

# Laboratorium 7

## Klasyfikacja metodą najbliższych sąsiadów

(3+2+1)

Użyjemy zestawu danych z zadania 3. Mamy trzy klasy danych *virginica*, *versicolor* i *setosa*. Każdy zestaw ma 50 egzemplarzy opisanych czterema atrybutami.

1. Tworzymy zbiór trenujący zawierający po 30 egzemplarzy każdej z klas, pozostałe dane posłużą jako zbiór testowy.  
Tworzymy obiekt klasyfikujący *ClassificationKNN.fit* (MATLAB 2012) Sprawdzamy ilość poprawnych i błędnych klasyfikacji na zbiorze trenującym i na zbiorze testującym  
Rysujemy kilka dwuwymiarowych rysunków ( $\text{atrybut}_i$ - $\text{atrybut}_j$ ), zaznaczając dla każdej z klas, przypadki treningowe, testujące poprawne i testujące błędne.
2. Powtarzamy wiele razy wybór podzbioru treningowego i testującego, i obliczamy średnią liczbę błędnych klasyfikacji (dla liczby sąsiadów  $k = 1, 3, 4$ ).
3. To samo sprawdzamy komendami *resubLoss*, *crossval*, *kfoldLoss*. Ten punkt wykonujemy biorąc wszystkie dane jako treningowe, a zmieniamy liczbę sąsiadów w algorytmie i/lub definicję odległości z euklidesowej na Mahalanobisa.