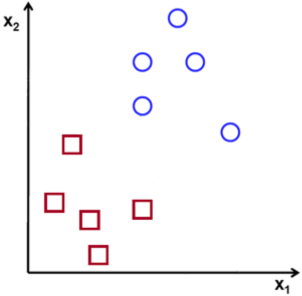
Support Vector Machine

Anna Przybyła

29.04.2020

Support Vector Machine

- Cel: Na podstawie danych treningowych znaleźć algorytm, który przyporządkuje każdemu elementowi ze zbioru treningowego kategorię, do której należy.



SVM - wprowadzenie

- Dana jest próbka treningowa

$$D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}, \text{ gdzie } y_i \in \{-1, 1\}, \ \mathbf{x}_i \in \mathbb{R}^p$$

- Klasyfikatorem liniowym nazywamy funkcję

$$f_{\mathbf{w},b}(\mathbf{x}) = sgn(\mathbf{w} \cdot \mathbf{x} + b)$$

- Próbka D jest liniowo seperowana jeśli istnieje klasyfikator liniowy $f_{\mathbf{w},b}(\ )$ taki, że $\forall i$

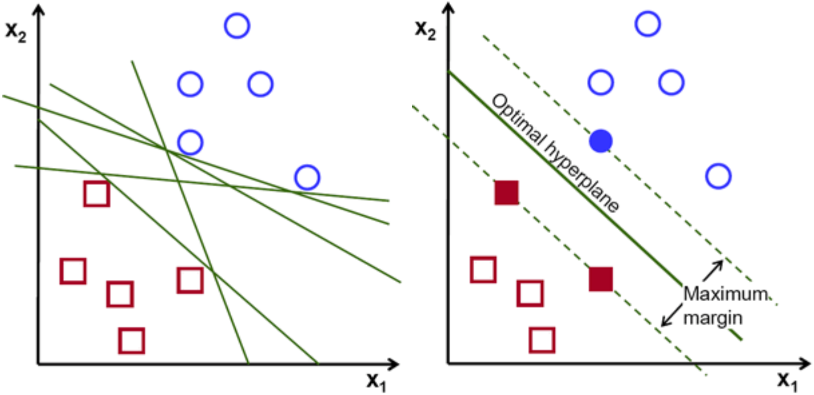
$$\mathbf{w} \cdot \mathbf{x}_i + b \geqslant +1, \text{ gdy } y_i = +1$$

$$\mathbf{w} \cdot \mathbf{x}_i + b < -1, \text{ gdy } y_i = -1$$

- Równoważnie

$$\forall i \ \ y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geqslant 1$$

SVM - najlepszy klasyfikator liniowy



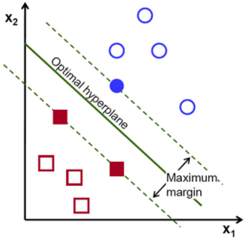
SVM - najlepszy klasyfikator liniowy

- Niech M - minimum odległości punktów od hiperpłaszczyzny rozdzielającej $\mathbf{w} \cdot \mathbf{x} + b = 0$
- Problem optymalizacyjny:

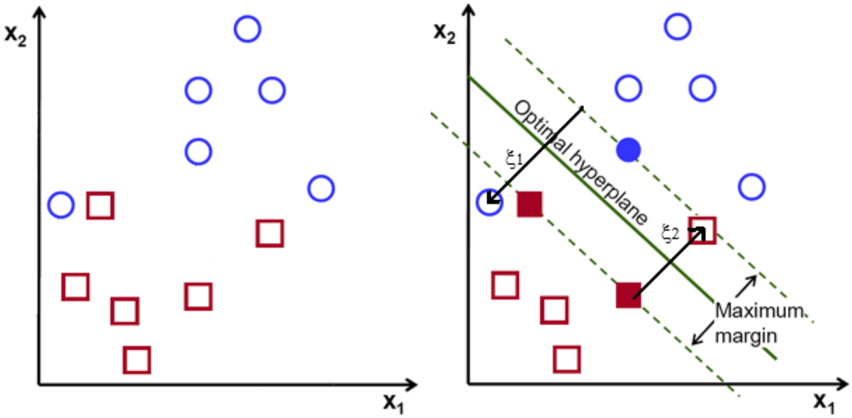
$$\max_{w,b,||\mathbf{w}||=1} M \text{ przy } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geqslant M, \quad i = 1, \ldots, N$$

- Równoważnie, przy założeniu $||\mathbf{w}|| = 1/M$

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 \text{ przy } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geqslant 1, \quad i = 1, \ldots, N$$



SVM - przypadek nieseparowalny liniowo - nieznaczna nieseparowalność



SVM - przypadek nieseparowalny liniowo - nieznaczna nieseparowalność

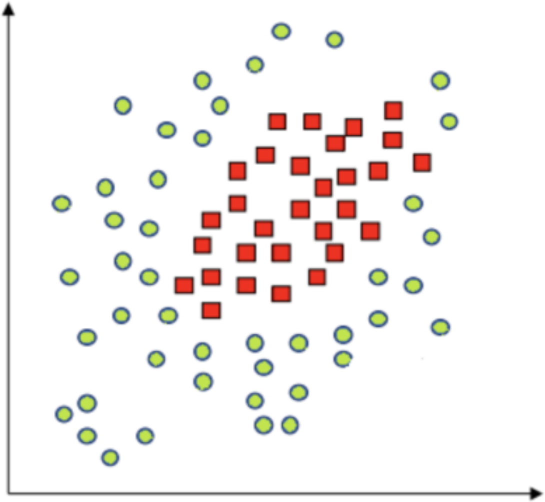
- Celem jest maksymalizacja M, ale przy dopuszczeniu niektórych punktów po złej stronie granicy
- Problem optymalizacyjny:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i$$

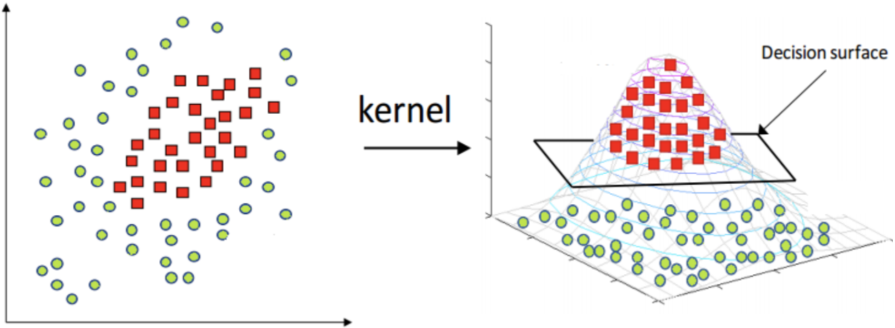
$$\text{przy} \;\; y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geqslant 1 - \xi_i, \;\; \xi_i \geqslant 0, \;\; i = 1, \ldots, N$$

- C - parametr kosztu - określa, czy ważniejsze przy optymalizacji jest zmniejszenie liczby źle sklasyfikowanych punktów, czy zwiększenie szerokości granicy

SVM - przypadek nieseparowalny liniowo - zmiana przestrzeni atrybutów



SVM - przypadek nieseparowalny liniowo - zmiana przestrzeni atrybutów



SVM - kernel trick

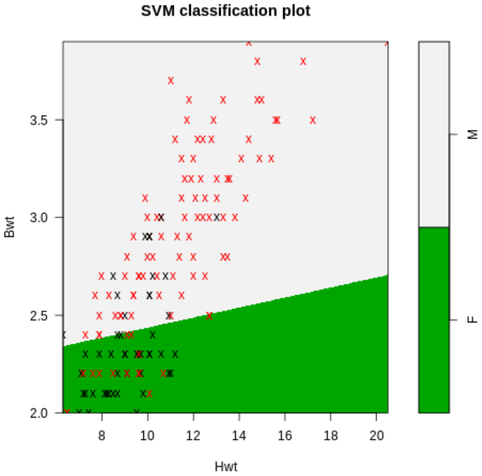
- Rozwiązaniem problemu optymalizacyjnego jest

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i < h(x), h(x_i) > +b$$

- Funkcja jądrowa (kernel function) $K(x, x') =< h(x), h(x') >$
- często stosowane funkcje jądrowe:
  - liniowa: $K(x, x') =< x, x' >$
  - wielomianowa: $K(x, x') = (1+ < x, x' >)^d$
  - gaussowska: $K(x, x') = exp(-\gamma ||x - x'||^2)$

Zbiór *Cats* zawierający dane o płci kotów, masie ciała i masie mięśnia sercowego.

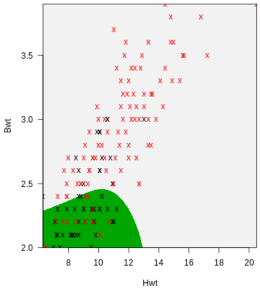
Funkcja jądrowa liniowa, C=50



Przykład - R

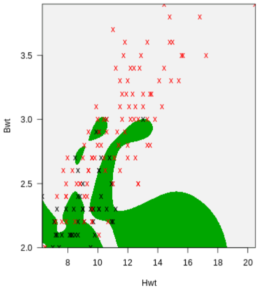
Funkcja jądrowa gaussowska

C=1



C=50

C=10e6



SVM - zalety i wady

- Zalety
  - SVM jest dobrą metodą przy braku wiedzy o danych.
  - Kernel trick umożliwia efektywne rozwiązywanie problemów nieliniowych.
  - Metoda działa efektywnie dla dużych wymiarów.
- Wady
  - Metoda SVM jest mocno zależna od parametrów, których wartość trudno określić.
  - Wybranie dobrej funkcji jądrowej jest trudne.
  - Dla dużych zbiorów danych czas uczenia jest długi.
  - Wyniki są trudne do interpretacji.

Literatura

- Laura Auria, Rouslan A. Moro, *Support Vector Machines (SVM) as a Technique for Solvency Analysis*
- Tony Van Gestel, David Martens, Bart Baesens, Daniel Feremans, Johan Huysmans, Jan Vanthienen, *Forecasting and Analyzing Insurance Companies' Ratings.* International Journal of Forecasting 23 (2007) 513–529
- Vipula Rawte, G Anuradha, *Fraud Detection in Health Insurance using Data Mining Techniques.* 2015 International Conference on Communication, Information Computing Technology (ICCICT), Jan. 16-17, Mumbai, India
- Kentaro Kuwata, Faizan Mahmood and Ryosuke Shibasaki, *Weather index for crop insurance to mitigate basis risk.*