

Laboratorium 9 (MASZYNY WEKTORÓW NOŚNYCH)

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

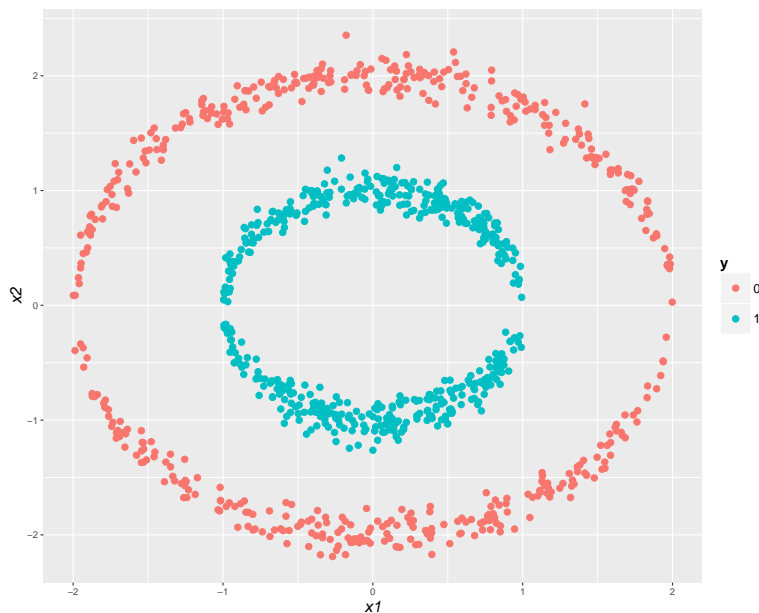
- pakiet: `e1071`, funkcja: `svm` (interfejs dla pakietu `libsvm`, język C, Chang and Lin, 2001),
- pakiet: `kernlab`, funkcja: `ksvm`,
- pakiet: `klaR`, funkcja: `svmlight`,
- pakiet: `svmlight`.

9.1

Wygeneruj dane w następujący sposób.

- Rozważamy dwie klasy $y = 1$ i $y = 0$, o licznosciach 500 obserwacji w każdej.
- Rozważamy $p = 2$ zmienne ilościowe x_1 i x_2 .
- Niech $x_1^2 + x_2^2 = 1$, w klasie $y = 1$, gdzie $x_1 \sim U[-1, 1]$.
- Niech $x_1^2 + x_2^2 = 4$, w klasie $y = 0$, gdzie $x_1 \sim U[-2, 2]$.
- Generując x_1 i x_2 dodaj niewielki szum.

Poniższy rysunek pokazuje przykładowe dane. Przetestuj działanie metody SVM z różnymi ją-



drami: liniowym, radialnym, wielomianowym.

9.2

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