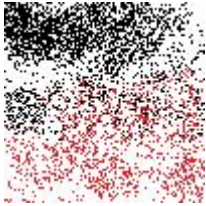


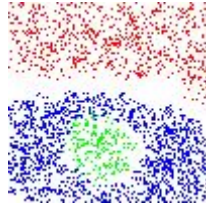
Anna Marciniak

Podstawy Uczenia Maszynowego lab 2 - **Metric Learning** **Raport**

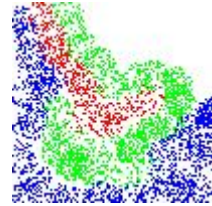
Na potrzeby zadania stworzyłam trzy zbiory danych korzystając z metody 'PAINT'.
Stworzyłam obrazki o wymiarach 100px na 100px.



'dataset 1'



'dataset 2'



'dataset 3'

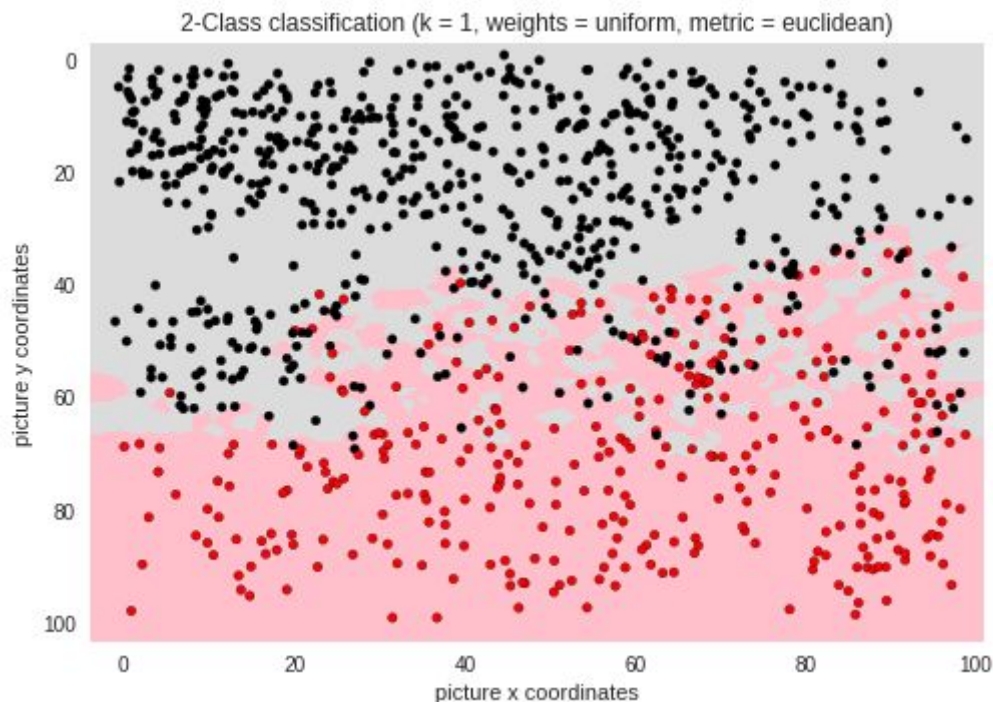
Następnie nałożyłam na nie szum, dodałam randomową liczbę z rozkładu normalnego.
Podzieliłam te zbiory na testowe i treningowe (70%, 30% wybrane randomowo).

Dla zbioru 1:

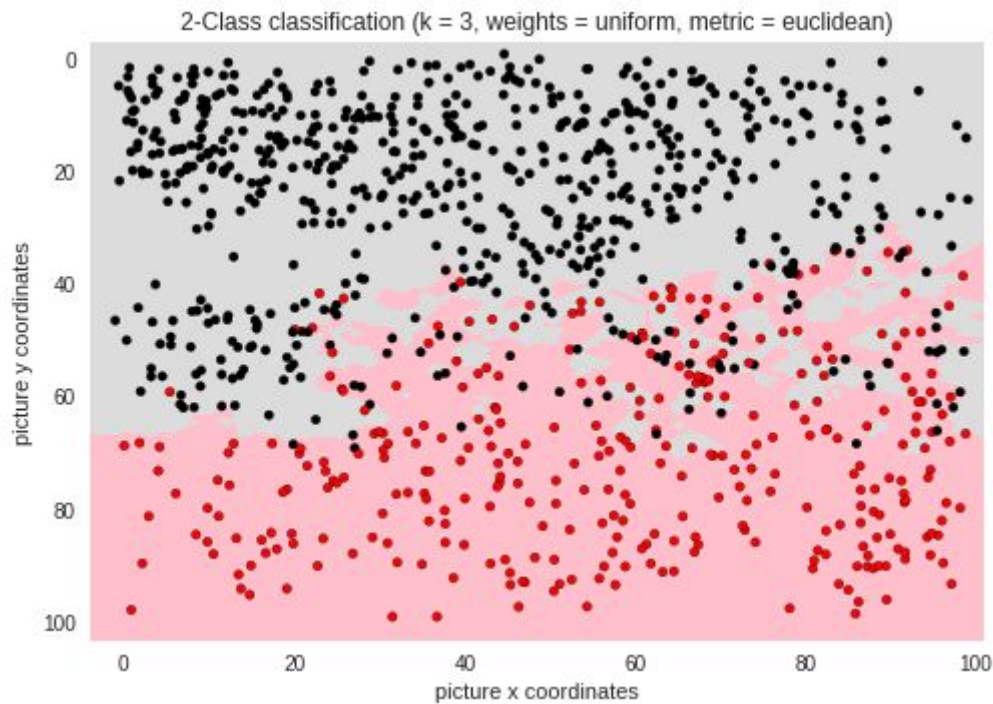
klasyfikator k-NN z $k=1$, głosowaniem większościowym i metryką Euklidesa

poprawność klasyfikatora: 89,69%, błąd 10,31%.

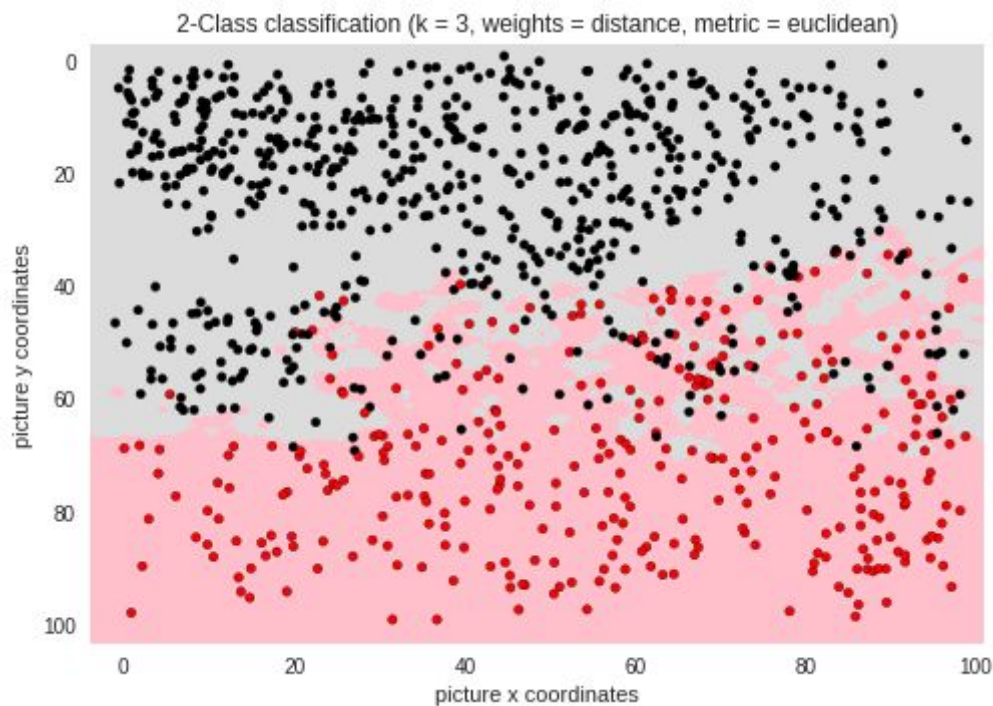
Kolor jasno czerwony(różowy) oznacza obszar dla czerwonych pikseli, natomiast szary - dla czarnych, miejsce, gdzie się kolory spotykają to granica decyzyjna. Naniesione punkty testowe, do zwizualnienia poprawności klasyfikatora.



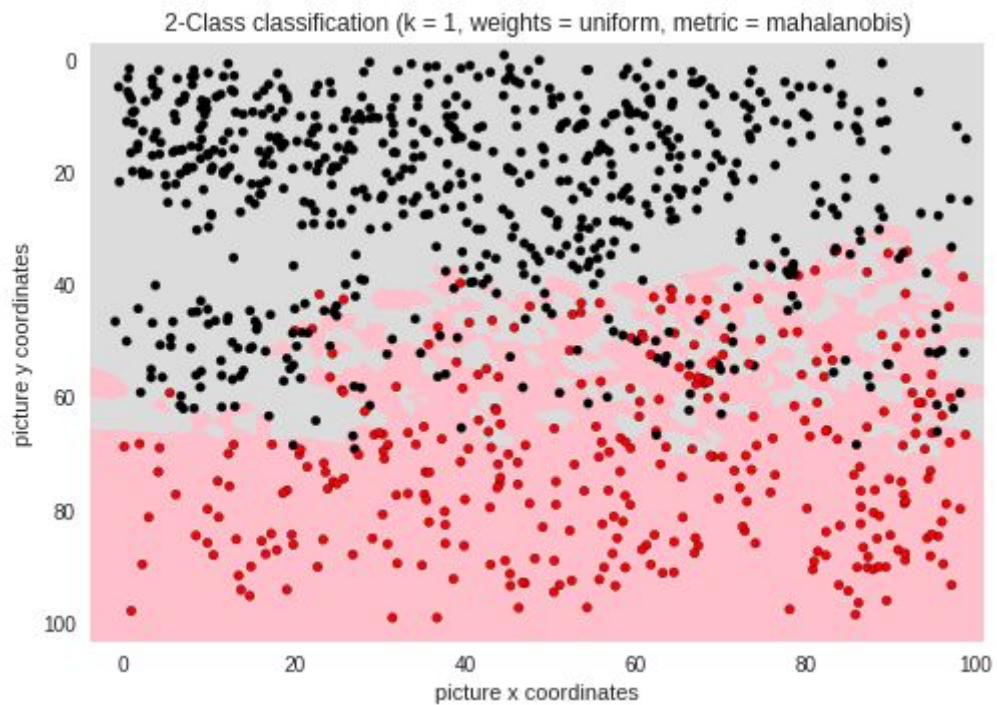
k-NN z $k=3$, głosowaniem większościowym i metryką Euklidesa
poprawność klasyfikatora: 89,38%, błąd 10,62%.



k-NN z $k=3$, głosowaniem ważonym odległością i metryką Euklidesa
poprawność klasyfikatora: 89,90%, błąd 10,10%.

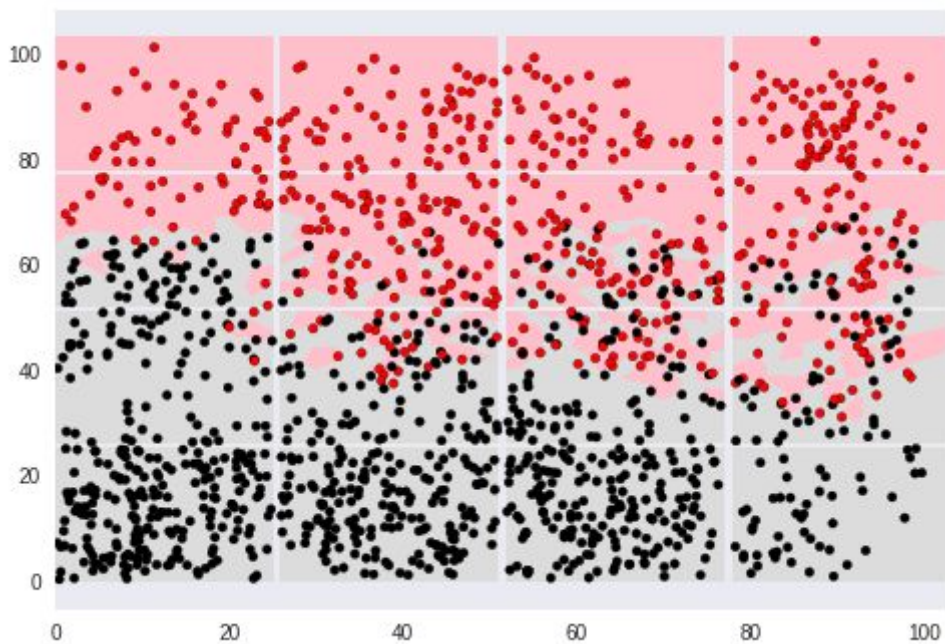


k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa
poprawność klasyfikatora: 89,38%, błąd 10,62%.



k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa zależną od regionu

średnia poprawność klasyfikatora: 88,36%, błąd 11,64%.

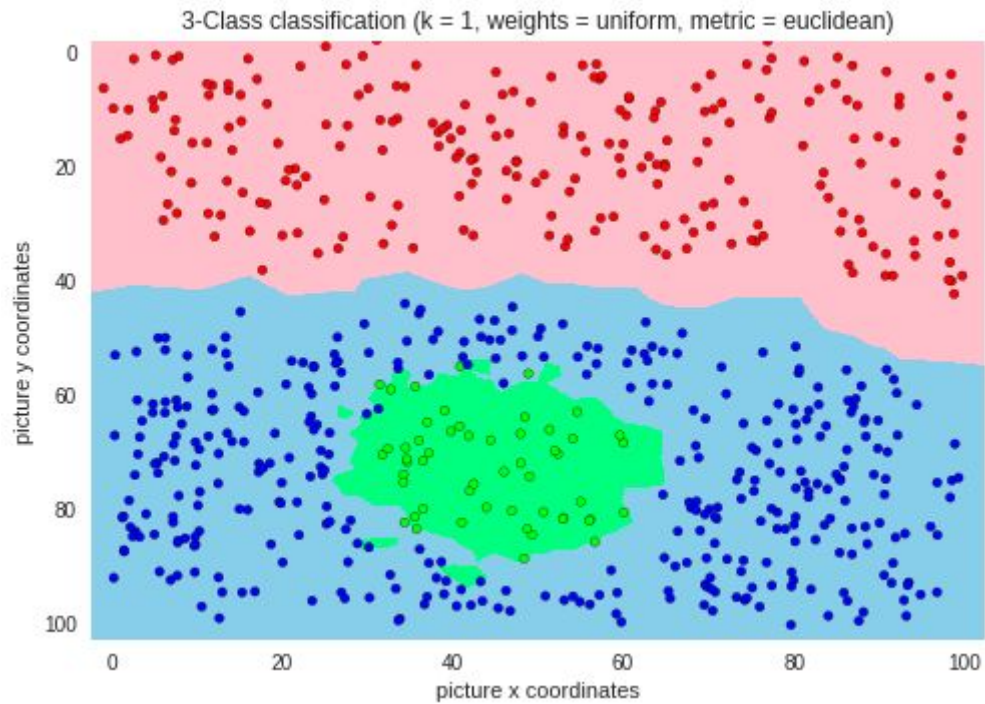


(plot odwrócony (lewy górny róg obrazka to punkt (0,0))

Dla zbioru 2:

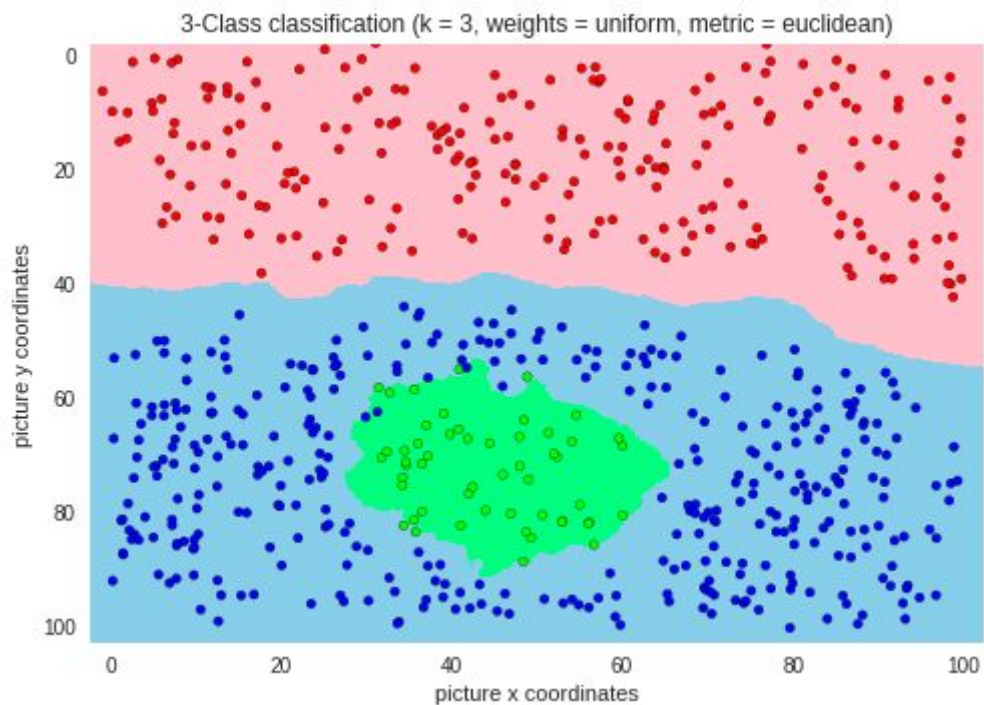
klasyfikator k-NN z $k=1$, głosowaniem większościowym i metryką Euklidesa

poprawność klasyfikatora: 98,43%, błąd 1,57%

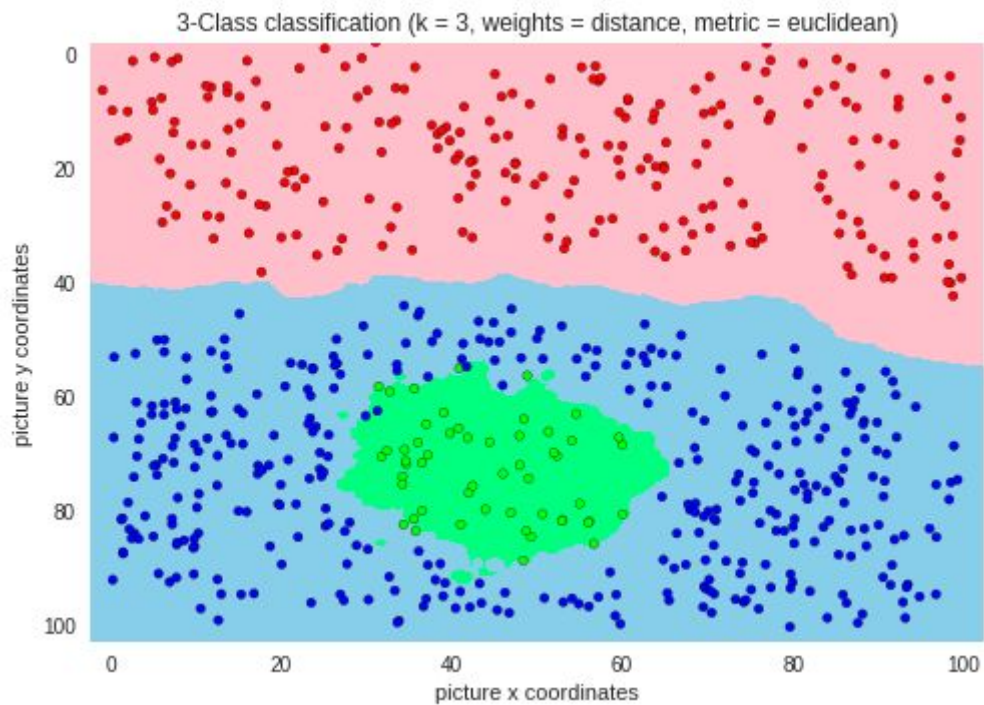


k-NN z $k=3$, głosowaniem większościowym i metryką Euklidesa

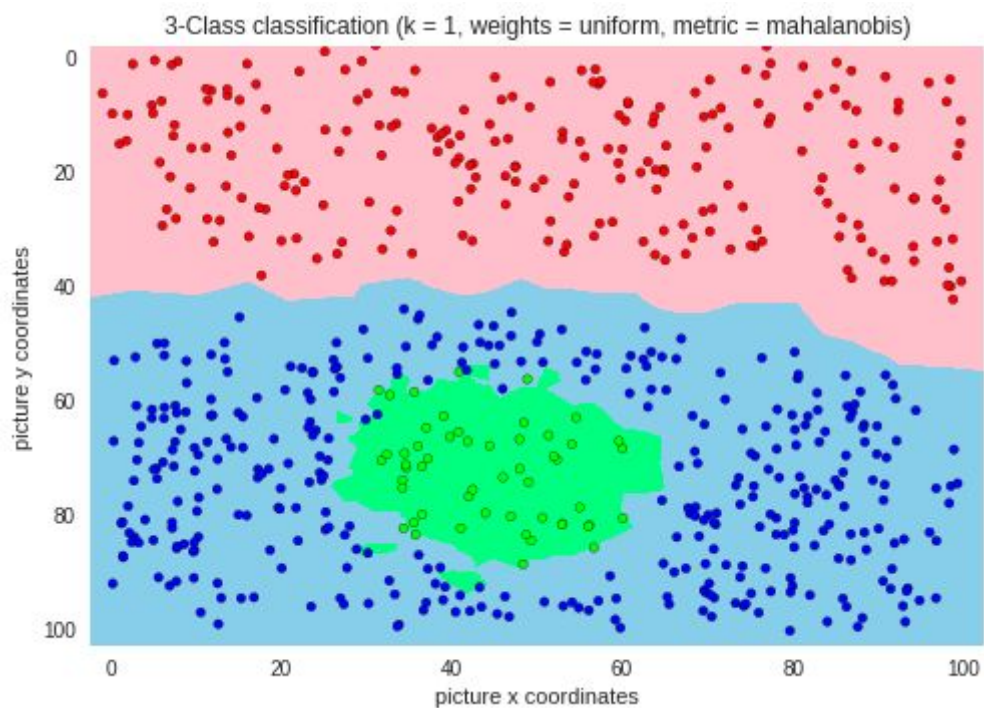
poprawność klasyfikatora: 98,59%, błąd 1,41%



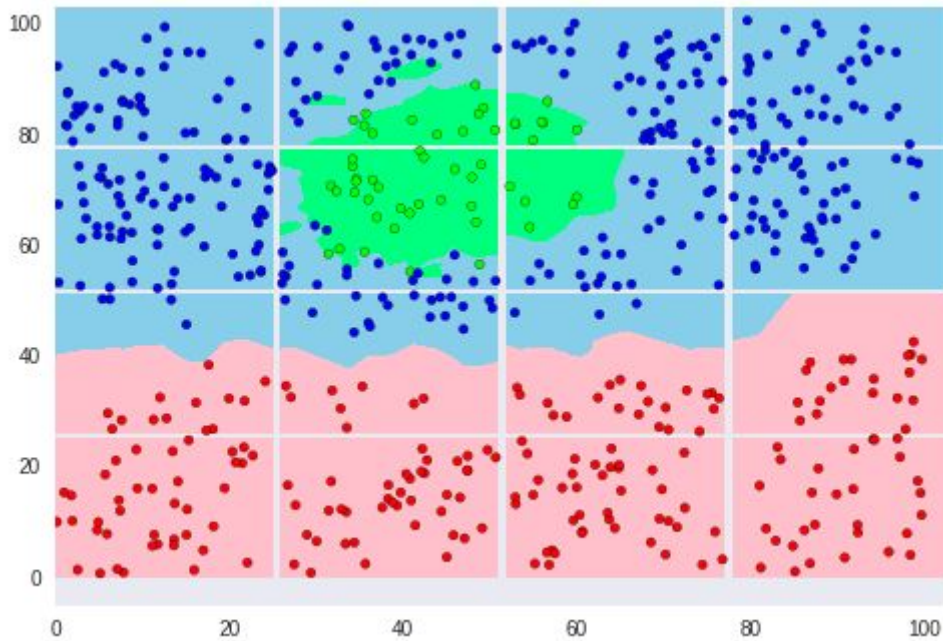
k-NN z $k=3$, głosowaniem ważonym odległością i metryką Euklidesa
poprawność klasyfikatora: 98,59%, błąd 1,41%



k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa
poprawność klasyfikatora: 98,43%, błąd 1,57%



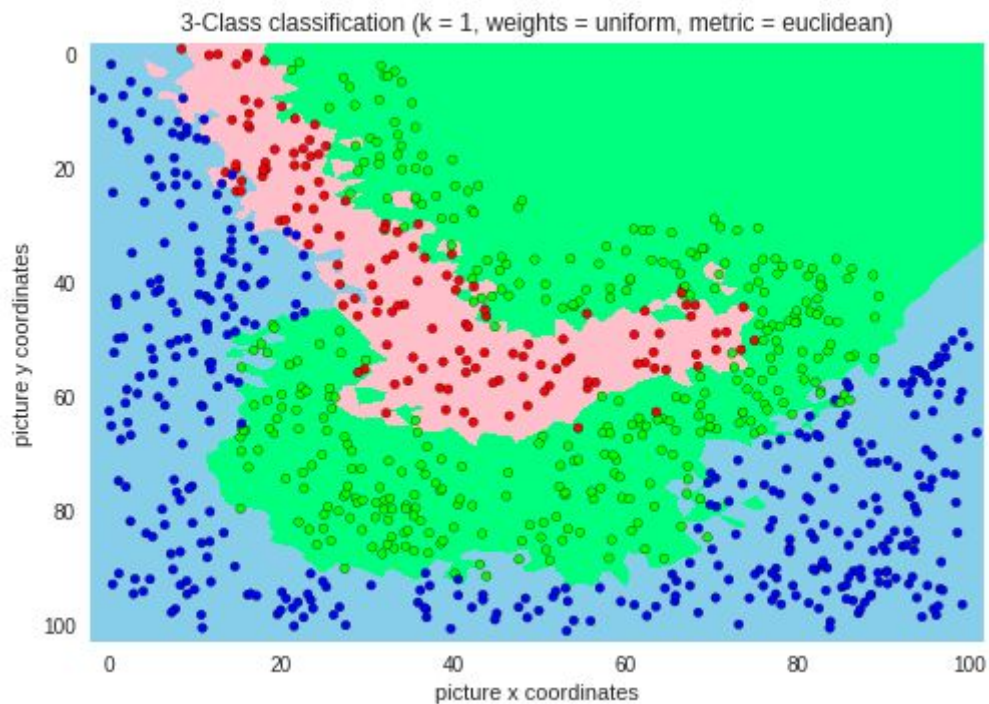
k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa zależną od regionu
średnia poprawność klasyfikatora: 98,42%, błąd 1,78%.



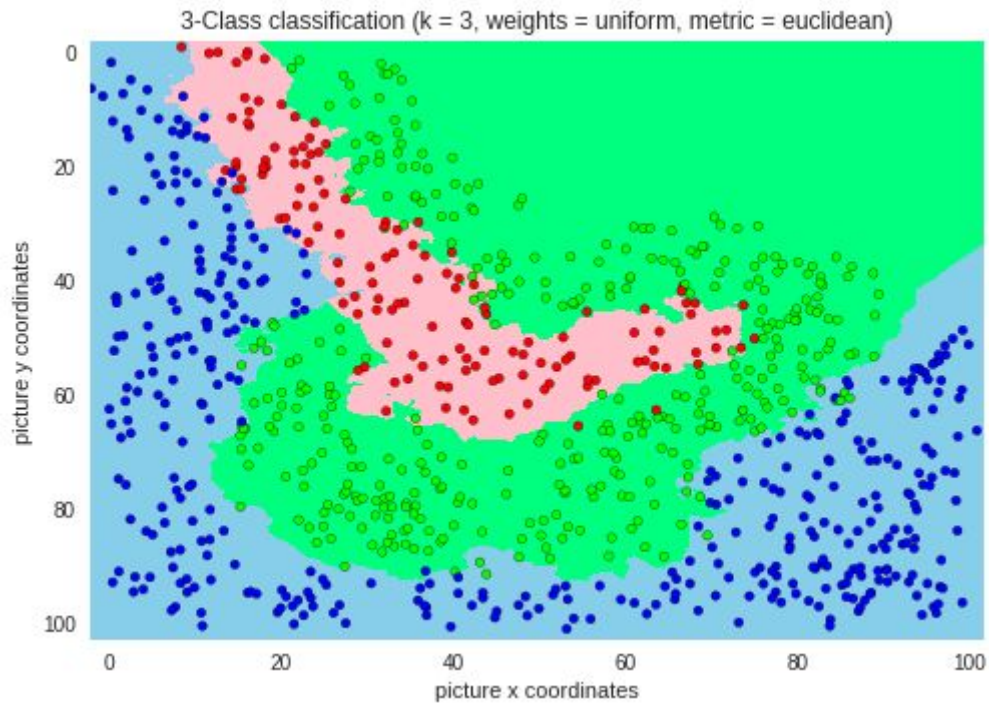
(plot odwrócony (lewy górny róg obrazka to punkt (0,0))

Dla zbioru 3:

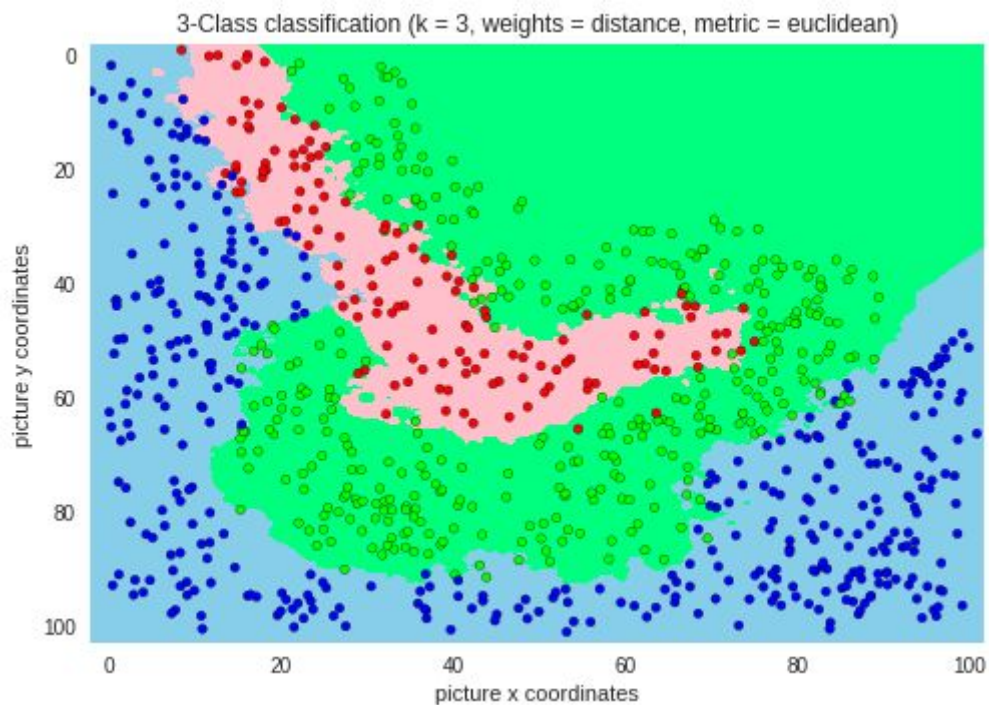
k-NN z $k=1$, głosowaniem większościowym i metryką Euklidesa.
poprawność klasyfikatora: 92,43%, błąd 7,57%



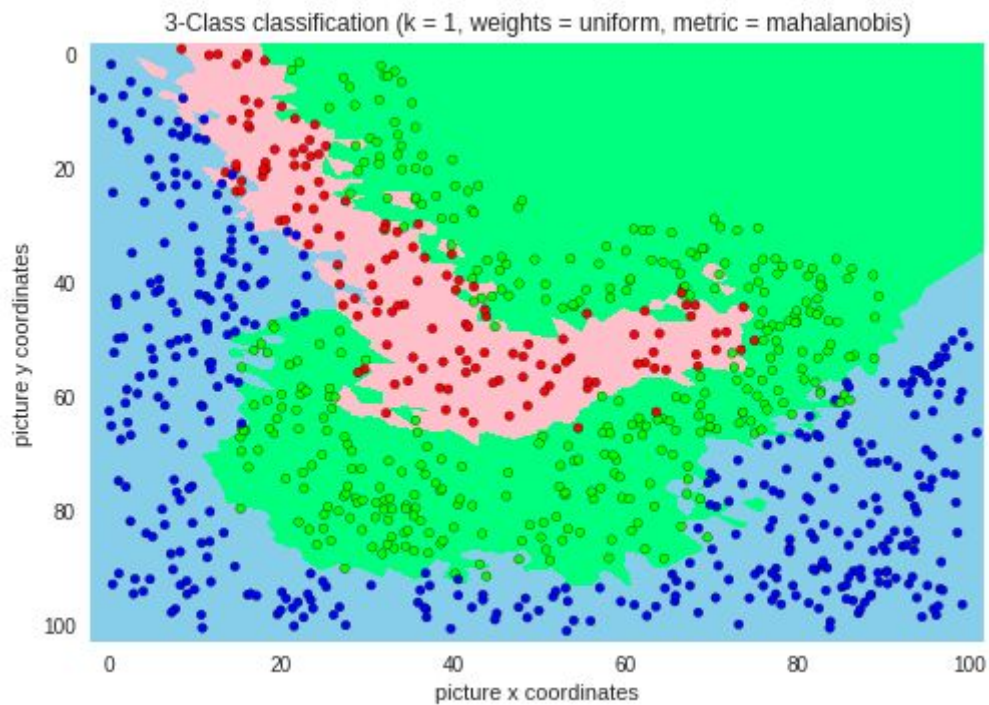
k-NN z $k=3$, głosowaniem większościowym i metryką Euklidesa
poprawność klasyfikatora: 94,01%, błąd 5,99%



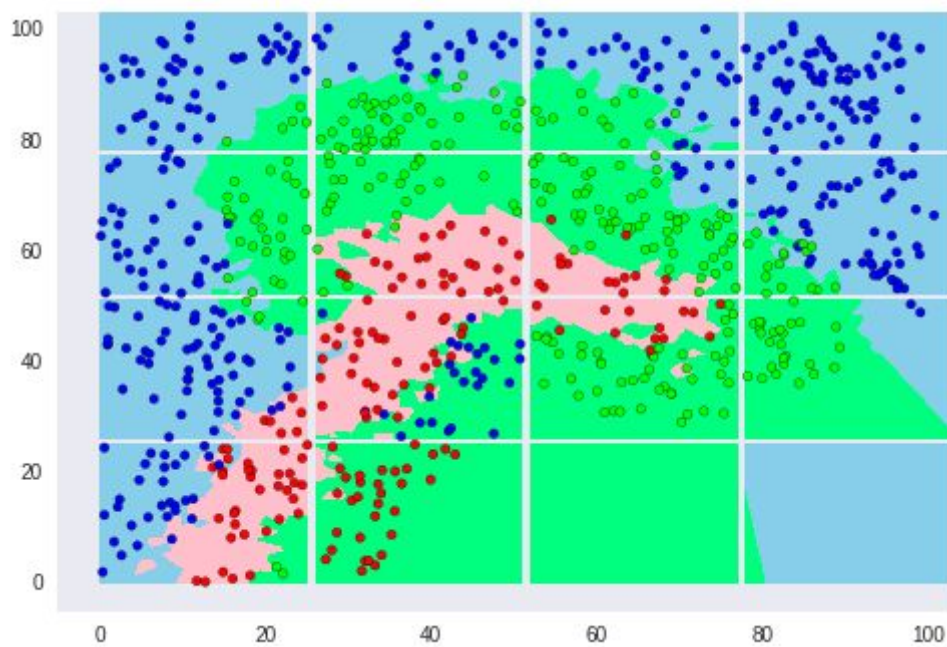
k-NN z $k=3$, głosowaniem ważonym odległością i metryką Euklidesa
poprawność klasyfikatora: 93,79%, błąd 6,21%



k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa
poprawność klasyfikatora: 92,20%, błąd 7,80%



k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa zależną od regionu
średnia poprawność klasyfikatora: 81,43%, błąd 18,56%



(plot odwrócony (lewy górny róg obrazka to punkt (0,0)))

Wnioski:

- Niezależnie od zastosowanego klasyfikatora, dla zbioru pierwszego błąd wynosi najwięcej, jest to prawdopodobnie za sprawą tego, że granica decyzyjna jest mocno przemieszana.
- Zależnie od zbioru głosowanie przez dystans czy większość w metryce euklidesa dawały przywilej, więc jednoznacznie nie może powiedzieć, które jest lepsze, więc należy wybierać adekwatnie do posiadanych danych.
- Dla $k=3$ można zauważyć, że granica decyzyjna nie tworzy tak wielu poszarpanych "wysepek", oraz granica jest gładsza.
- W żadnym przypadku klasyfikator z metryką mahalanobisa nie był lepszy od klasyfikatora z metryką euklidesa - gdy nie dzieliłam na obszary.
- Średni błąd klasyfikatora z metryką mahalanobisa zależną od jego regionu jest dość duży, ale należy zwrócić uwagę, że w niektórych obszarach dopasował dane w 100% natomiast znalazły się też obszary gdzie błąd sięgał ponad 50%.