

AMO - Projekt 2, zestaw danych saheart

Jakub Postępski

17 stycznia 2019

1 Zadanie optymalizacji

Metoda SVM pozwala na klasyfikację binarną danych. W metodzie należy zminimalizować normę $\|\vec{w}\|$ opisujących hiperpłaszczyznę oddzielającą dane dla zbioru danych uczących. W projekcie przyjęto twardy margines pomiędzy danymi.

Zadanie prymalne można zapisać jako:

$$\min \|\vec{w}\|$$

przy ograniczeniach:

$$y_i(\vec{w} * \vec{x}_i - b)$$

Dla $i \in 1..n$ gdzie n to ilość próbek mamy: \vec{x}_i to wsp. kolejnych próbek a $y_i \in -1, 1$ to wartości klasyfikatorów.

Zadanie dualne można zapisać jako:

$$\min \|f(c_i)\| = - \sum_{i=1}^n + \frac{1}{2} \sum i = 1n \sum_{j=1}^n y_i c_i x_i x_j y_j c_j$$

przy ograniczeniach:

$$\sum_{i=1}^n c_i y_i = 0$$
$$c_i \geq 0$$

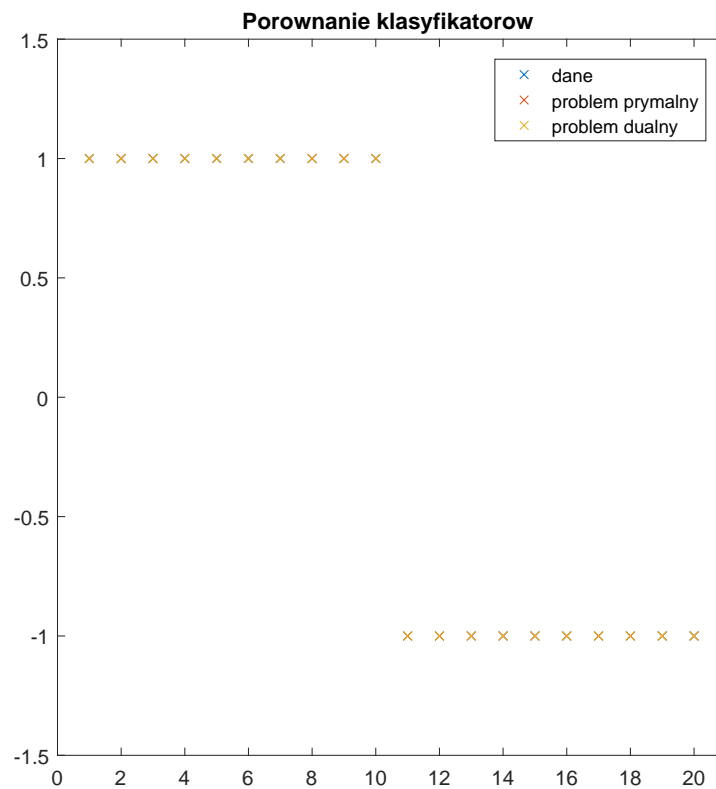
W celu uzyskania klasyfikacji należy obliczyć:

$$\vec{w} \vec{x}_i - b$$

Do implementacji problemu odpowiednio prymalnego i dualnego wykorzystano funkcje *fmincon* i *quadprog* w programie Matlab.

2 Testowe dane losowe

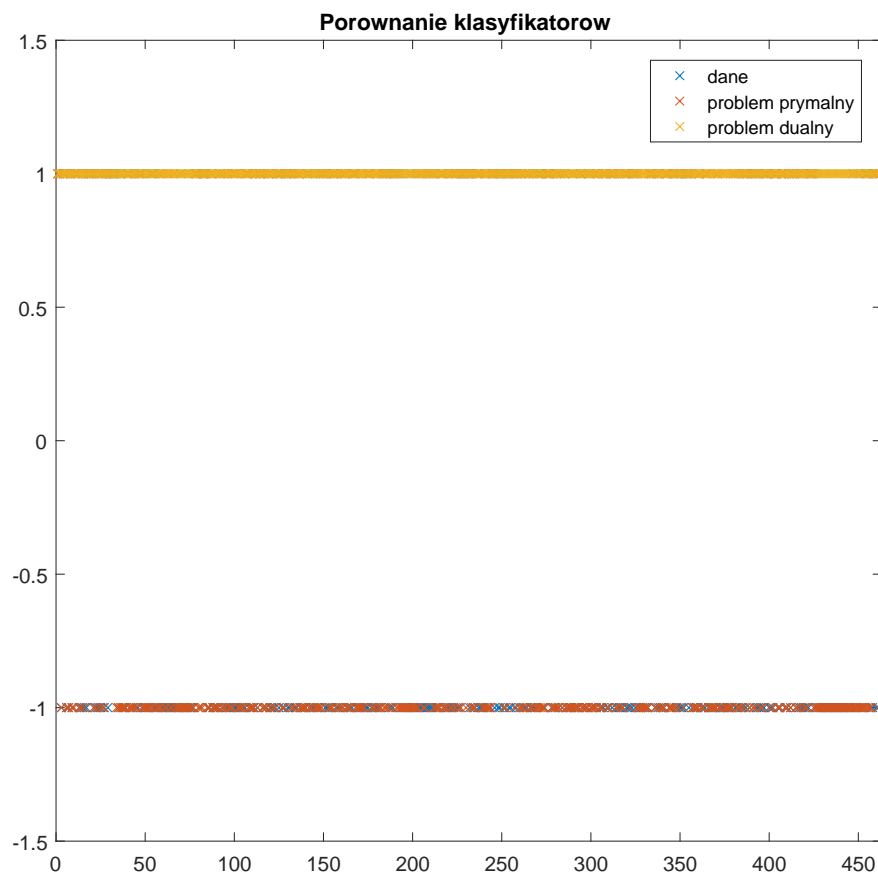
Wygenerowano (funkcja *randn*) zbiór danych pięciowymiarowych o wielkości 20 próbek (rys. 1). Przyporządkowano je do dwóch klas poprzez zamianę ostatniej współrzędnej (pierwsza klasa wartości dodatnie, druga ujemne). Rozwiązanie prymalne i dualne wygenerowały ten sam wektor. Klasyfikacja była całkowicie zgodna z przyporządkowaniem.



Rysunek 1: Wyniki klasyfikacji danych testowych

3 Rzeczywiste dane

Rzeczywiste dane (rys. 2) nie miały widocznej granicy pomiędzy płaszczyznami. Wykorzystanie optymalizacji problemu primalnego pozwoliło na uzyskanie dokładności ok. 66 %. Optymalizacja problemu dualnego nie udała się.



Rysunek 2: Wyniki klasyfikacji danych rzeczywistych

4 Podsumowanie

Zadanie prymalne pozwala na prostsze zapisanie problemu. Dodatkowo algorytm jest bardziej skuteczny od zastosowanego w rozwiązaniu problemu dualnego. Problem prymalny można obliczać przy pomocy tych samych solverów co problem dualny. W idealnym przypadku oba algorytmy dają to samo rozwiązanie.