



Politechnika
Wrocławska

Metody Systemowe i Decyzyjne L

Klasyfikacja przy użyciu Maszyny Wektorów Nośnych

Piotr Kawa

W4N, K46

sem. letni 2022/23



Agenda

- 1 Problem klasyfikacji
- 2 Maszyna wektorów nośnych
- 3 Główna idea SVM
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne
- 6 Klasyfikacja wieloklasowa
 - One vs one
 - One vs rest

Problem klasyfikacji

Klasyfikacja:

- przypisanie nowych danych do określonych kategorii lub klas na podstawie wcześniej nauczonych wzorców (przykładów).

Przykłady:

- określenie czy email jest spamem,
- przeżycie rejsu Titanic na podstawie miejsca na statku oraz wieku,
- klasyfikacja zwierząt na obrazie (np. pies vs. kot).

Agenda

- 1 Problem klasyfikacji
- 2 **Maszyna wektorów nośnych**
- 3 Główna idea SVM
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne
- 6 Klasyfikacja wieloklasowa
 - One vs one
 - One vs rest

Maszyna wektorów nośnych

Maszyna wektorów nośnych (ang. support vector machine, SVM):

- model uczenia maszynowego,
- służy do klasyfikacji i regresji danych,

Zadanie SVM to znalezienie hiperpłaszczyzny:

- najlepiej rozdzielającą dane wejściowe na różne klasy (klasyfikacja),
- najlepiej pasującą do danych (regresja).

Agenda

- 1 Problem klasyfikacji
- 2 Maszyna wektorów nośnych
- 3 Główna idea SVM**
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne
- 6 Klasyfikacja wieloklasowa
 - One vs one
 - One vs rest

Główna idea SVM

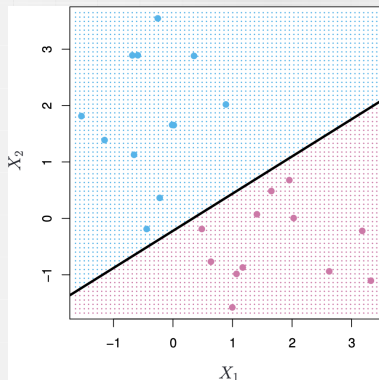


Figure: Hiperpłaszczyzna rozdzielająca punkty na należące do klasy niebieskiej i czerwonej. Kolorowe siatki punktów oznaczają płaszczyznę decyzyjną. Próbkę testową która znajdzie się na niebieskiej części zostanie przypisana do klasy niebieskiej, czerwona - do czerwonej. [1].

Główna idea SVM

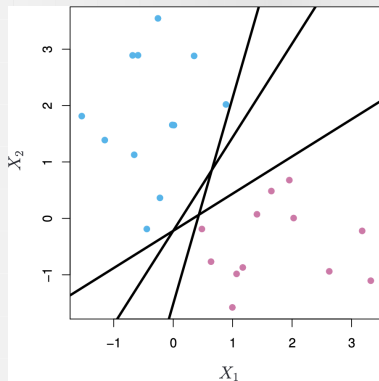
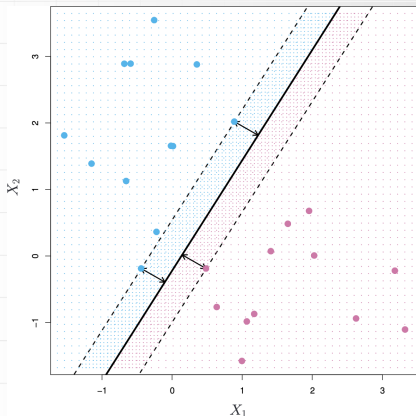


Figure: Istnieje wiele hiperpłaszczyzn przy użyciu których jesteśmy w stanie podzielić punkty - na jakiej podstawie wybieramy tą właściwą? [1].

Główna idea SVM



- hiperpłaszczyzna - granica decyzyjna, która rozdziela zbiór na różne klasy,
- wektory nośne - próbki najbliższe hiperpłaszczyźnie,
- margines - odległość prostopadła od hiperpłaszczyzny do wektorów nośnych.

Główna idea SVM

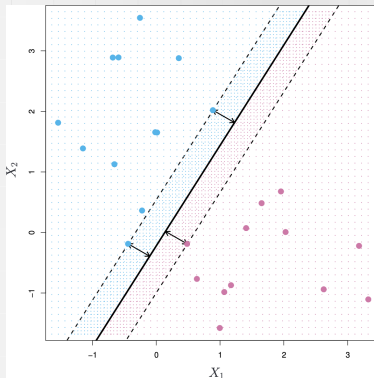


Figure: Celem SVM jest znalezienie granicy decyzyjnej, która maksymalizuje margines, czyli odległość między granicą a punktami. Dzięki temu ma tendencję do generowania granic decyzyjnych, które są maksymalnie oddalone od najbliższych punktów, a tym samym są najbardziej odporne na szum i zapewniają lepszą zdolność do generalizacji. [1]

Agenda

- 1 Problem klasyfikacji
- 2 Maszyna wektorów nośnych
- 3 Główna idea SVM
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne
- 6 Klasyfikacja wieloklasowa
 - One vs one
 - One vs rest

Parametry SVM

- C - stała regularyzacji, kontroluje balans między dopasowaniem do danych treningowych a minimalizacją błędów klasyfikacji. Niskie wartości C - większy margines, (ale i błędy klasyfikacji na danych treningowych); wysokie wartości C mogą prowadzić do overfittingu,
- jądro (ang. kernel) - funkcja która przekształca dane wejściowe do przestrzeni o wyższej wymiarowości.

Rodzaje:

- liniowe,
- wielomianowe,
- radialne - RBF (ang. radial basis function),
- γ (dla RBF i wielomianowych) - parametr określający zakres wpływu jednego punktu danych na granicę decyzyjną. Niskie wartości - płaskie granice decyzyjne; wysokie wartości - granice bardziej złożone i dostosowane do danych treningowych.

Agenda

- 1 Problem klasyfikacji
- 2 Maszyna wektorów nośnych
- 3 Główna idea SVM
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne**
- 6 Klasyfikacja wieloklasowa
 - One vs one
 - One vs rest

Dane liniowo nieseparowalne

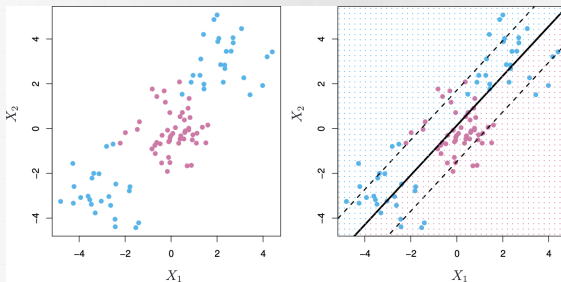


Figure: Nie wszystkie dane jesteśmy w stanie poprawnie zaklasyfikować przy użyciu liniowej granicy decyzyjnej. [1]

Dane liniowo nieseparowalne

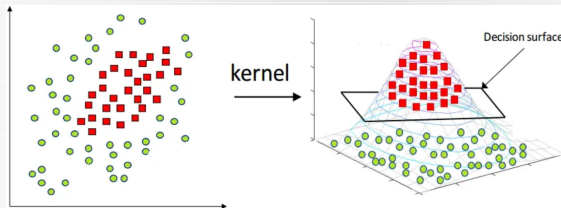


Figure: Kernel Trick (pol. sztuczka jądrowa) pozwala nam na przekształcenie danych wejściowych do przestrzeni o wyższej wymiarowości. Opieramy się w niej na użyciu funkcji jądra - możliwe jest zastosowanie liniowych metod klasyfikacji w przestrzeni o wyższej wymiarowości, które są w stanie rozwiązać problemy klasyfikacyjne, które nie byłyby rozwiązywalne w przestrzeni oryginalnych danych wejściowych. [1].

Dane liniowo nieseparowalne

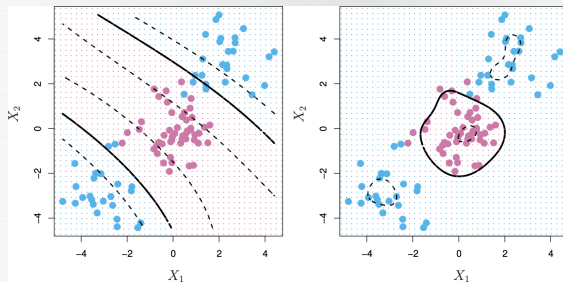


Figure: Porównanie - wielomianowe jądro 3 stopnia (lewo), jądro funkcji radialnej (prawo). [1].

Agenda

- 1 Problem klasyfikacji
- 2 Maszyna wektorów nośnych
- 3 Główna idea SVM
- 4 Parametry SVM
- 5 Dane liniowo nieseparowalne
- 6 Klasyfikacja wieloklasowa**
 - One vs one
 - One vs rest

SVM jest w stanie rozwiązywać problemy wieloklasowe. Może to robić na dwa sposoby:

- one vs one,
- one vs rest.

One vs one

One vs one:

- wykonujemy separację każdej pary klas (jednocześnie pomijając pozostałe klasy),
- otrzymujemy $\frac{N(N-1)}{2}$ klasyfikatorów - dla klas od 0 do N mamy następujące klasyfikatory "0 vs 1", "0 vs 2", ... "0 vs N", "1 vs 2", ... "N-1 vs N",
- wynik otrzymujemy np. poprzez głosowanie większościowe (ang. majority voting) - wskazujemy najczęściej wybraną klasę.

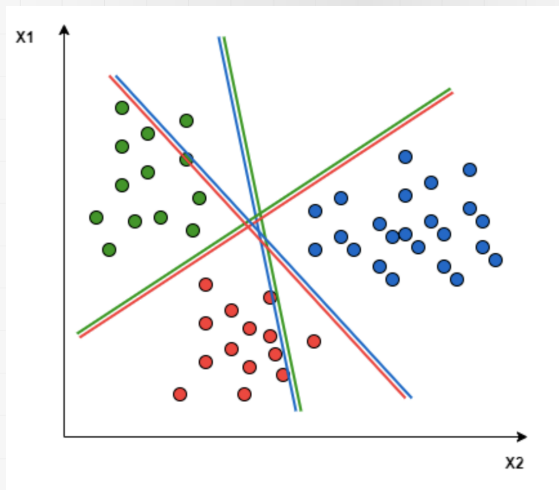


Figure: Wieloklasowa klasyfikacja SVM - one vs one. Źródło: [2].

One vs rest

One vs rest:

- wykonujemy separację każdej klasy kontra reszta klas (mamy dwa rodzaje próbek - jedna grupa to aktualna klasa, druga - wszystkie inne),
- otrzymujemy N klasyfikatorów,
- wynik otrzymujemy np. analizę prawdopodobieństwa przynależności do danej klasy. Jako odpowiedź wybieramy tę klasę która ma największe prawdopodobieństwo przynależności.

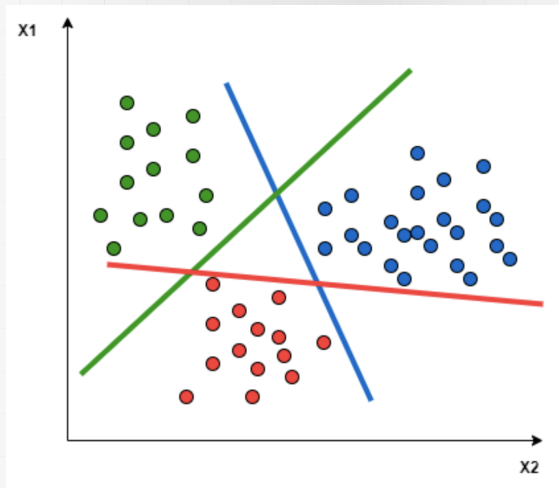


Figure: Wieloklasowa klasyfikacja SVM - one vs rest. Źródło: [1].



Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani.
An Introduction to Statistical Learning: with Applications in R.
2013.



Baeldung.com.
Multiclass classification using support vector machines.

Dziękuję za uwagę!