

Laboratorium Rozpoznawania Obrazów - Ćwiczenie #1

Klasyfikacja najbliższego sąsiedztwa

Terminy ćwiczeń: **25.02 28.02 4.03 7.03**

Państwa zadaniem jest przygotowanie klasyfikatora najbliższego sąsiedztwa rozróżniającego dwa gatunki kosaćców (*iris*). Dane są zawarte w plikach `iris2.txt` i `iris3.txt`, zawierających odpowiednio pomierzone cechy kwiatów z gatunku *iris versicolor* i *iris virginica*.

Każda linia pliku zawiera:

1. Kolejny numer próbki
2. Długość zielonej części okrywającej kwiat
3. Szerokość zielonej części okrywającej kwiat
4. Długość płatka
5. Szerokość płatka

Sam klasyfikator jest bardzo prosty (byłbym zdziwiony, gdyby jego implementacja była dłuższa niż 10 linii kodu w Octave):

1. Mając punkt do sklasyfikowania, trzeba wyznaczyć jego odległości do wszystkich punktów w zbiorze uczącym – tutaj warto przygotować oddzielną funkcję, żeby sam klasyfikator (jakkolwiek prosty by nie był, nie zależał od metryki, którą chcemy wykorzystać w klasyfikacji).
2. Wybrać próbkę (a dokładniej etykietę klasy próbki) o najmniejszej odległości jako decyzję klasyfikacyjną.

Prawdziwa praca czeka Państwa przy ocenie klasyfikatora. Ponieważ mamy tylko 100 próbek, nie możemy podzielić tego zbioru na część uczącą i testującą bez poważnych konsekwencji dla jakości klasyfikatora. Aby zmierzyć oczekiwaną współczynnik błędu klasyfikatora skorzystacie Państwo z metody *leave-one-out* (znanej także jako metoda *jackknife* – czyli scyzoryk):

1. Wybieramy do klasyfikacji jedną próbkę ze zbioru uczącego.
2. Resztę zbioru uczącego wykorzystujemy do skonstruowania klasyfikatora (zostało 99 próbek, więc klasyfikator będzie bardzo podobny do końcowego, używającego wszystkich 100 próbek).
3. Wybrana w kroku 1-szym próbka jest klasyfikowana, a wynik zapisywany.
4. Kroki 1-3 są powtarzane dla wszystkich 100 próbek w zbiorze uczącym. Zakumulowane wyniki klasyfikacji są zupełnie dobrym przybliżeniem oczekiwanej jakości końcowego klasyfikatora.

Proszę ocenić jakość klasyfikacji dla 2, 3 i 4 cech. Dla 2 i 3 cech proszę użyć po 2 różne zestawy cech.

Pasjonaci rozpoznawania obrazów mogą wybrać *najlepsze* 2 i 3 elementowe zestawy cech.