POLITECHNIKA GDAŃSKA

Monitorowanie i Diagnostyka w Systemach Sterowania

Wydział Elektrotechniki i Automatyki
Katedra Elektrotechniki, Systemów Sterowania i Informatyki
Dr inż. Michał Grochowski

Monitorowanie i Diagnostyka w Systemach Sterowania

na studiach II stopnia specjalności: Systemy Sterowania i Podejmowania Decyzji

Maszyny Wektorów Nośnych

(Support Vector Machines - SVM)

na podstawie:

Yaser S. Abu-Mostafa. Learning from the Data. Course at California Institute of Technology

Alpaydin, E. Introduction to Machine Learning. The MIT Press Cambridge, Massachusetts London, England 2010.

Haykin, S. Neural Networks and Machine Learning., Pearson Prentice Hall, 2009.

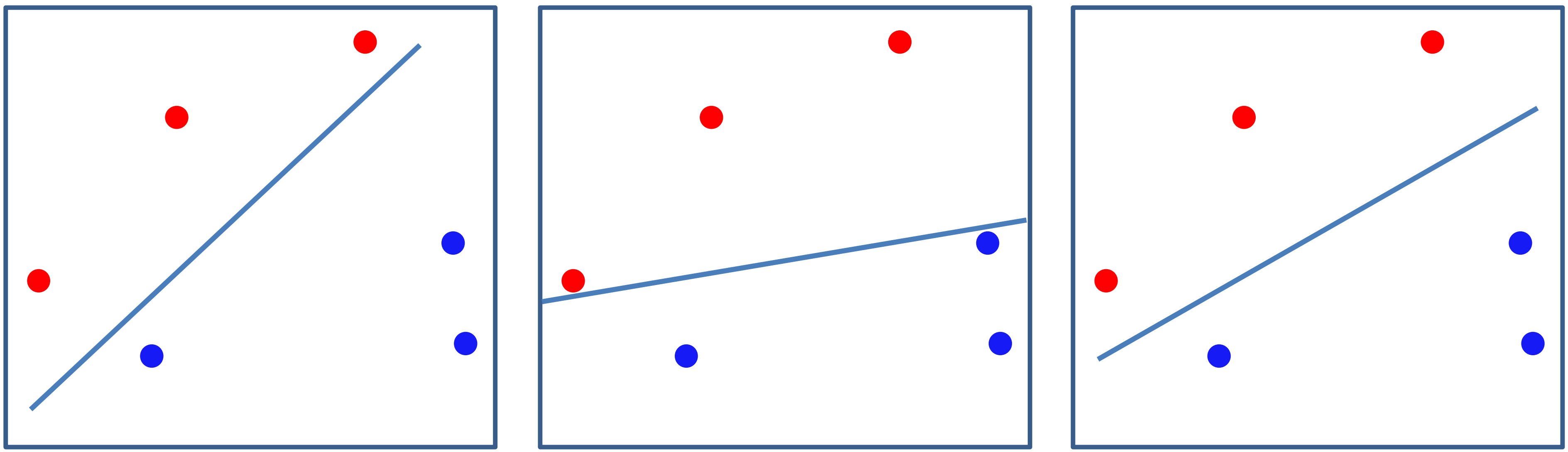
Opracował: dr inż. Michał Grochowski

kiss.pg.mg@gmail.com

michal.grochowski@pg.gda.pl

Charakterystyka problemu

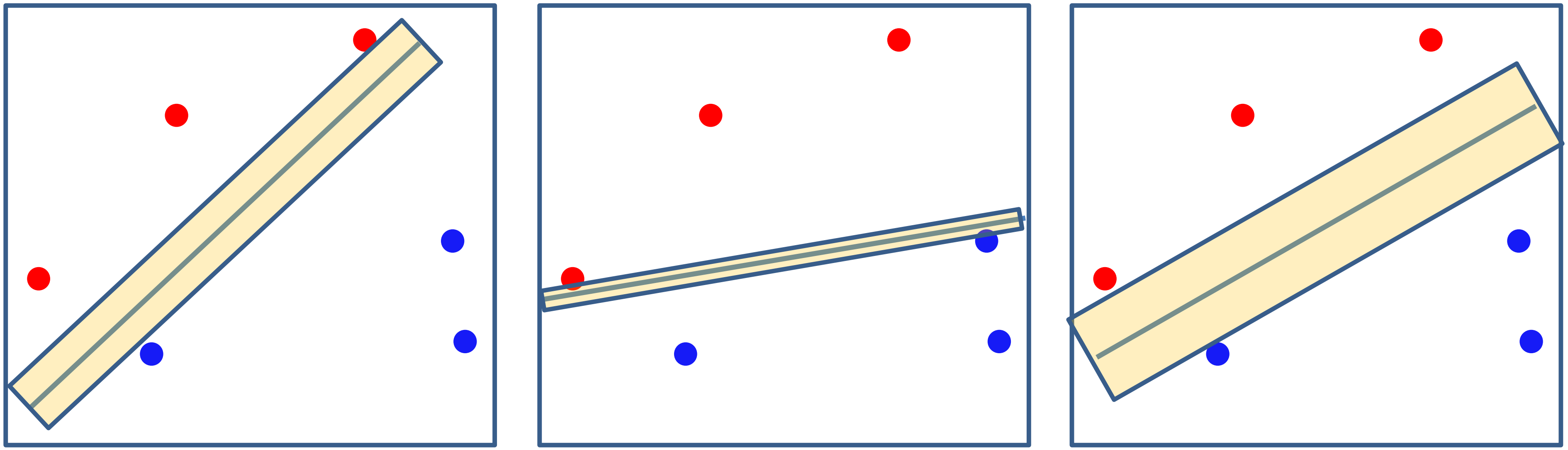
Jak rozdzielić dane ??



Nieskończenie wiele możliwości !!!

Charakterystyka problemu

Który margines jest najlepszy ??



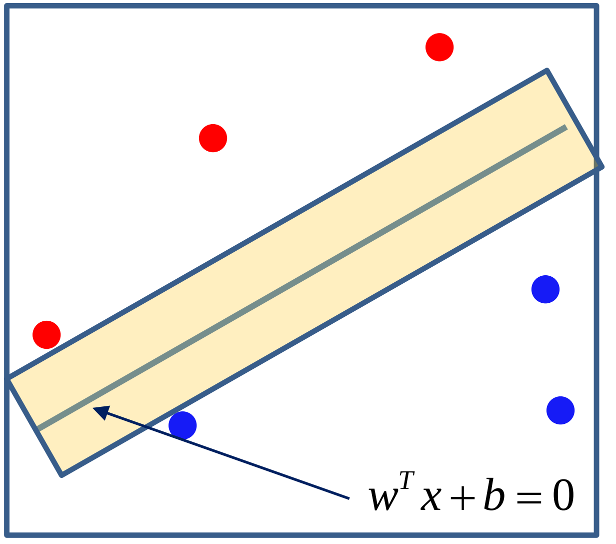
Która z linii maksymalizuje ten margines ??

Charakterystyka problemu

Niech: $\{(x_i, y_i)\}_{i=1}^{N}$ oznacza zbiór danych,

gdzie x_i jest wzorcem wejściowym, a y_i oznacza przynależność *i* -tego wejścia do danej klasy, „+1" lub „-1".

Załóżmy (na chwilę) że punkty są liniowo separowalne.



Linia (płaszczyzna, hiperpłaszczyzna) decyzyjna opisana jest zależnością:

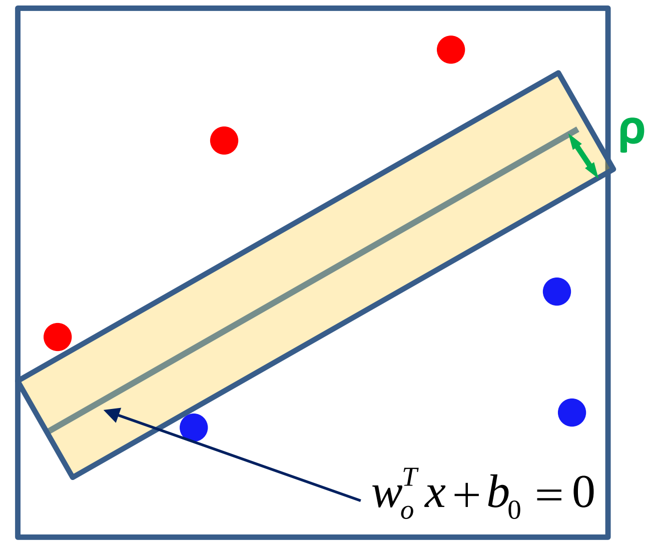
$$w^T x + b = 0 \tag{1}$$

gdzie ***x*** jest wektorem wejściowym, ***w*** jest „nastawialnym" wektorem (np. wag), ***b*** jest progiem.

Charakterystyka problemu

Możemy zapisać:

$$w^T x_i + b \geq 0 \qquad dla \qquad y_i = +1$$
$$w^T x_i + b < 0 \qquad dla \qquad y_i = -1 \qquad \textbf{(2)}$$



Optymalna linia (płaszczyzna, hiperpłaszczyzna) decyzyjna maksymalizująca margines ρ opisana jest zależnością:

$$w_o^T x + b_0 = 0 \qquad \textbf{(3)}$$

Nieco podstawowych przekształceń ...

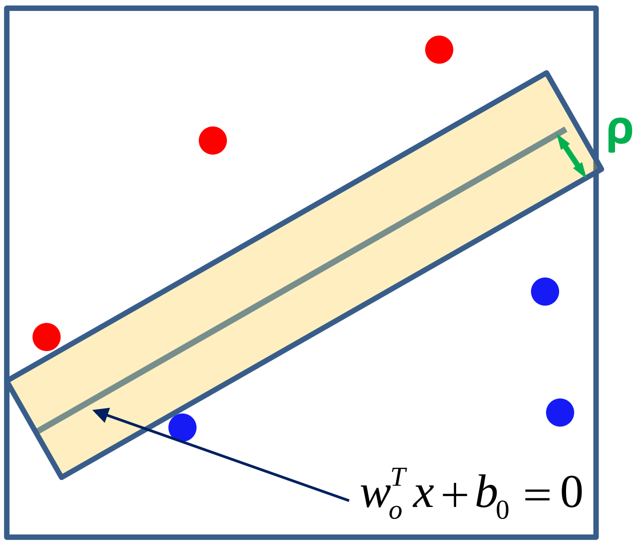
$$w^T x_i + b \geq 0 \quad dla \quad y_i = +1$$
$$w^T x_i + b < 0 \quad dla \quad y_i = -1$$

Wiemy że (z (2)):

$$\left| w^T x + b \right| \geq 0 \qquad (4)$$

Możemy więc przyjąć że (normalizacja):

$$\left| w^T x + b \right| = 1 \qquad (5)$$



Możemy więc zapisać:

$$w_o^T x_i + b_o \geq +1 \quad dla \quad y_i = +1$$
$$w_o^T x_i + b_o < -1 \quad dla \quad y_i = -1 \qquad (6)$$

lub:

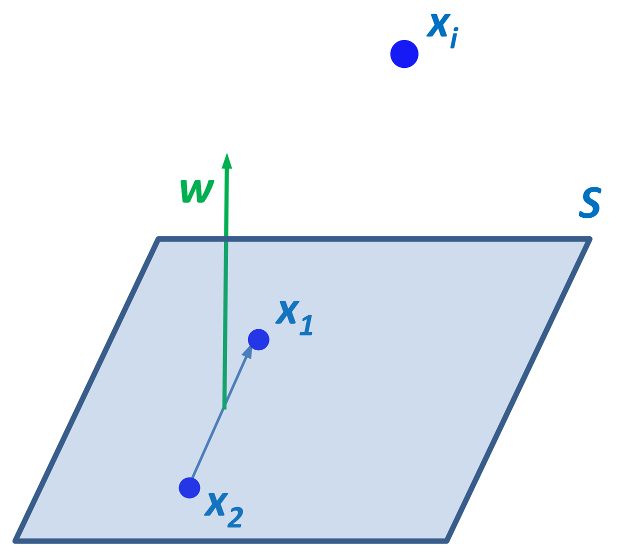
$$y_i (w_o^T x_i + b_o) = +1$$

Nieco podstawowych przekształceń ...

Odległość od płaszczyzny decyzyjnej

Niech x_i będzie punktem leżącym najbliżej płaszczyzny: $w^T x + b = 0$

oraz: $\left| w^T x + b \right| = 1$



Wektor **w** jest prostopadły do płaszczyzny ***S*** w przestrzeni wejść ***X***.

Dowolne punkty x_1 oraz x_2 należące do płaszczyzny ***S***, spełniają jej równanie, więc:

$$w^T x_1 + b = 0 \qquad w^T x_2 + b = 0$$

więc: $w^T (x_1 - x_2) = 0$

Stąd **w** jest ortogonalny do jakiegokolwiek wektora na płaszczyźnie ***S***

$$\left| w^T x + b \right| \geq 0$$

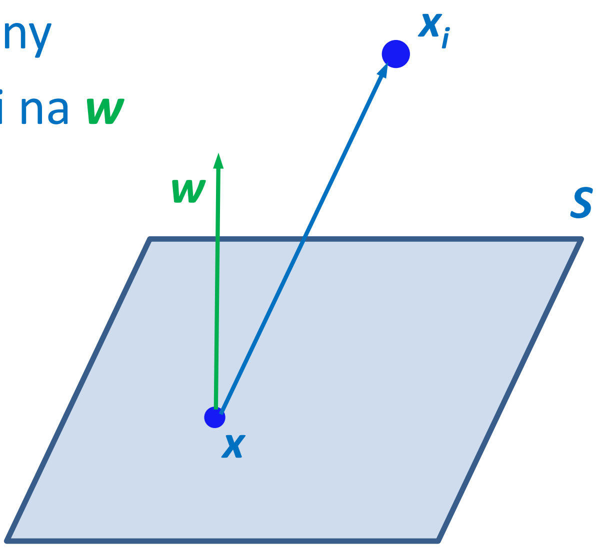
Nieco podstawowych przekształceń …

Odległość od płaszczyzny decyzyjnej

Odległość pomiędzy $\boldsymbol{x}_i$ i płaszczyzną:

… weźmy jakikolwiek punkt $\boldsymbol{x}$ należący do płaszczyzny

… utwórzmy wektor $\boldsymbol{x}_i - \boldsymbol{x}$ i dokonajmy jego projekcji na $\boldsymbol{w}$



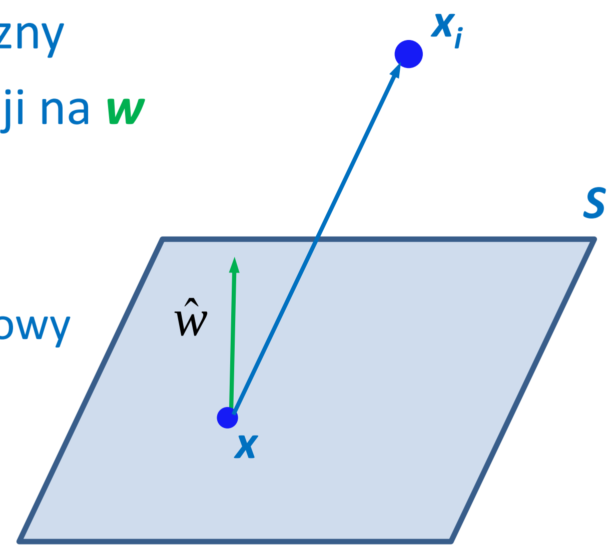
Nieco podstawowych przekształceń …

Odległość od płaszczyzny decyzyjnej

Odległość pomiędzy $\boldsymbol{x_i}$ i płaszczyzną (***L***):

… weźmy jakikolwiek punkt ***x*** należący do płaszczyzny

… utwórzmy wektor $\boldsymbol{x_i - x}$ i dokonajmy jego projekcji na ***w***



Matematyka …

$$\hat{w} = \frac{w}{\|w\|}$$

$\hat{w}$ - wektor jednostkowy

odległość ***L***:

$$L = \left|\hat{w}^T (x_i - x)\right|$$

$$L = \frac{1}{\|w\|}\left|w^T x_i - w^T x\right| = \frac{1}{\|w\|}\left|w^T x_i + b - w^T x - b\right| = \frac{1}{\|w\|}$$

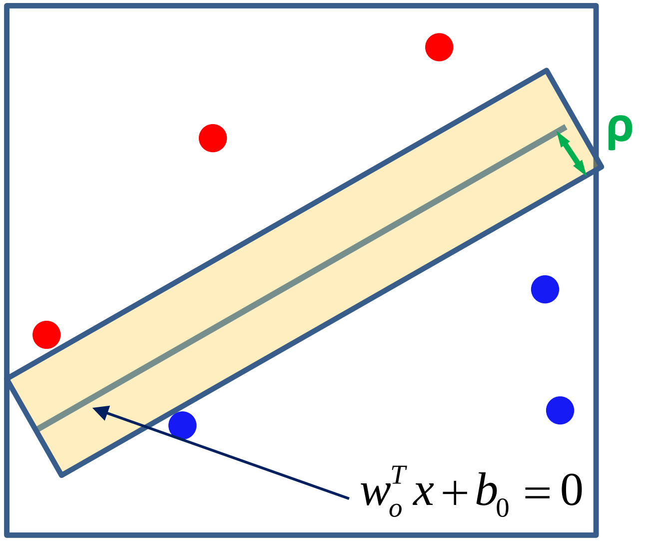
ponieważ:

$$\left|w^T x + b\right| = 0 \quad \text{oraz} \quad \left|w^T x_i + b\right| = 1$$

Problem optymalizacyjny

Zmaksymalizować odległość L = ρ

$$\max \frac{1}{\|w\|} = \min \; w^T w$$



Zadanie optymalizacji:

$$\min_{w \in R} \; g(w) = \frac{1}{2} w^T w$$

przy ograniczeniach (z równania 6):

$$y_i (w^T x_i + b) \geq 1 \qquad dla \qquad i = 1, 2, \ldots, N$$

gdzie:

$$w \in R^d, \; b \in R$$

Problem optymalizacyjny

Problem optymalizacji kwadratowej z ograniczeniami - liniowo kwadratowy

Zadanie optymalizacji (kwadratowej, g(w) jest wypukła):

$$\min_{w \in R} \; g(w) = \frac{1}{2} w^T w$$

przy ograniczeniach (liniowych względem ***w***):

$$y_i (w^T x_i + b) \geq 1 \qquad dla \qquad i = 1, 2, \ldots, N$$

Tak sformułowany problem jest prymalnym zagadnieniem optymalizacyjnym.

Możemy go rozwiązać stosując metodę mnożników Lagranga
oraz warunków Kukna – Tuckera (lub Karush – Kukna – Tuckera, w skrócie *KKT*).

Problem optymalizacyjny

Sformułowanie prymalne

Funkcja Lagranga:

$$L(w,b,\lambda)=\frac{1}{2}w^T w-\sum_{i=1}^{N}\lambda_i\left[y_i\left(w^T x_i+b\right)-1\right]$$

$\boldsymbol{\lambda_i}$ - mnożniki Lagranga

Minimalizujemy względem ***w*** oraz ***b*,**
a jednocześnie maksymalizujemy względem **$\boldsymbol{\lambda}$** ***($\lambda_i \geq 0$) !!!***

Warunki optymalności:

$$\frac{\delta L(w,b,\lambda)}{\delta w}=0$$

$$\frac{\delta L(w,b,\lambda)}{\delta b}=0$$

stąd:

$$w=\sum_{i=1}^{N}\lambda_i y_i x_i \quad \text{oraz} \quad \sum_{i=1}^{N}\lambda_i y_i=0$$

Problem optymalizacyjny

Sformułowanie dualne

Możemy podstawić:

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i \quad \text{oraz} \quad \sum_{i=1}^{N} \lambda_i y_i = 0$$

do:

$$L(w,b,\lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \lambda_i \left[y_i \left(w^T x_i + b \right) - 1 \right]$$

otrzymując ($Q(\lambda)=L(w,b,\lambda)$):

$$\underset{\lambda}{Q(\lambda)} = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left[y_i y_j \lambda_i \lambda_j x_i^T x_j \right]$$

Którego szukamy maksimum po λ_i, przy ograniczeniach:

$$\lambda_i \geq 0, \quad i = 1, \dots, N \quad \text{oraz} \quad \sum_{i=1}^{N} \lambda_i y_i = 0$$

Problem optymalizacyjny

Sformułowanie dualne

Niech: $\{(x_i, y_i)\}_{i=1}^{N}$ oznacza zbiór danych (treningowych),

Znajdź współczynniki Lagranga $\{\lambda_i\}_{i=1}^{N}$, które minimalizują funkcję:

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[y_i y_j \lambda_i \lambda_j x_i^T x_j\right] - \sum_{i=1}^{N}\lambda_i$$

przy ograniczeniach:

$$\lambda_i \geq 0, \quad i = 1, \dots, N$$

$$\sum_{i=1}^{N}\lambda_i y_i = 0$$

Problem optymalizacyjny

Sformułowanie dualne

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\lambda^T \begin{bmatrix} y_1 y_1 x_1^T x_1 & y_1 y_2 x_1^T x_2 & \cdots & y_1 y_N x_1^T x_N \\ y_2 y_1 x_2^T x_1 & y_2 y_2 x_2^T x_2 & \cdots & y_2 y_N x_2^T x_N \\ \cdots & \cdots & & \cdots \\ y_N y_1 x_N^T x_1 & y_N y_2 x_N^T x_2 & \cdots & y_N y_N x_N^T x_N \end{bmatrix} \lambda + (-1^T)\lambda$$

przy ograniczeniach:

$$y^T \lambda = 0$$

$$0 \le \lambda \le \infty$$

Problem optymalizacyjny

Sformułowanie dualne

Uwagi:

- problem dualny oparty jest całkowicie o dane wejściowe (treningowe, pomiarowe);
- funkcja optymalizowana zależy wyłącznie od iloczynu skalarnego wejść.

$$\left\{ x_i^T \cdot x_j \right\}_{i,j=1}^{N}$$

Cechy wektorów nośnych

Sformułowanie dualne

Rozwiązanie:

$$\lambda_o = \lambda_{o,1}, \ldots, \lambda_{o,N}$$

***λ*$_o$** – optymalne mnożniki Lagranga

Z warunków *KKT*:

$$\lambda_i \left[y_i \left(w^T x_i + b \right) - 1 \right] = 0$$

Wszystkie niezerowe mnożniki Lagranga ***λ*$_{o,i}$** określają
wektory nośne (podtrzymujące) x$_i$.
Takich wektorów jest ***N$_S$***

$$\lambda_{o,i} > 0 \Rightarrow x_i$$

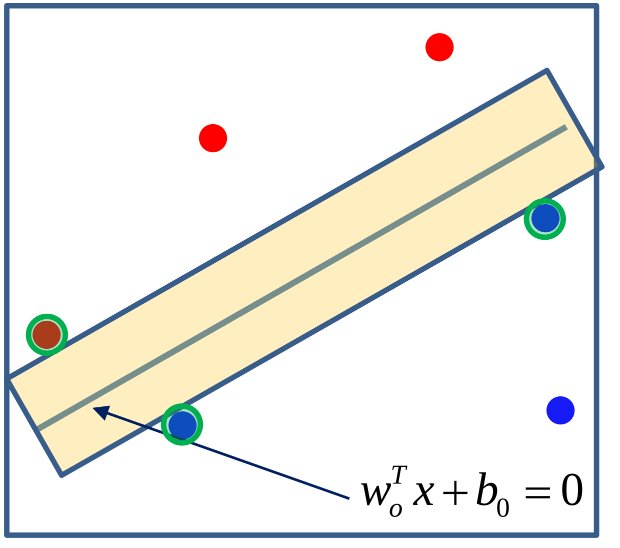
wektory nośne (podtrzymujące) SV

Cechy wektorów nośnych

Sformułowanie dualne

Rozwiązanie:

$\boldsymbol{x_i}$ leżące najbliżej optymalnej linii (płaszczyzny, hiperpłaszczyzna) decyzyjnej są wektorami nośnymi (podtrzymującymi).
$\boldsymbol{x_i}$ „dotykają” marginesu ρ.



Wektory nośne spełniają:

$$y_i\left(w^T x_i + b\right) = 1$$

Cechy wektorów nośnych

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i \qquad \sum_{i=1}^{N} \lambda_i y_i = 0$$

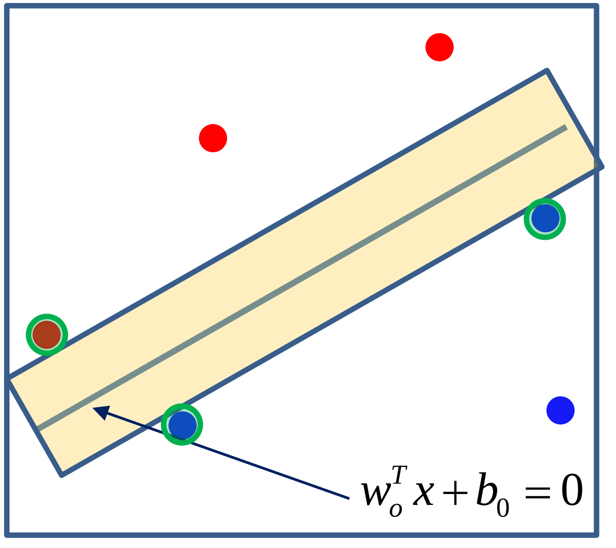
Sformułowanie dualne

Rozwiązanie:

Wyznaczamy optymalny wektor $\boldsymbol{w_o}$:

$$w_o = \sum_{i=1}^{Ns} \lambda_{o,i} y_i x_i$$

N_S – ilość niezerowych $\boldsymbol{\lambda_{o,i}}$



Wyznaczamy optymalny wektor $\boldsymbol{b_o}$:

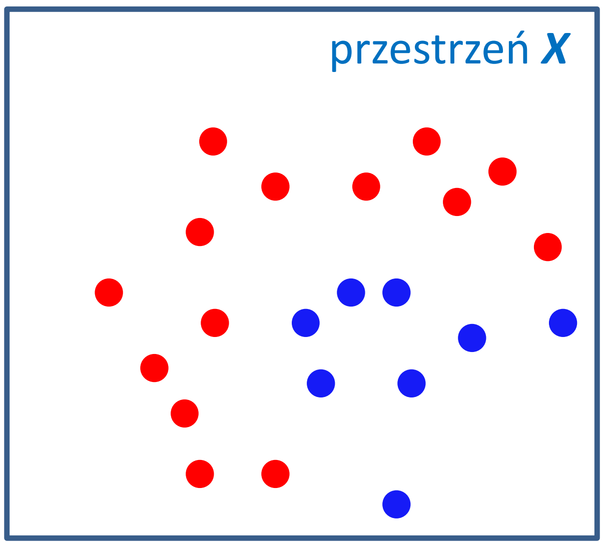
$$b_0 = 1 - w_o^T x^{(S)} \qquad dla \;\; y^{(S)} = 1$$

$$b_0 = 1 - \sum_{i=1}^{Ns} \lambda_{o,i} y_i x_i x^{(S)}$$

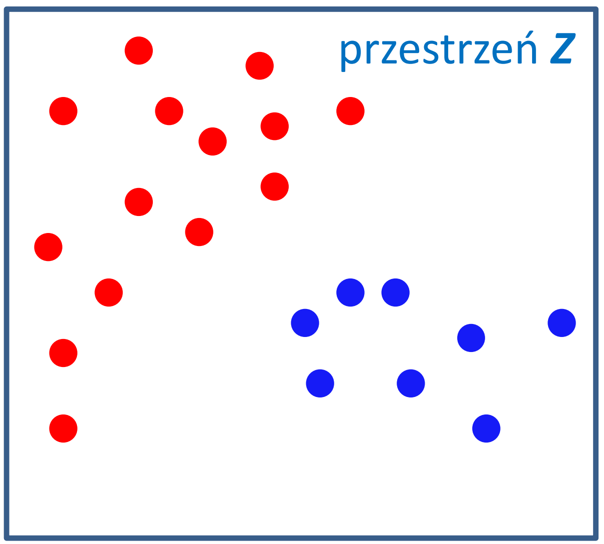
Problemy liniowo nieseparowalne

Transformacja do innej przestrzeni przestrzeni (*X* → *Z*)

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[y_i y_j \lambda_i \lambda_j x_i^T x_j\right] - \sum_{i=1}^{N}\lambda_i$$



$$X \rightarrow Z$$

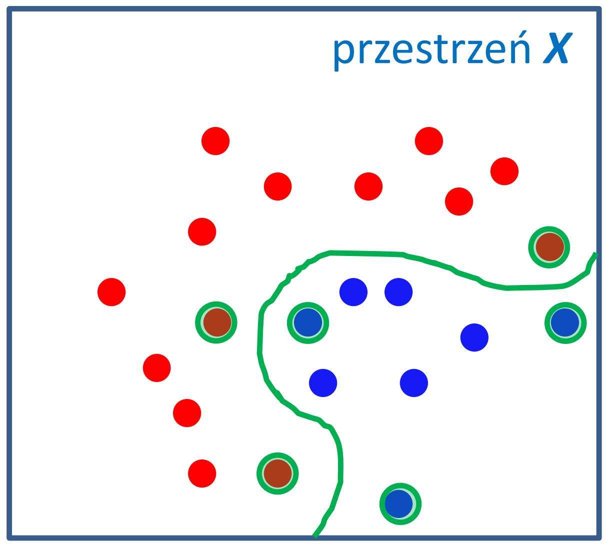


$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[y_i y_j \lambda_i \lambda_j z_i^T z_j\right] - \sum_{i=1}^{N}\lambda_i$$

Problemy liniowo nieseparowalne

Cechy wektorów nośnych

Wektory nośne leżą w przestrzeni ***Z !!!***



W przestrzeni ***X*** możemy jedynie obserwować ich „obraz”***!!!***

Również marginesy określane są w przestrzeni ***Z !!!***

Problemy liniowo nieseparowalne

Co potrzebujemy z przestrzeni *Z* ?

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[y_i y_j \lambda_i \lambda_j z_i^T z_j\right] - \sum_{i=1}^{N}\lambda_i$$

Uwaga: Nie potrzebujemy wektorów $\boldsymbol{z_i}$ i $\boldsymbol{z_j}$ a jedynie ich iloczynów skalarnych !!!

Ograniczenia:

$$\lambda_i \geq 0, \quad i = 1, \ldots, N$$

$$\sum_{i=1}^{N}\lambda_i y_i = 0$$

Otrzymujemy:

$$f(x) = sign(w_o z^T + b_o)$$

gdzie:

$$w_o = \sum_{i=1}^{Ns}\lambda_{o,i} y_i z_i$$

potrzebujemy: $z_i^T z_j$

$$b_0: \quad y_i(w_o^T z_j^{(S)} + b_o) = 1$$

potrzebujemy: $z_i^T z_j$

Problemy liniowo nieseparowalne

Obliczenia w przestrzeni *Z (kernel trick)*

Dla dowolne punktów $\boldsymbol{x_1}$ oraz $\boldsymbol{x_2}$ należących do $\boldsymbol{X}$, chcemy policzyć $z_1^T z_2$

Niech:

$$z_1^T z_2 = k(x_1, x_2)$$

Będzie jądrem (kernel) – „iloczyn skalarny" $\boldsymbol{x_1}$ oraz $\boldsymbol{x_2}$

Np: Gaussa:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\delta^2}\right), \text{ przy } \delta > 0$$

Problemy liniowo nieseparowalne

Obliczenia w przestrzeni *Z (kernel trick)*

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\lambda^T \begin{bmatrix} y_1 y_1 \underline{x_1^T x_1} & y_1 y_2 \underline{x_1^T x_2} & \cdots & y_1 y_N \underline{x_1^T x_N} \\ y_2 y_1 \underline{x_2^T x_1} & y_2 y_2 \underline{x_2^T x_2} & \cdots & y_2 y_N \underline{x_2^T x_N} \\ \cdots & \cdots & & \cdots \\ y_N y_1 \underline{x_N^T x_1} & y_N y_2 \underline{x_N^T x_2} & \cdots & y_N y_N \underline{x_N^T x_N} \end{bmatrix} \lambda + (-1^T)\lambda$$

$$X \rightarrow Z$$

$$\min_{\lambda} Q(\lambda) = \frac{1}{2}\lambda^T \begin{bmatrix} y_1 y_1 k\underline{(x_1, x_1)} & y_1 y_2 k\underline{(x_1, x_2)} & \cdots & y_1 y_N k\underline{(x_1, x_N)} \\ y_2 y_1 k\underline{(x_2, x_1)} & y_2 y_2 k\underline{(x_2, x_2)} & \cdots & y_2 y_N k\underline{(x_2, x_N)} \\ \cdots & \cdots & & \cdots \\ y_N y_1 k\underline{(x_N, x_1)} & y_N y_2 k\underline{(x_N, x_2)} & \cdots & y_N y_N k\underline{(x_N, x_N)} \end{bmatrix} \lambda + (-1^T)\lambda$$

Problemy liniowo nieseparowalne

Obliczenia w przestrzeni *Z (kernel trick)*

Wyraźmy f(x) przy pomocy K(-,-):

$$z_1^T z_2 = k(x_1, x_2)$$

$$f(x) = sign(w_o z^T + b_o)$$

$$w_o = \sum_{i=1}^{Ns} \lambda_{o,i} y_i z_i$$

to:

$$f(x) = sign(\sum_{i=1}^{Ns} \lambda_i y_i K(x_i, x) + b_o)$$

gdzie:

$$b_0 = y_i - \sum_{i=1}^{Ns} \lambda_i y_i K(x_i, x)$$

dla każdego niezerowego λ_i :

$$\lambda_i > 0$$

Dziękuję za uwagę