

Uwaga: W środowisku R metoda svm jest zaimplementowana w następujących pakietach:

- pakiet: `e1071`, funkcja: `svm` (interfejs dla pakietu `libsvm`, język C, Chang and Lin, 2001),
- pakiet: `kernlab`, funkcja: `ksvm`,
- pakiet: `klaR`, funkcja: `svmlight`,
- pakiet: `svmlight`.

## 7.1

a) Dane liniowo separowalne: wygeneruj następujące dane:

```
data1 = seq(1,10,by=2)
classes1 = c('a','a','a','b','b')
test1 = seq(1,10,by=2) + 1
```

Zastosuj metodę Maszyn Wektorów Nośnych (SVM) z jądrem liniowym używając funkcji `svm(e1071)`. Wyznacz tabelę reklasyfikacji oraz dokonaj predykcji na zbiorze `test1`.

b) Dane liniowo nieseparowalne: wygeneruj następujące dane:

```
data2 <- seq(1,10)
classes2 <- c('b','b','b','a','a','a','a','b','b','b')
```

Zastosuj metodę Maszyn Wektorów Nośnych (SVM) z jądrem liniowym, oblicz tabelę reklasyfikacji. Zmień jądro na jądro radialne i porównaj wyniki.

## 7.2

a) Narysuj jądro radialne (definicja: `?svm`) biorąc jako argument  $|u - v|$ . Jak zmienia się kształt wykresu przy zmianie parametru `gamma=0.1, 1, 5`?

b) Wczytaj dane `cats(MASS)`. Zastosuj metodę SVM z jądrem radialnym (etykietą klas jest zmienna `Sex`). Wyznacz krzywą rozdzielałą klasy. Jak zmienia się jej kształt przy zmianie parametru `gamma=0.1, 1, 5`?

c) Zastosuj metodę SVM z jądrem wielomianowym stopnia  $d$ . Wyznacz krzywą rozdzielałą klasy. Jak zmienia się jej kształt przy zmianie stopnia wielomianu  $d = 1, 3, 5$ ?

## 7.3

Dane `breast-cancer-wisconsin.data` zawiera informację o 699 pacjentkach z podejrzeniem nowotworu piersi. Celem analizy jest stwierdzenie czy dany guz jest złośliwy (zmienna `Class="malignant"`) czy łagodny (zmienna `Class="benign"`). W analizie należy pominąć pierwszą zmienną (`Samplecodenumber`).

Podzielić zbiór danych na dwa zbiory: treningowy (pierwsze 400 obserwacji) oraz testowy (pozostałe obserwacje).

a) Zastosuj metodę SVM z parametrami domyślnymi i oblicz frakcję poprawnych klasyfikacji na zbiorze testowym, precyzję i czułość.

b) Stosując krosvalidację 10-krotną na zbiorze treningowym wyznacz optymalne wartości parametrów `gamma` oraz `cost` (współczynnik kary za niespełnienie "idealnych organiczeń"). Skorzy-

staj z funkcji: `tune.svm` oraz `tune.control`. Oblicz frakcję poprawnych klasyfikacji na zbiorze testowym, precyzję i czułość.

#### 7.4

Program Weka. Na wybranym zbiorze danych porównaj działanie metody svm dla różnych wartości parametrów. Na wybranym zbiorze danych porównaj działanie svm z metodami: drzewo klasyfikacyjne, model logistyczny, metoda najbliższego sąsiada, naiwny klasyfikator Bayesa stosując krosvalidację.