzednim kroku przypadków.

Numer

0

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16
17
18
19
20
21
22
23
24
25
26
27
28
29

Przypadek testowy: wzrost = 1.68 m, waga = 51 kg

Najbliżsi sąsiedzi: próbki 19 (niedowaga), 1 (niedowaga), 9 (ok), 16 (niedowaga), 11 (ok)

Czy to są dobrze przygotowane dane?

Lektura dodatkowa:

- <https://www.ibm.com/topics/knn>
- https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm