

Metody Rozpoznawania Obrazu

KNN algorithm

Grzegorz Miejski

1. Wykorzystaliśmy algorytm KNN do klasyfikacji danych. Algorytm przetestowałem na zbiorze danych dostępnym na stronie:

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>

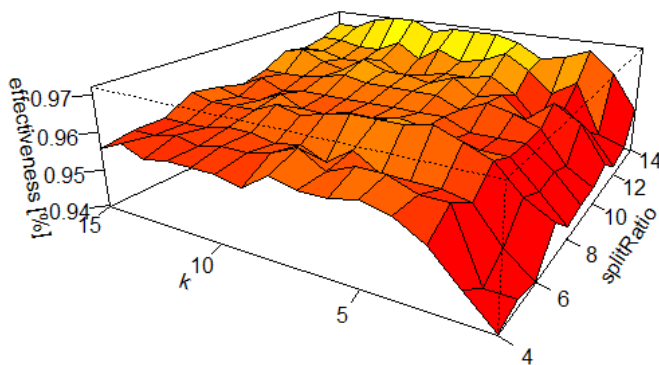
Wykorzystując wektor cech starałem się sklasyfikować przypadek raka jako złośliwy lub nie.

Sprawiłem skuteczność procentową klasyfikacji za pomocą tego algorytmu dla różnych parametrów:

- a) K - ilość sąsiadów wykorzystanych do sklasyfikowania konkretnego wektora cech
- b) Split ratio - Relację wielkościową pomiędzy wielkościami training i test setu

Aby sprawdzić średnią skuteczność klasyfikacji najpierw wybierałem stosunek w jakim mają być ze sobą ilość danych train setu do test setu (co określałem parametrem split ratio). Następnie dzieliłem dane wejściowe na tyle części, gdzie jeden podzbiór zawsze był moim test setem, a pozostałe (split ratio – 1) danych stanowiły mój training set. Następnie obliczałem średnią skuteczność klasyfikacji na test secie dla każdego podziału.

Wynik Przedstawiłem na wykresie 3d:



Wnioski :

- Można zauważyć, że wpływ wartości parametru k jest znaczący jeżeli bierzemy pod uwagę jego małe wartości (1-5). Następnie jego wpływ stabilizuje się.
- Dzieląc cały zbiór danych na większą ilość fragmentów uzyskiwałem lepsze skuteczności.

Dodatkowo przygotowałem również porównanie źle sklasyfikowanych punktów (kolor żółty) na tle pozostałych danych z wykorzystaniem PCA:

