

Implementacja i badanie skuteczności algorytmu k najbliższych sąsiadów (k-NN)

Konspekt

Kinga Pilch, Hubert Rosiak

27 lutego 2018

Krótki opis algorytmu

Algorytm k najbliższych sąsiadów (ang. *k nearest neighbours*) jest prostym algorytmem służącym do rozwiązywania problemu klasyfikacji lub regresji. Idea algorytmu jest następująca: mając zbiór danych uczących w pewnej przestrzeni, dla nowej, nieznanej wcześniej obserwacji prognozujemy wartość będącą średnią arytmetyczną wartości jej k najbliższych (według pewnej metryki, najczęściej euklidesowej) sąsiadów ze zbioru uczącego. Gdy mamy do czynienia z problemem klasyfikacji, nowej obserwacji przypisujemy klasę najliczniej występującą wśród jej k najbliższych sąsiadów.

Algorytm k najbliższych sąsiadów można łatwo modyfikować. Wybór parametru k może w znaczący sposób wpływać na zachowanie algorytmu. Podobnie zamiast średniej arytmetycznej można zastosować inny sposób wyboru nowej wartości/klasy na podstawie informacji uzyskanych od sąsiadów - jedną z popularnych modyfikacji tego rodzaju jest zastosowanie średniej ważonej. Używana w algorytmie metryka służąca obliczaniu odległości między obserwacjami również może podlegać zmianie.

Wybrany język programowania

Projekt stworzony zostanie w języku Python.

Literatura

- [1] Sahibsingh A. Dudani, *The Distance-Weighted k-Nearest-Neighbour Rule*, IEEE Transactions on Systems, Man, and Cybernetics (Vol: SMC-6, Issue:4, April 1976)