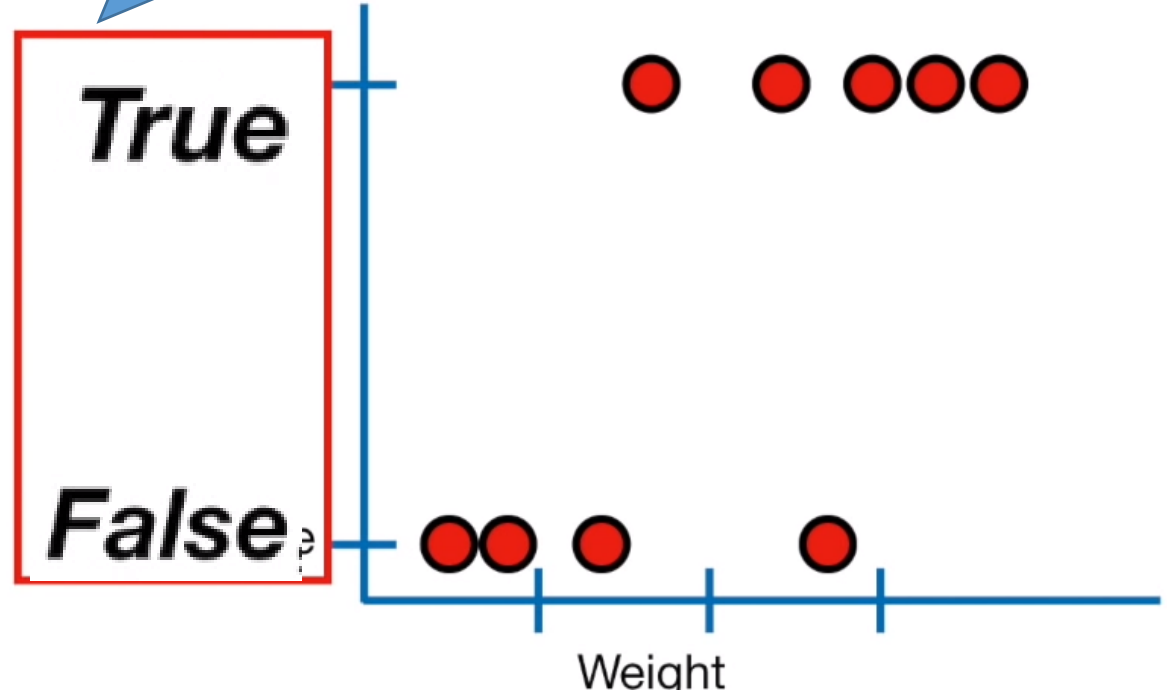
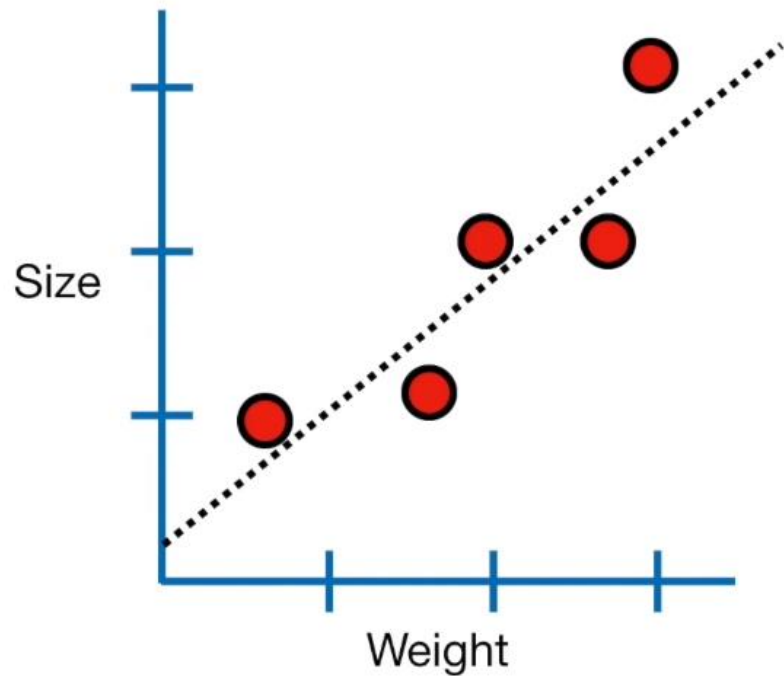


Logistic Regression

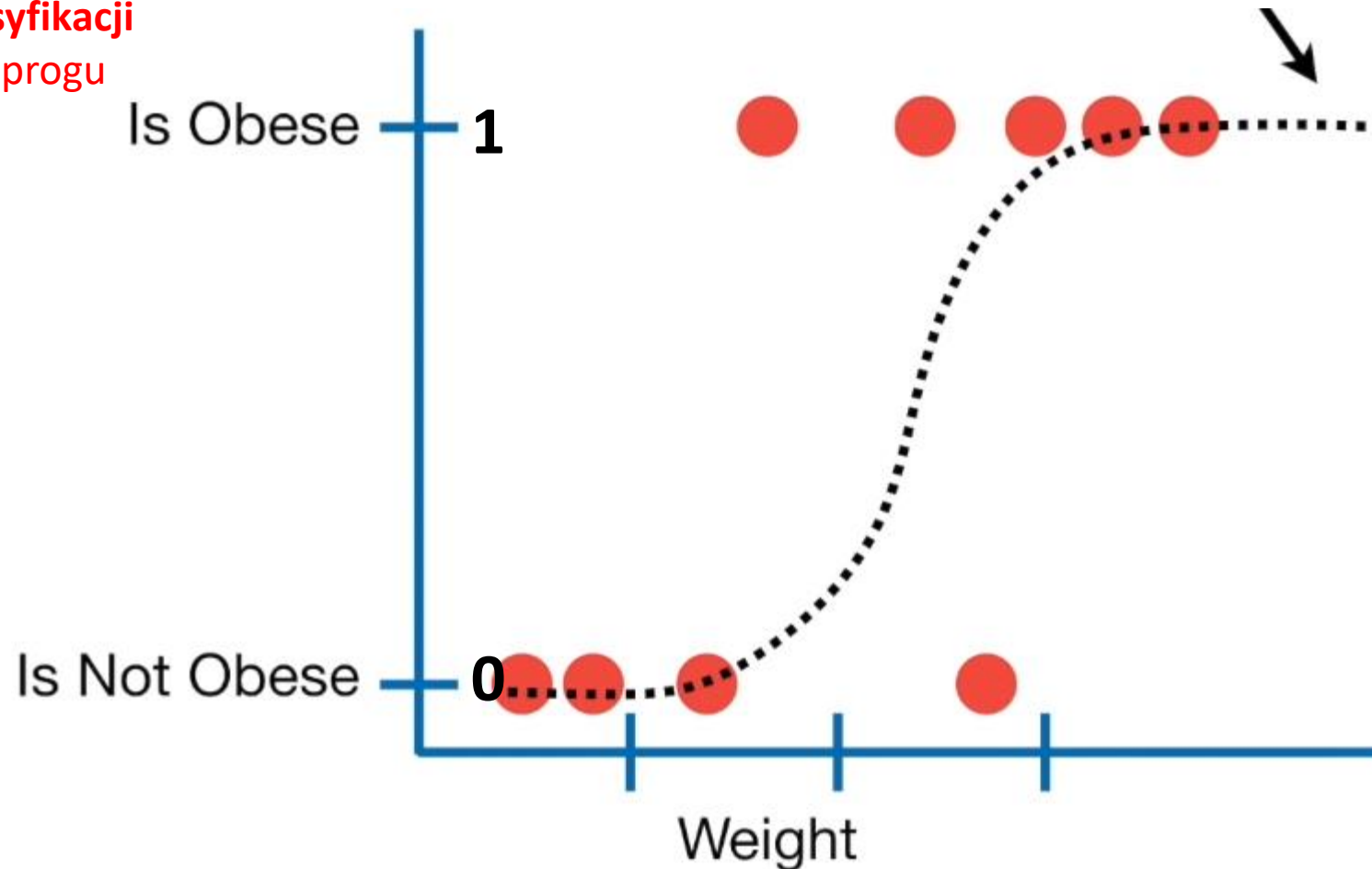
Tradycyjna technika statystyczna także popularna jako narzędzie ML.



1. Obliczenie R^2 i określenie stopnia korelacji
2. Obliczenie **p-value** aby określić czy korelacja jest istotna
3. Wykorzystanie linii do predykcji **Size = f(Weight)**

Logistic regression przewiduje czy coś jest **True** czy **False**

Logistic regression daje
prawdopodobieństwo,
ale najczęściej jest
wykorzystywana do **klasyfikacji**
(jeśli powyżej pewnego progu
prawd.)

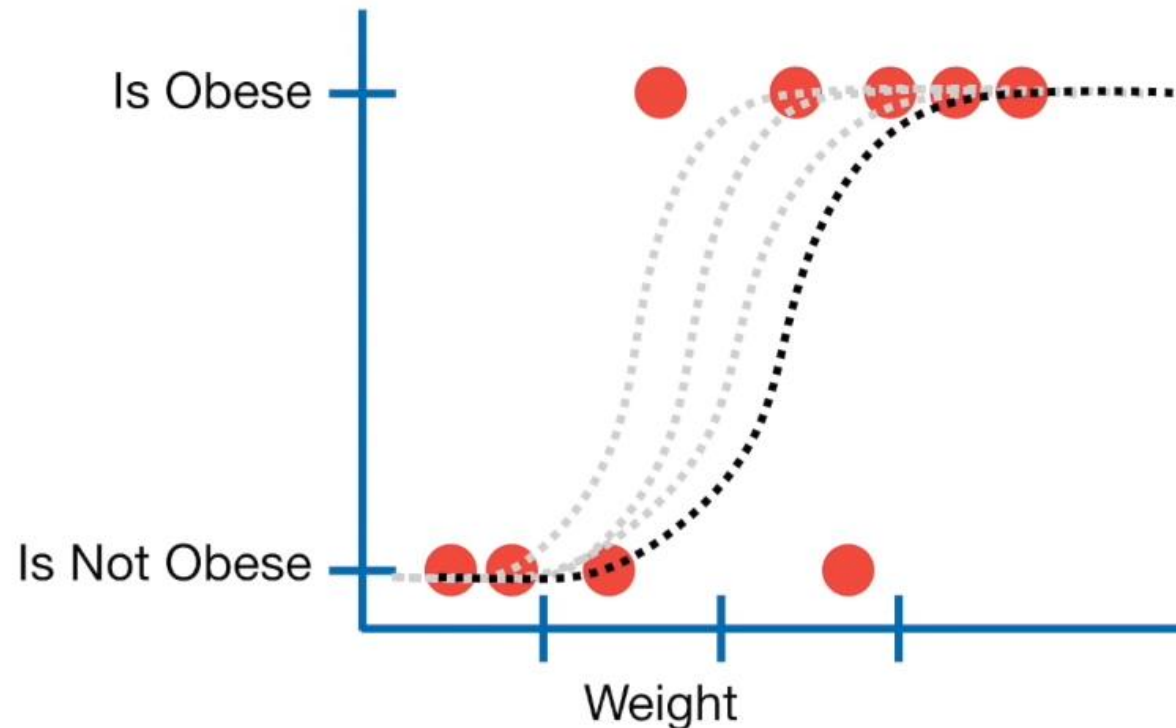


Zamiast dopasowywać linię do danych, logistic regression dopasowuje „S” kształtną „logistic function”. Wartość zmienia się od 0 do 1 i opisuje prawdopodobieństwo, otyłości w zależności od wagi.

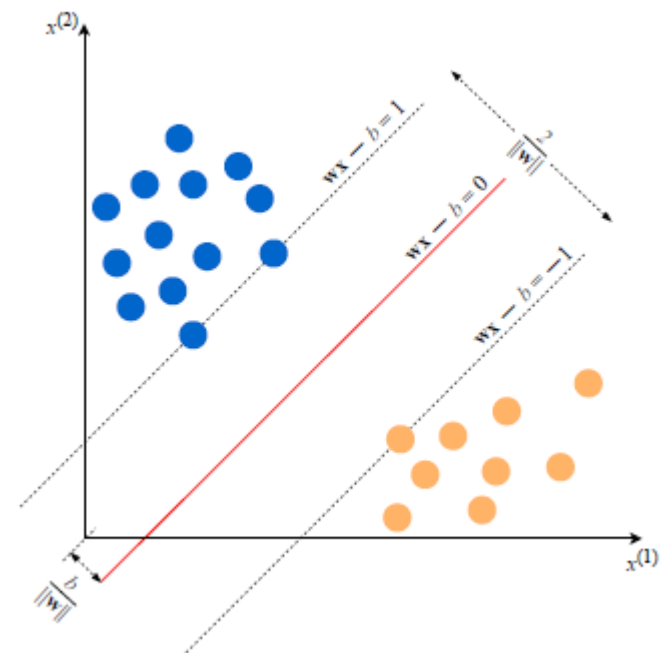
Porównując oba rodzaje regresji (liniową i logistic) istnieje jedna znacząca różnica. W Logistic brak jest miary do porównywania modeli (w przypadku wielu zmiennych) a tylko ocena czy wpływ danej zmiennej jest znacząco różny od zera.

Do tego celu używa się Wald's Test

W Logistic Regression do dopasowania linii używa się metody „maximum likelihood”, dla krzywej kandydata sprawdza się prawdopodobieństwo wszystkich i mnoży przez siebie (z największym jest wybierane).



SVM (Support Vector Machine) - nowa generacja „learning” algorytmów



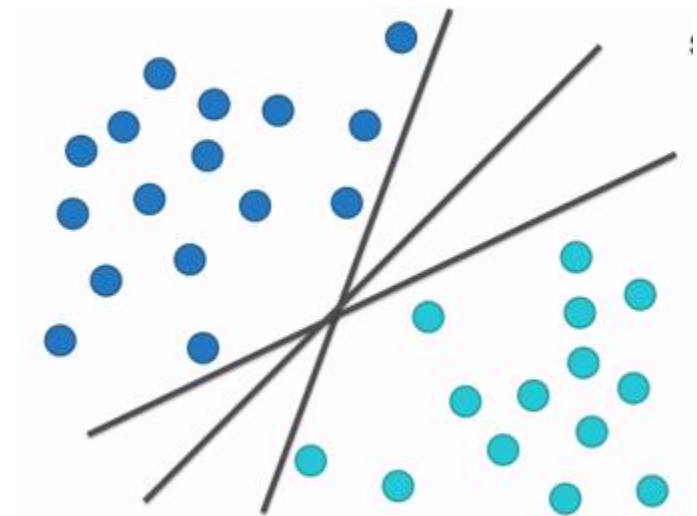
SVM (Support Vector Machine) - nowa generacja „learning” algorytmów

Dwie nowe podstawowe idee.

Nowy efektywny sposób rozdzielania nie-liniowych regionów za pomocą „kernel functions”

Wykorzystanie „quadratic optimization” do rozwiązania problemu „local minimum”

1. Support Vectors – to - **data points** które leżą najbliżej **decision surface** (lub **hyperplane**)
1. Są to najczęściej data points, które są najtrudniejsze do klasyfikacji
2. Mają one bezpośredni wpływ na lokalizację **decision surface**
3. **SVM – znajduje optymalne rozwiązanie**



2D – linia, wyższy wymiar -hyperplanes

SVM maksymalizuje „margin” (według innych terminologii „road” lub „street” naokoło rozdzielającej obie klasy „hyperplane”).

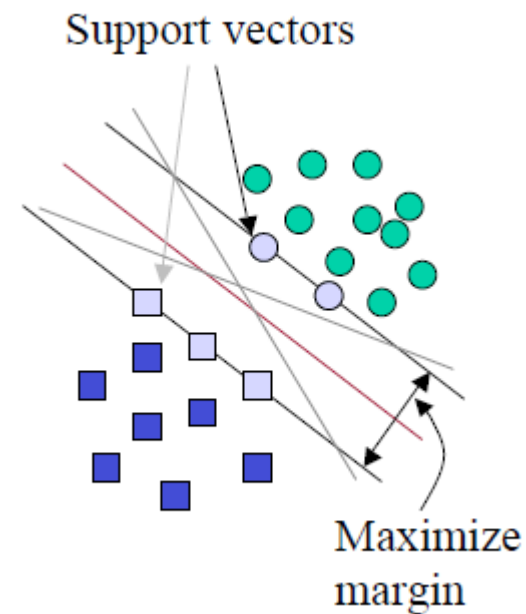
„Decision function” jest całkowicie określona przez podzbiór danych treningowych (support vectors).

Problem może być rozwiązany standardową metodą (Quadratic programming problem).

Pierwsze przybliżenie – linear separability

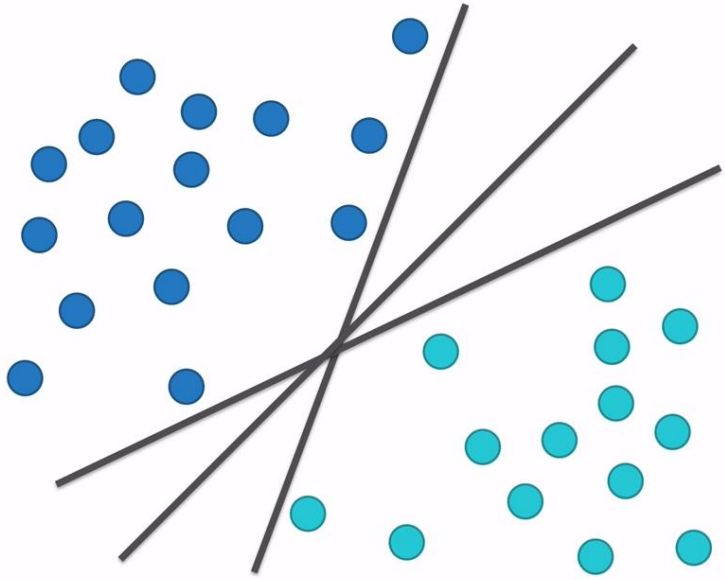
Input – zbiór input sample features x_1, x_2, \dots, x_n i ich rezultatów y (najczęściej jest ich dużo).

Output – zbiór wag \mathbf{w} (lub w_i) jedna dla każdego features, których liniowa kombinacja daje predykcję wartości y



Support Vector Machines

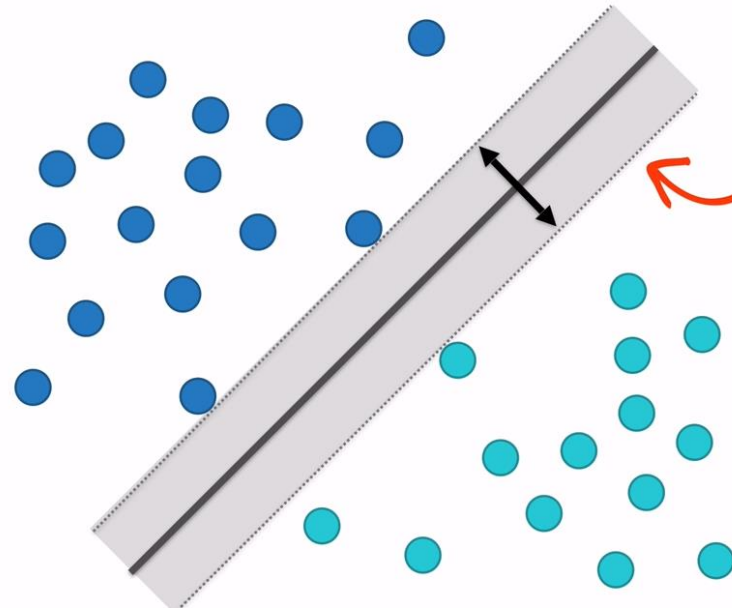
Split the data



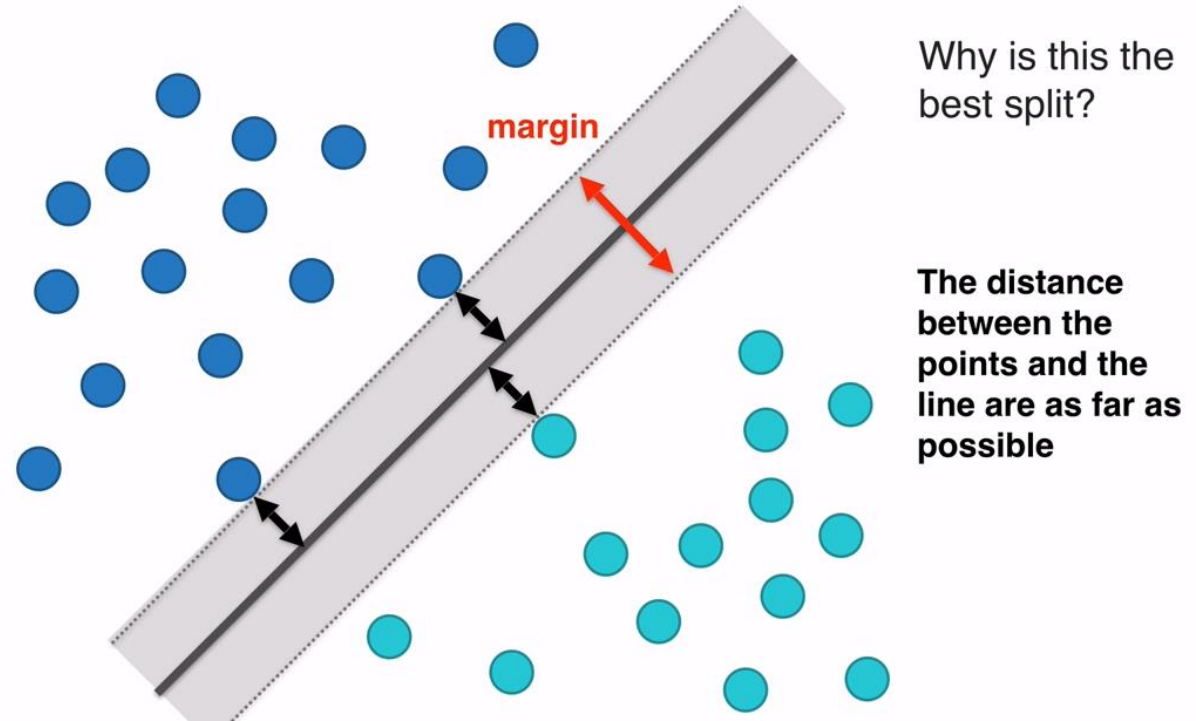
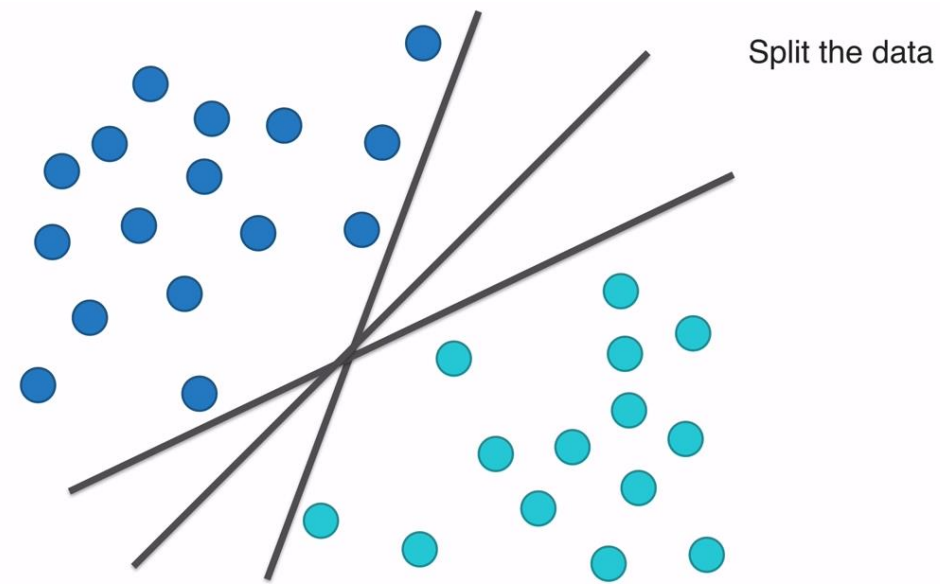
Why is this the best split?

This is the widest road that separates the two groups

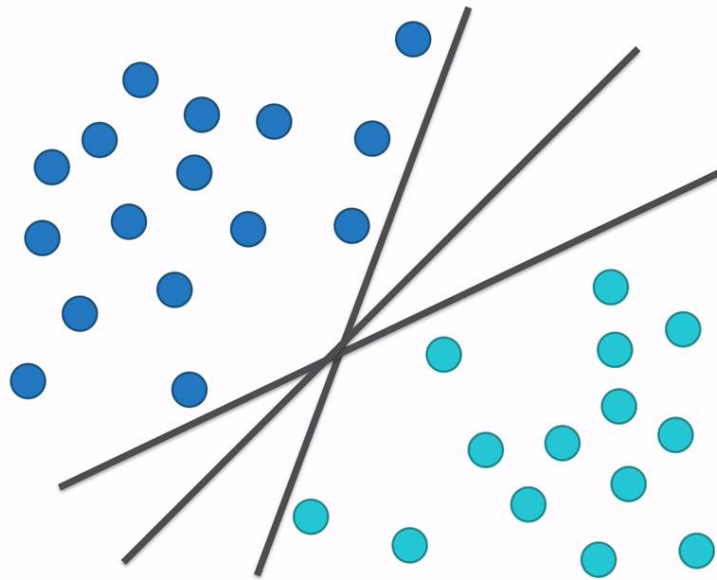
This is the widest **margin** that separates the two groups



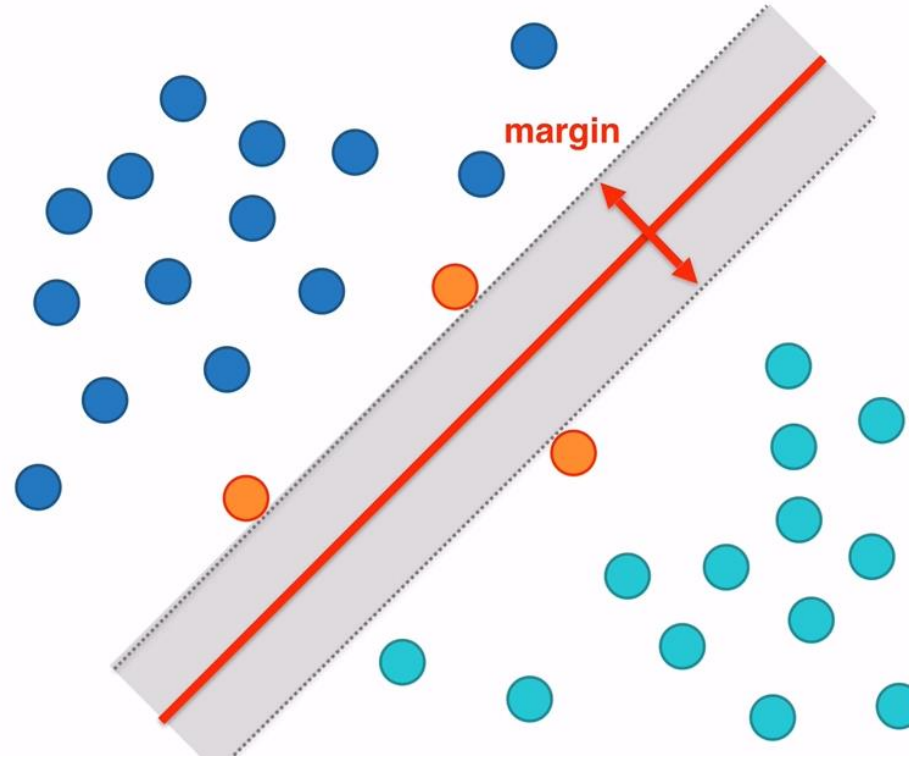
Support Vector Machines



Support Vector Machines

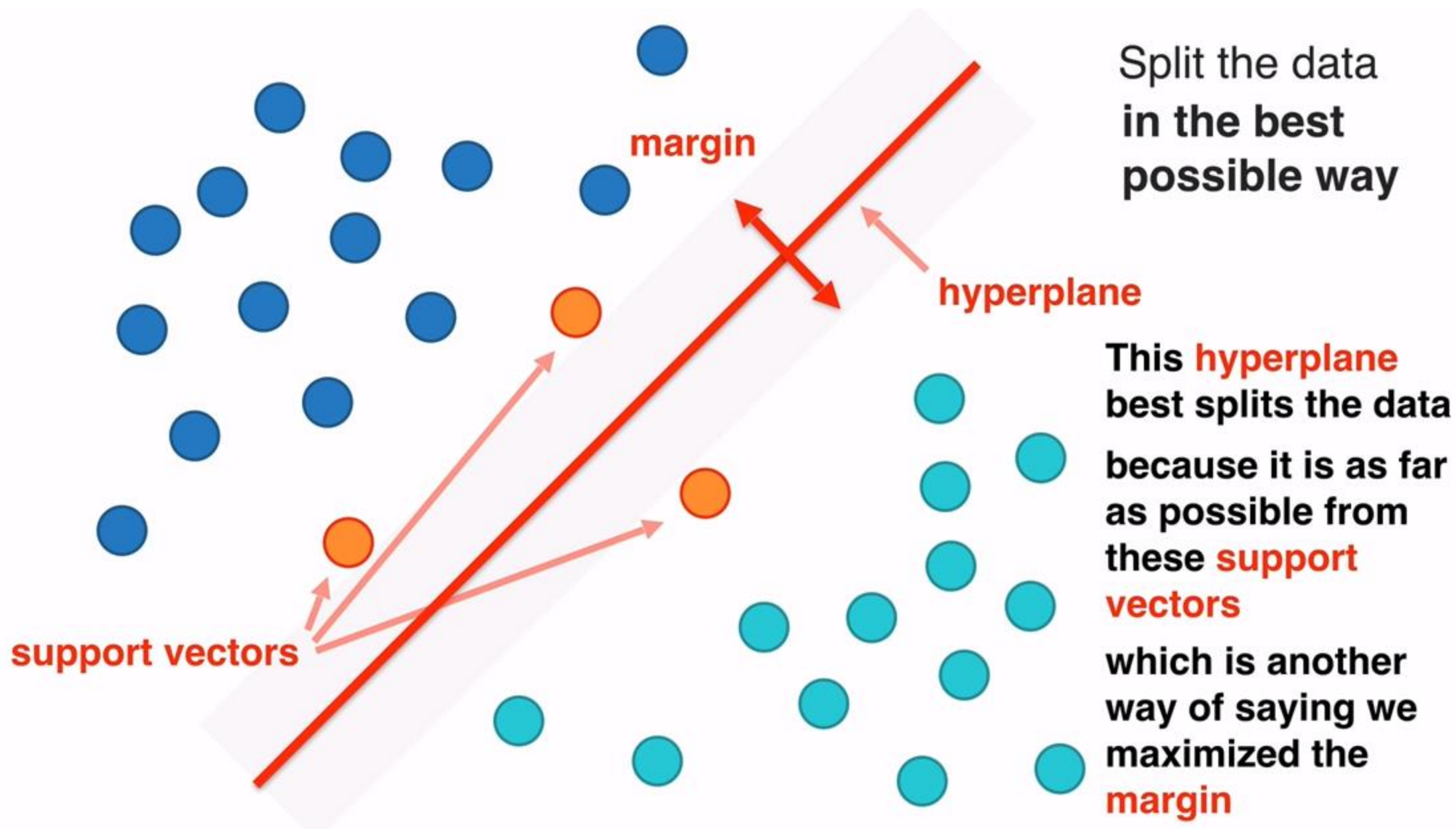


Split the data



Why is this the best split?

The distance between the **support vectors** and the **hyperplane** are as far as possible



Maksymalizacja of margin jest to **constrained optimization** problem, który może być rozwiązany za pomocą **Lagrange Multipliers** technique

Ważne:

SVM jest wrażliwy na feature scales.
Powinno być przeprowadzane feature scaling (StandardScaler)

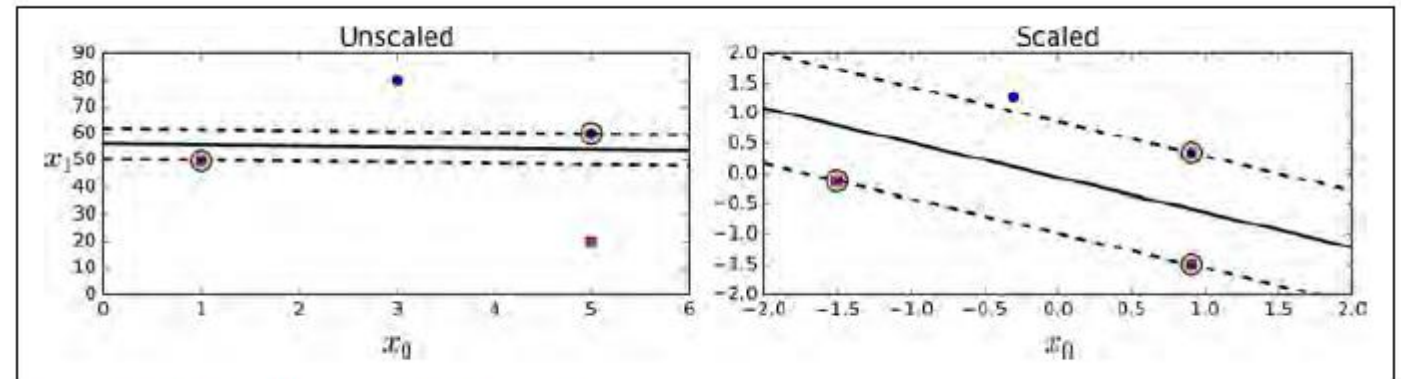


Figure 5-2. Sensitivity to feature scales

Soft and hard margin classification
(równowaga pomiędzy szerokością
pasa i ograniczeniem margin
violations).

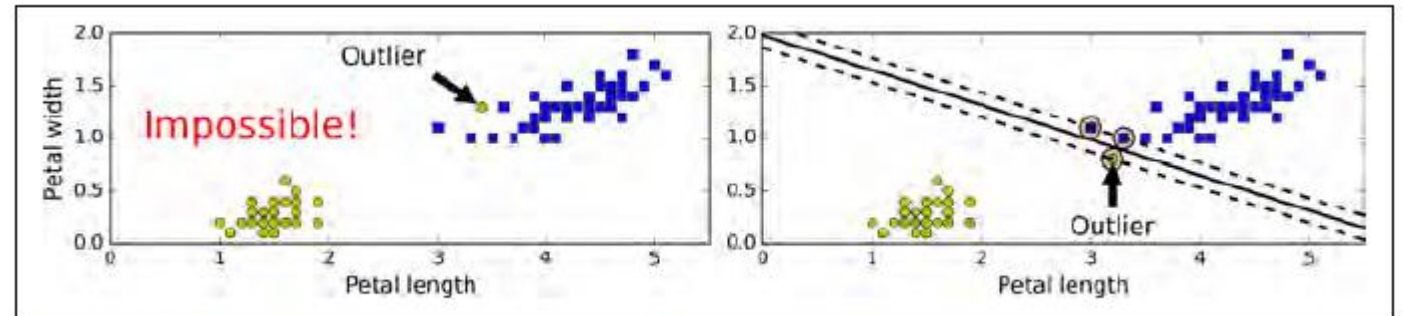
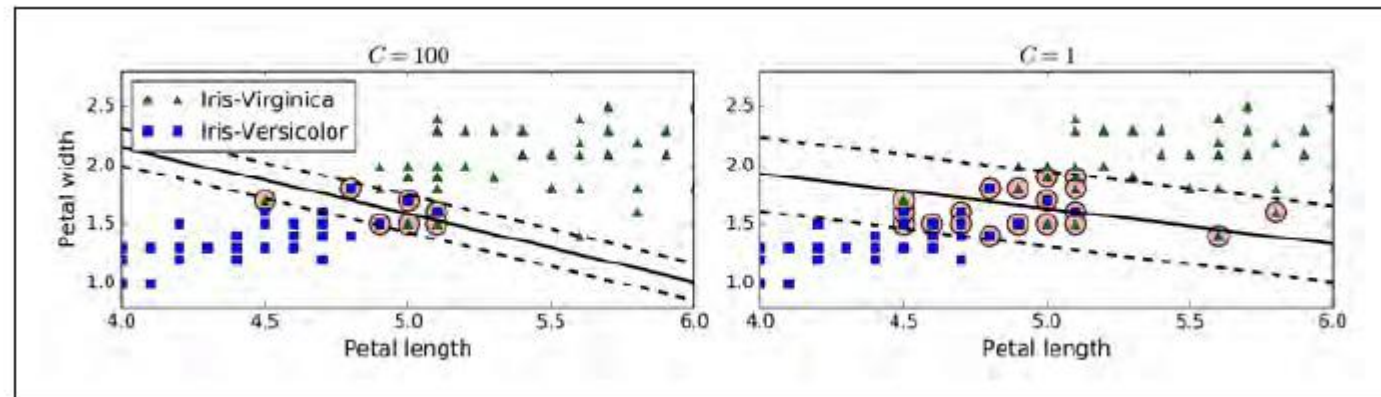


Figure 5-3. Hard margin sensitivity to outliers

C – hyperparameter (male – szeroki pas większe violations)



Nonlinear SVM Classification

Adding polynomial features

Generate a new **feature** matrix consisting of all **polynomial** combinations of the **features** with degree less than or equal to the specified degree

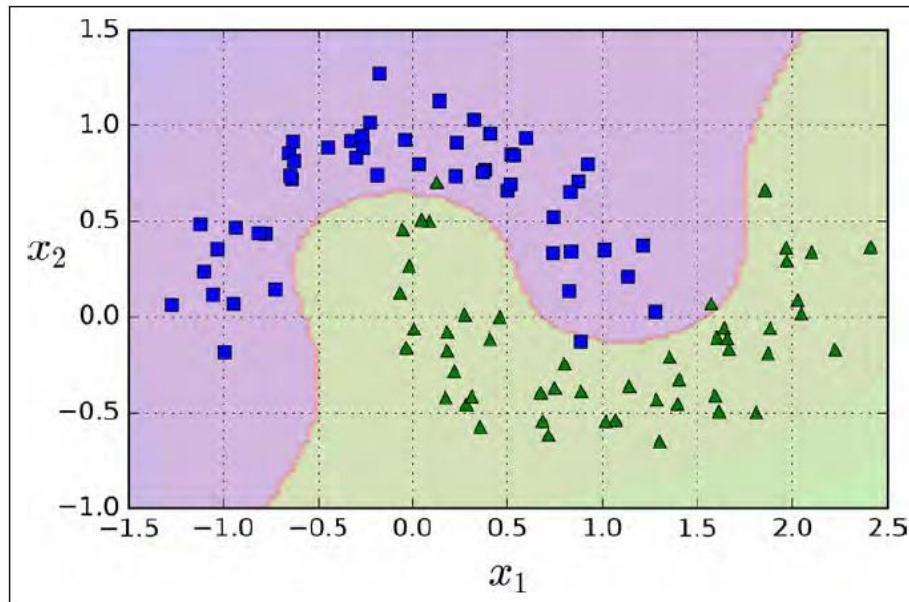


Figure 5-6. Linear SVM classifier using polynomial features

kernel trick

analogiczny rezultat

hyperparameter coef0

Kontroluje w jakim stopniu na algorytm mają wielomiany różnego stopnia

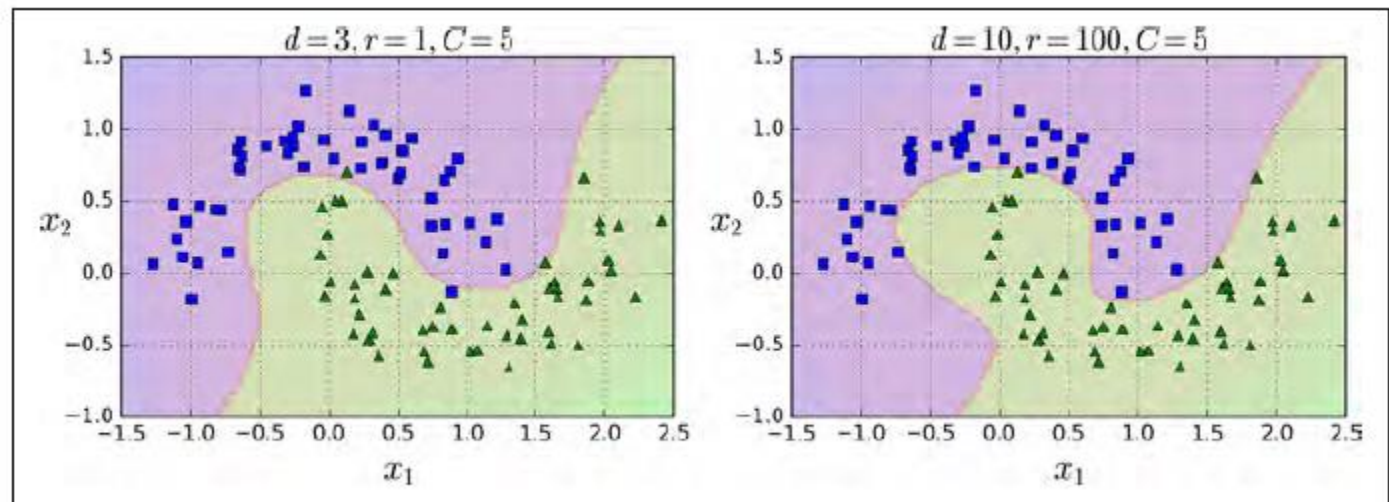


Figure 5-7. SVM classifiers with a polynomial kernel