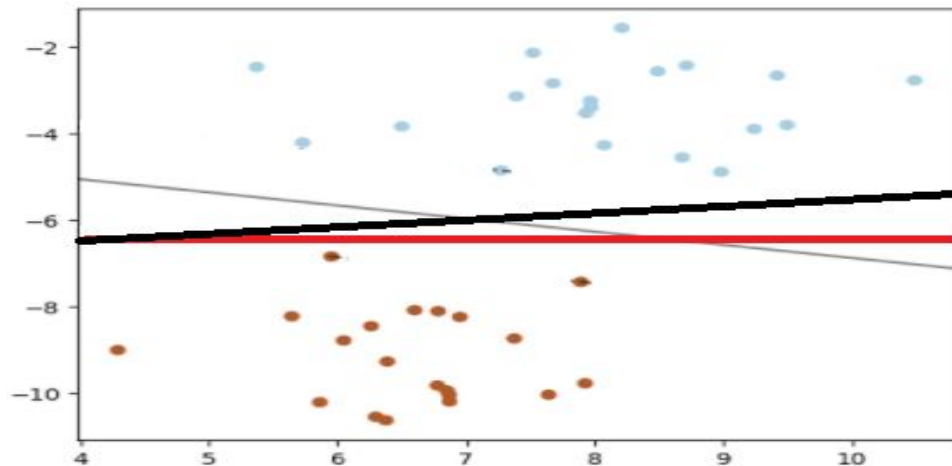


Klasyfikacja i regresja przy użyciu Support Vector Machine, krzywe ROC, Kernel Trick

Robert Bielas
Michał Kosowski



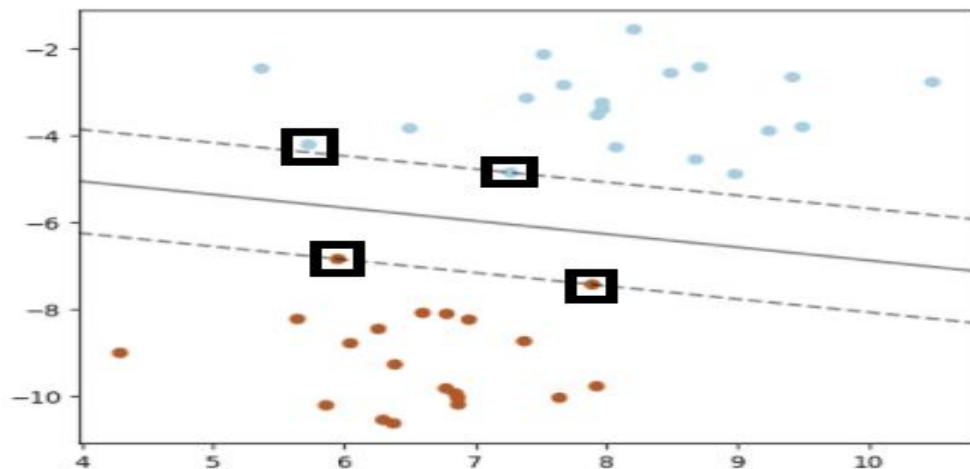
Klasyfikacja prostą, problem



Mamy dany zbiór binarny
(tzn. dwuklasowy)

Która prosta jest
najlepszym separatorem
obu klas?

Klasyfikacja - słownik pojęć



Punkty otoczone kwadratami - support vectors

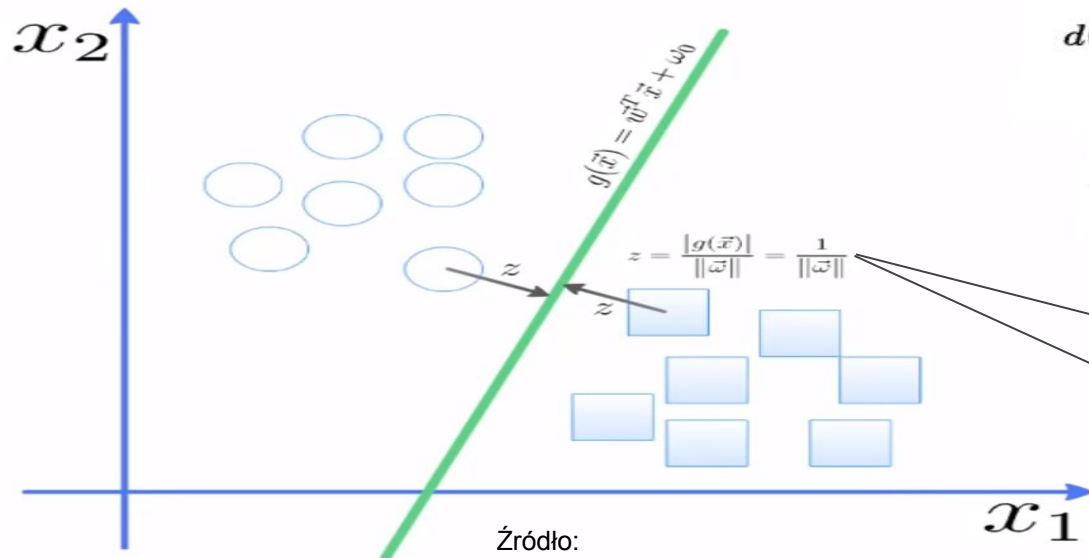
Pomarańczowe - tzw. positive samples

Niebieskie - tzw. negative samples

Support
Vector
Machine

Margins

Klasyfikacja c.d.



Źródło:
<https://www.youtube.com/watch?v=1NxnPkZM9bc&t=195s>

$$d(P, k) = \frac{|A x_P + B y_P + C|}{\sqrt{A^2 + B^2}}$$

$$g(\vec{x}) \geq 1, \quad \forall \vec{x} \in \text{class 1}$$

$$g(\vec{x}) \leq -1, \quad \forall \vec{x} \in \text{class 2}$$

$$\vec{\omega} = \sum_{i=0}^N \lambda_i y_i \vec{x}_i$$

$$\sum_{i=0}^N \lambda_i y_i = 0$$

$y_i = -1$ lub 1 jeśli odpowiednio
neg. lub pos.

$$\text{Klasyfikacja} = \text{sgn}\left(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + \rho\right)$$

Krzywe ROC - potrzebne pojęcia

Screening Tests Terminology

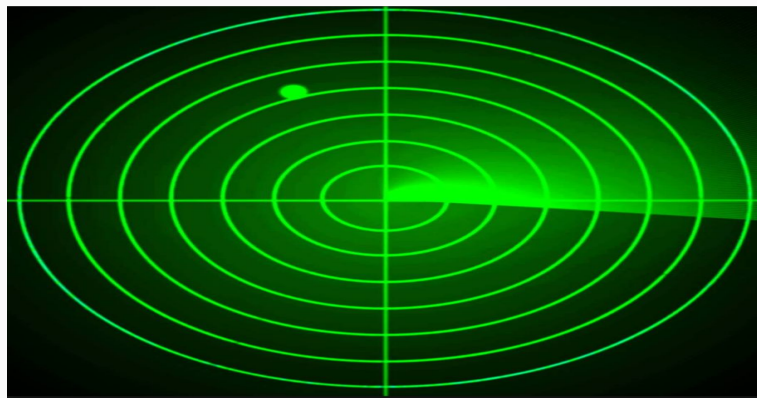
Sensitivity: Probability the test is positive given you have the disease

Specificity: Probability the test is negative given you do not have the disease

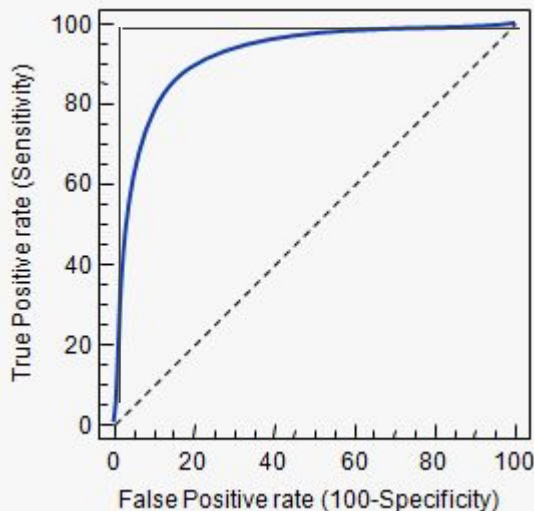
False Negative Rate (FNR): Probability the test is negative given you have the disease

False Positive Rate (FPR): Probability the test is positive given you do not have the disease

ROC - Receiver Operator Curve, etymologia - II wojna światowa, receiver - operator radaru



Ocena klasyfikatorów przy użyciu krzywej ROC



Algorytm:

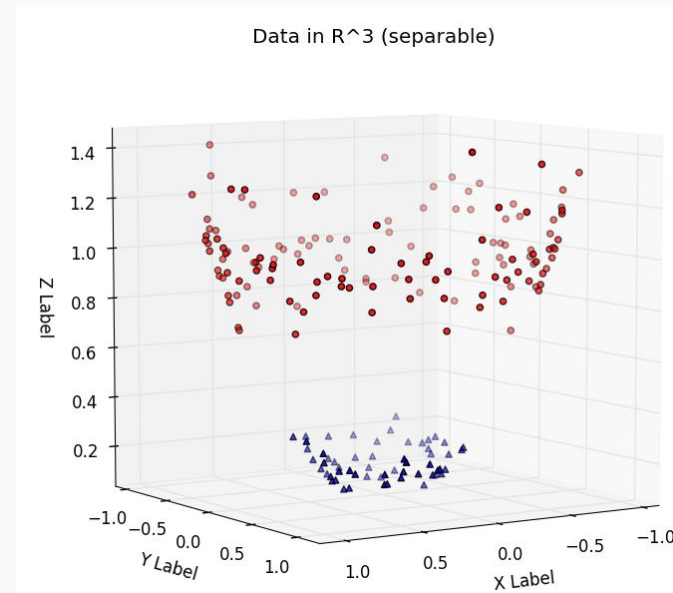
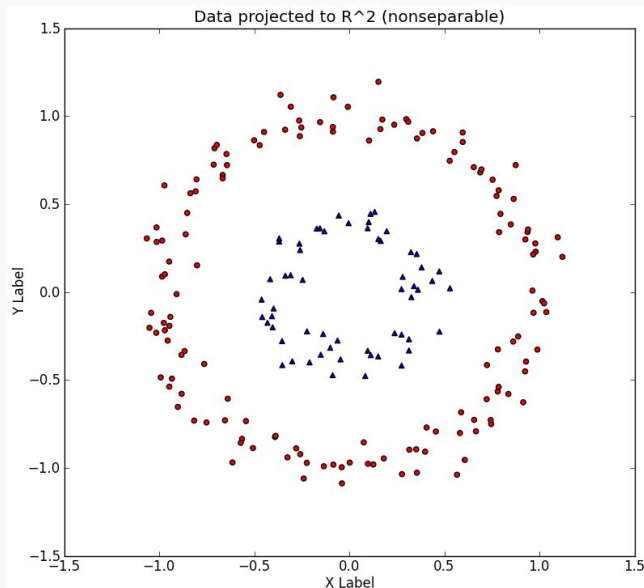
- Weź zbiór danych, którego wektory należą do jednej z dwóch różnych klas
- Dla każdego wektora liczymy jego score klasyfikacji badanego klasyfikatora.
- Sortujemy punkty wg ich wyniku malejąco
- Ideał jest taki, że wyniki klasyfikacji wszystkich “prawdziwie negatywnych” próbek leżą pod wynikami “prawdziwie pozytywnych” (idealny klasyfikator)
- Ustalamy wartość tzw. cutoff, posortowane wartości poniżej tego progu uznajemy za negatywne, powyżej za pozytywne. Znając prawidłowe przynależności wektorów do klas liczymy Sensitivity i False Positive Rate, odkładamy te wartości na wykresie
- Powtarzamy poprzedni krok dla wielu wartości cutoff
- Jakość klasyfikatora określamy przez pole powierzchni pod wykresem krzywej ROC (ang. Area Under Curve, AUC)

Dążymy do tego, żeby AUC wypełniło wykres.

Uznaje się, że “najgorszy klasyfikator” jest przedstawiany przez przerywaną krzywą ROC - klasyfikator po prostu zgaduje, ale UWAGA: można mieć jeszcze gorszy klasyfikator, gdzie już lepiej dla nas by było żeby klasyfikować rzutami monety.

Kernel Trick

Zastosowanie funkcji kernela, która sprowadza nieliniową postać iloczynu skalarnego wektorów do postaci liniowej w większej ilości wymiarów.



(Left) A dataset in \mathbb{R}^2 , not linearly separable. (Right) The same dataset transformed by the transformation:
$$[x_1, x_2] = [x_1, x_2, x_1^2 + x_2^2].$$

Często stosowane kernele

- Wielomianowy

$$k(x, y) = (\gamma x^\top y + c_0)^d$$

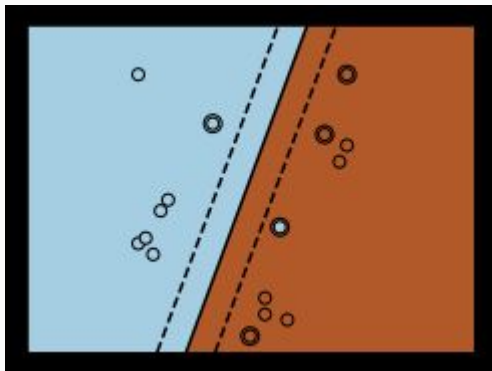
- Gaussowski (RBF - Radial Basis Function)

$$k(x, y) = \exp(-\gamma \|x - y\|^2)$$

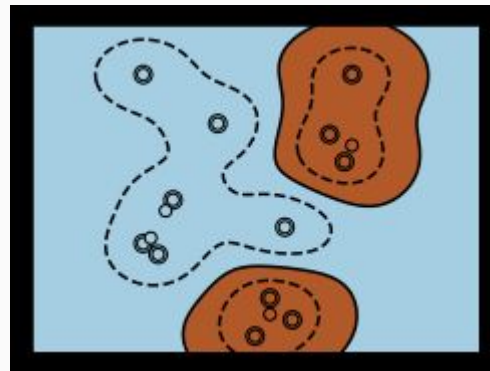
- Sigmoidalny

$$k(x, y) = \tanh(\gamma x^\top y + c_0)$$

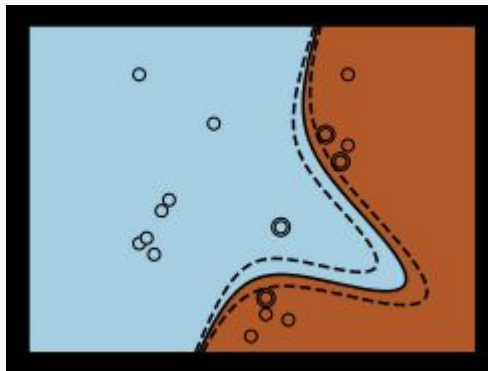
Kernele



Liniowy

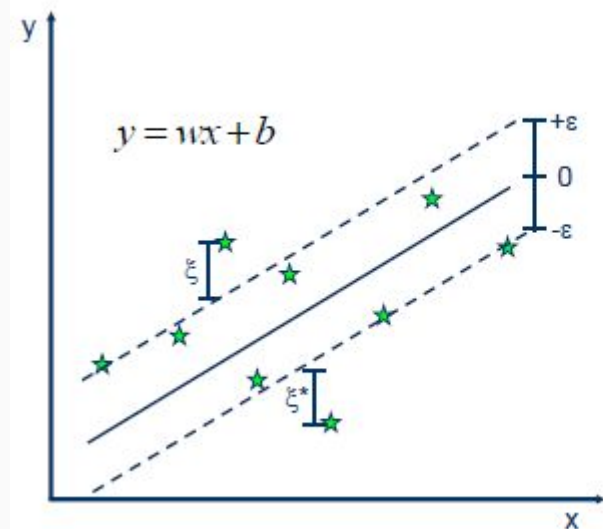


Gaussowski



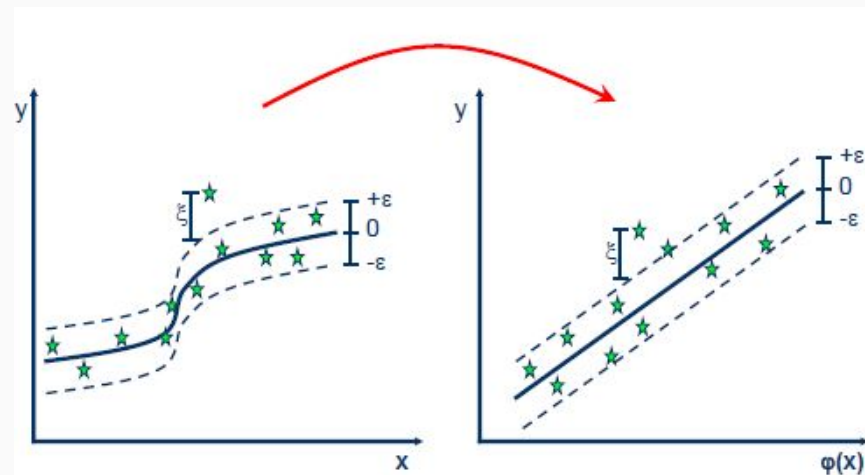
Wielomianowy

Regresja SVM



Linear SVR

- Minimize:
$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$
- Constraints:
$$y_i - wx_i - b \leq \epsilon + \xi_i$$
$$wx_i + b - y_i \leq \epsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \geq 0$$



Non-linear SVR

Linki do pouczających tutoriali

<https://www.youtube.com/watch?v=KTeVOb8gaD4>

<https://www.youtube.com/watch?v=N1vOgolbjSc>

<https://www.youtube.com/watch?v=ik7E7r2a1h8&t=>

<https://www.youtube.com/watch?v=gYlIKUP2hk0&t=9s>

<https://www.youtube.com/watch?v=OAl6eAyP-yo&t=10s>