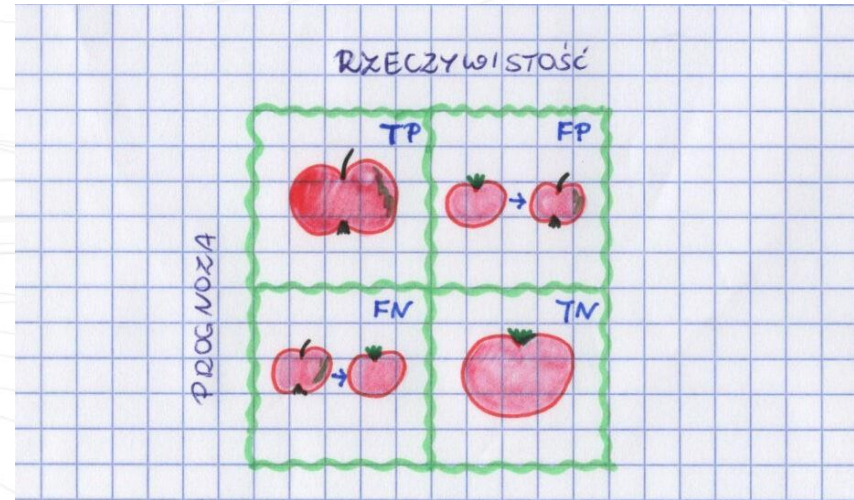


# SVM



# Algorytmy klasyfikacyjne

Algorytmy klasyfikacyjne wskazują jedną z wartości kategoriycznych. Jeśli w danym problemie algorytm prognozuje jedną z dwóch wartości (prawda/fałsz, 0/1, tak/nie), mówimy o klasyfikacji dwuklasowej (ang. *two-class classification*). W przypadku problemów, w których zmienna predykcyjna może przyjąć więcej niż dwie wartości, mówimy o klasyfikacji wieloklasowej (ang. *multiclass classification*).





## Macierz pomyłek – confusion matrix

	precision	recall	f1-score	support
english	1.00	0.92	0.96	25
polish	0.82	1.00	0.90	14
spanish	1.00	0.93	0.97	15
accuracy			0.94	54
macro avg	0.94	0.95	0.94	54
weighted avg	0.95	0.94	0.95	54

**Support** – w raporcie mówi ile przykładów z danej kategorii zostało wylosowanych do grupy testowej.

**Accuracy** – dowiadujemy się, jaka część tekstów, ze wszystkich zaklasyfikowanych, została zaklasyfikowana poprawnie. Czyli sumę poprawnych klasyfikacji z danej kategorii (TP) oraz poprawnej z innych kategorii (TN) dzielimy przez liczbę wszystkich klasyfikowanych przypadków.

**Recall** – Czulość mówi nam o tym, jaki jest udział prawidłowo zaprognozowanych przypadków pozytywnych (TP) wśród wszystkich przypadków pozytywnych (również tych, które błędnie zostały zaklasyfikowane do negatywnych – FN).

**Precision** – TP+FP. W efekcie dowiadujemy się, ile wśród przykładów zaprognozowanych pozytywnie jest rzeczywiście pozytywnych.

**F1-score** to średnia harmoniczna pomiędzy precyzją (precision) i czulością (recall). Im bliższa jest jedynki, tym lepiej to świadczy o algorytmie klasyfikującym. W najlepszym przypadku przyjmuje wartość 1, kiedy mamy do czynienia z idealną czulością i precyzją.

The logo consists of three circles: a small cyan circle at the top left, a small purple circle at the bottom left, and a larger red circle to their right.

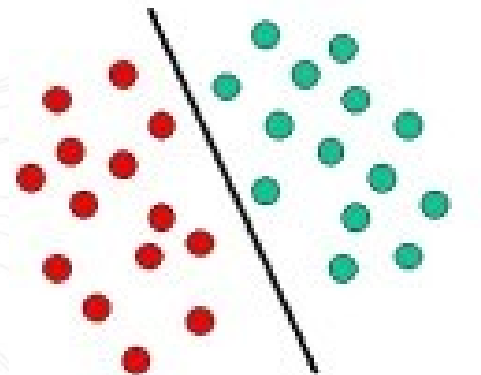
# Support Vector Machine

Maszyna Wektorów Nośnych



Opracowane w AT&T Bell Laboratories przez Vladimira Vapnika z kolegami w latach 90 tych oparte na statystycznych ramach uczenia się lub teorii VC zaproponowanej przez Vapnika (1982, 1995) i Chervonenkisa (1974).

Mając zestaw przykładów uczących, każdy oznaczony jako należący do jednej z dwóch kategorii, algorytm uczący SVM buduje model, który przypisuje nowe przykłady do jednej lub drugiej kategorii (lub też do większej ich ilości), czyniąc go nieprobabilistycznym binarnym klasyfikatorem liniowym.







“Wyobraźmy sobie, że na stole są poustawiane piłki o dwóch kolorach. Czerwone po jednej stronie stołu a niebieskie po drugiej. Naszym zadaniem jest takie położenie kijka na stole, żeby oddzielał te dwie grupy.

Trzeba to tak zrobić, aby odległość między kijkiem a każdą piłką była możliwie największa. A więc to proste! Jednak po chwili ktoś dokłada niebieską kulkę w miejsce czerwonych. No to teraz nie znajdziemy na świecie takiego kijka, który by nam rozdzielił te dwie grupy. Musiałby być niesamowicie wygięty.

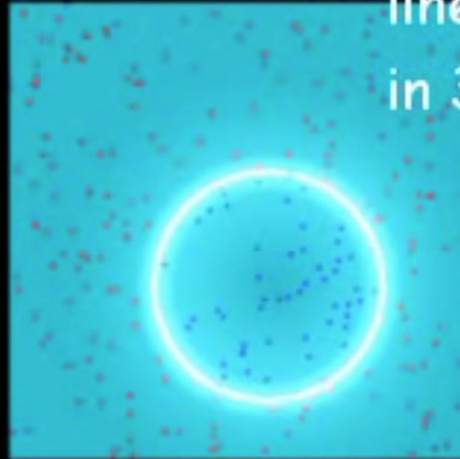
Ale na szczęście możemy zrobić pewną sztuczkę: wywrócimy stół, a gdy wszystkie piłki będą w powietrzu przetniemy je kartką papieru ruchem sprawnego wojownika Ninja. W ten sposób rozdzielimy piłki na dwie grupy.

Poważni dorośli ludzie na piłki mówią dane, na kijek i kartkę – klasyfikator albo hiperpłaszczyzna, na ułożenie kijka w największej odległości od piłek – optymalizacja, a na wywrócenie stołu – transformacja za pomocą funkcji jądrowych.”

<https://tomaszkacmajor.pl/index.php/2016/04/17/support-vector-machine/>



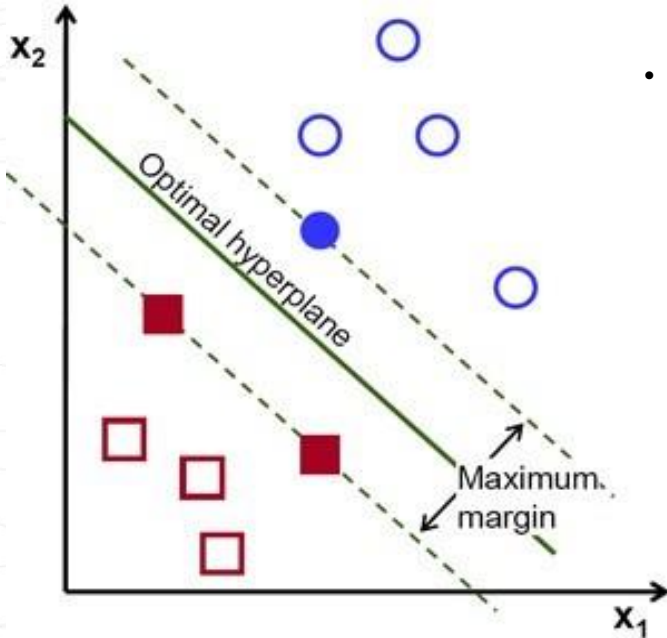
Using a polynomial  
kernel they can be  
linearly separated  
in 3D space





# Hiperplane – hiperpłaszczyzna

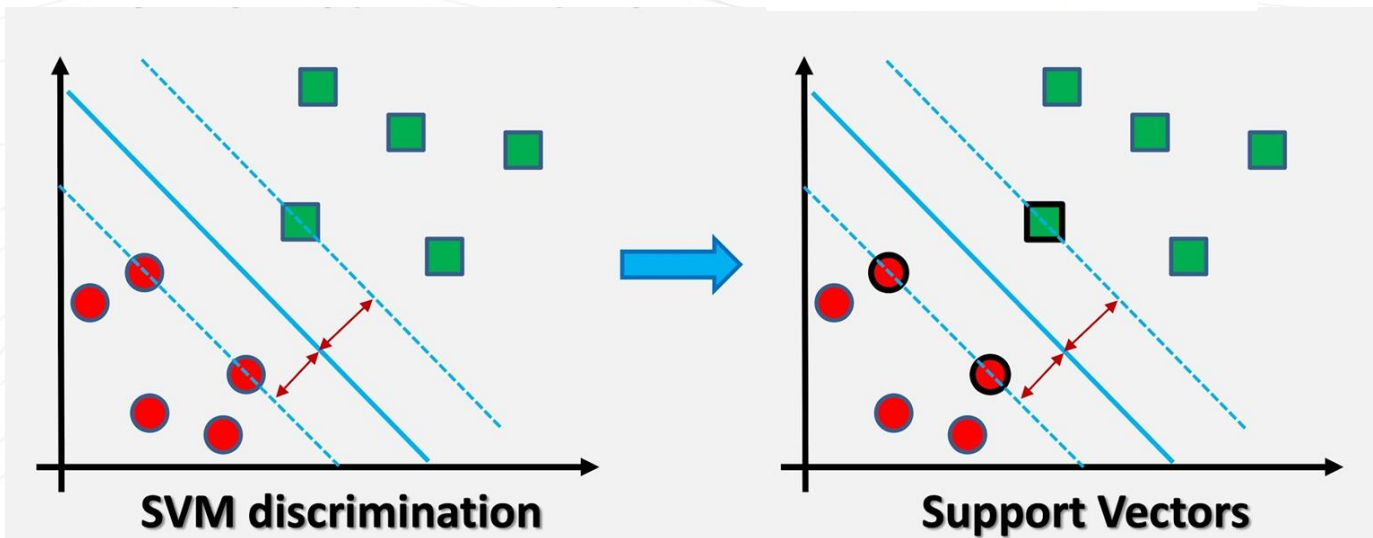
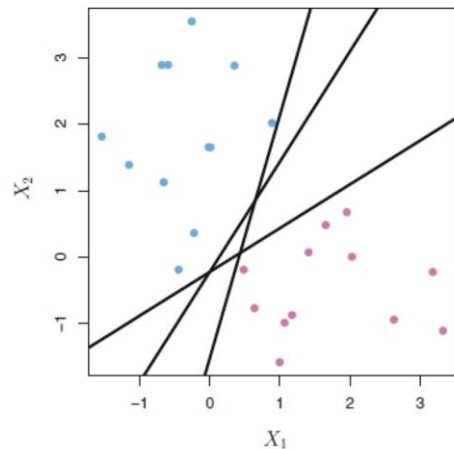
- Linia oddzielająca klasy to optimal hiperplane – hiperpłaszczyzna.
- Wektory na których się ją “wspiera” to tzw. Support vectors – wektory nośne. Są to separowalne punkty w przestrzeni, które są najczęściej skrajnymi elementami klas.
- Maximum margin – maksymalny margines. Celem algorytmu jest uzyskanie największego marginesu pomiędzy klasami. Im większy tym lepiej bo są wtedy maksymalnie “różne”.

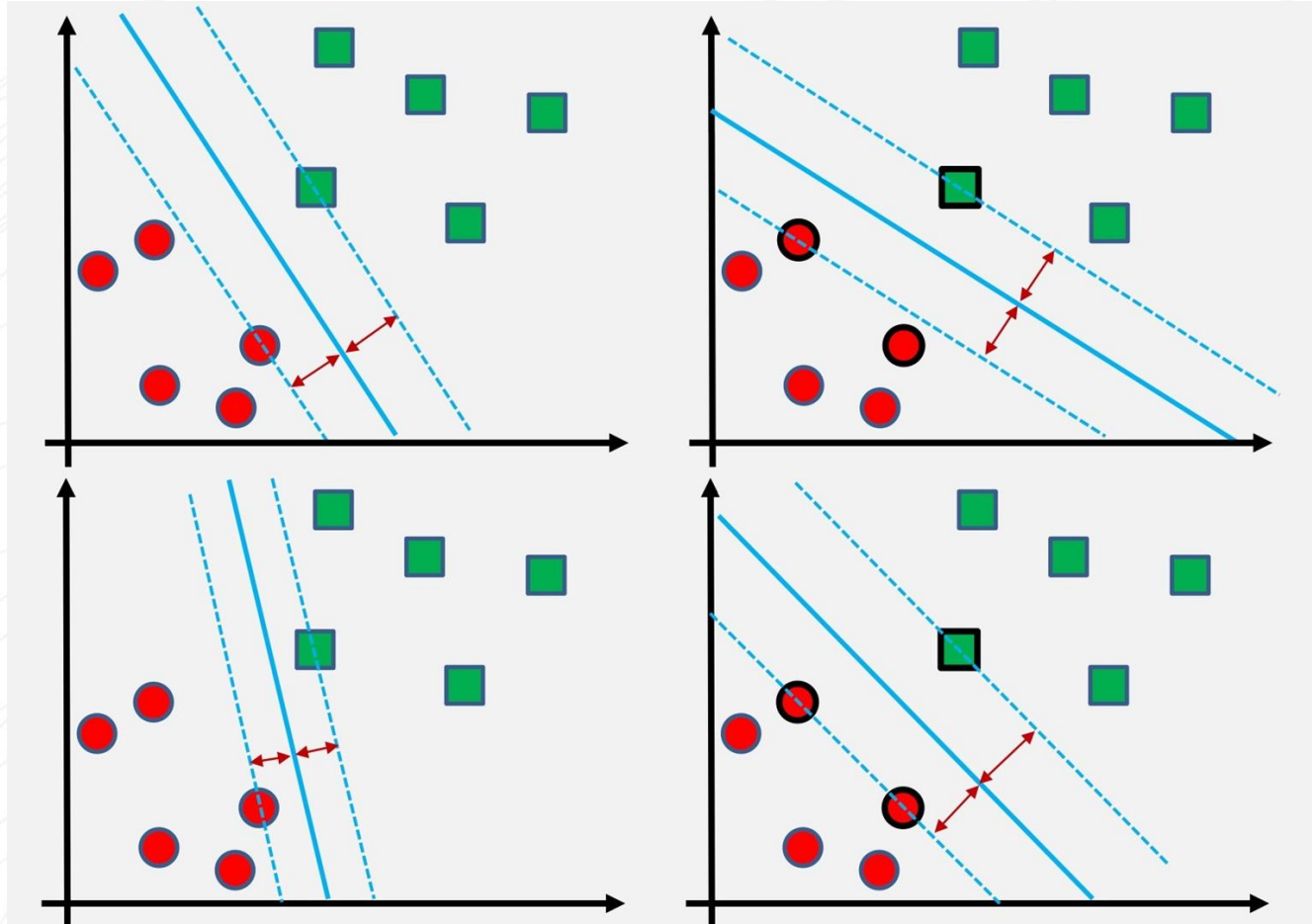


SVM odwzorowuje przykłady treningowe na punkty w przestrzeni, aby zmaksymalizować szerokość luki między dwiema kategoriami. Nowe przykłady są następnie mapowane w tej samej przestrzeni i przewiduje się, że będą należeć do kategorii na podstawie tego, po której stronie luki się znajdują.

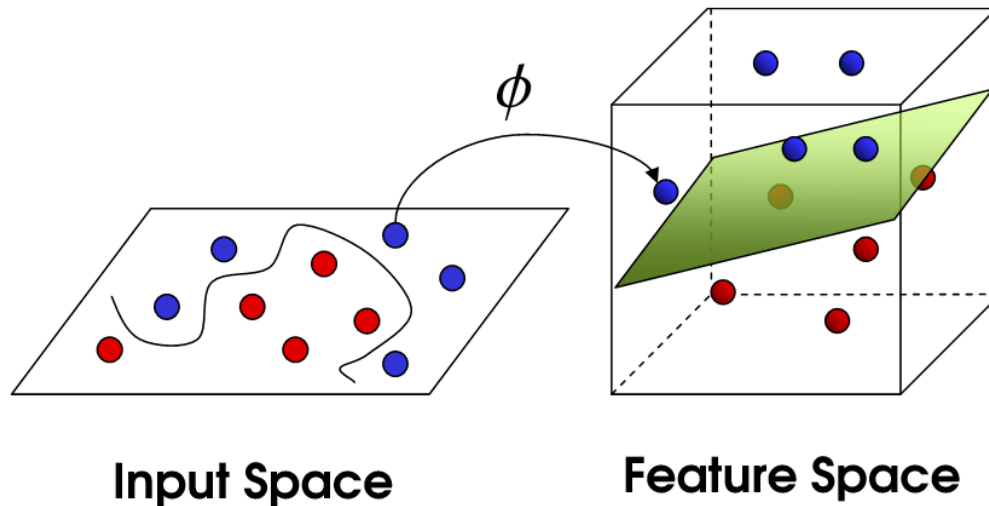


Możliwe jest narysowanie wielu hiperpłaszczyzn (rys. po prawej) ale algorytm szuka optymalnej.





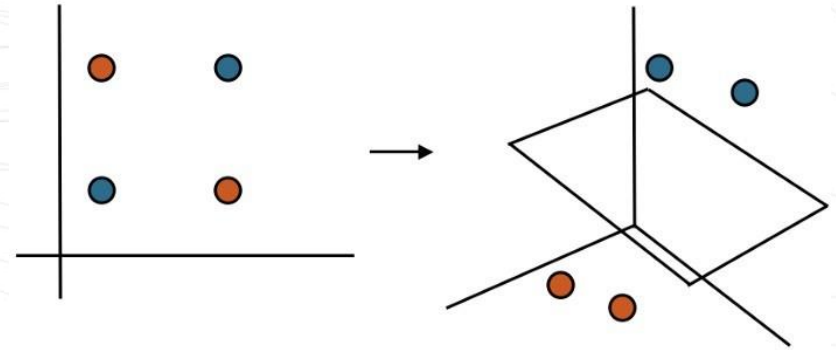
# Kernel Trick (SVM)



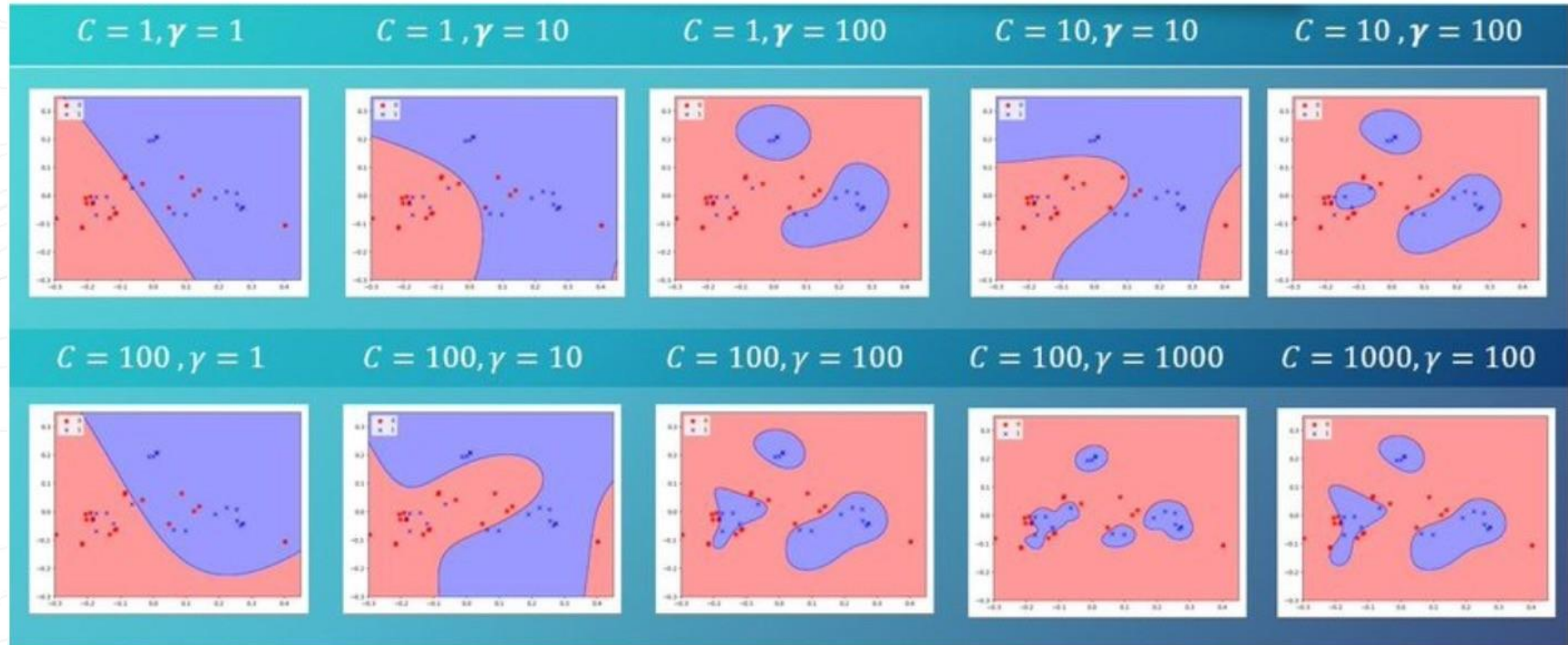
Jądro – funkcja matematyczna służąca do przekształcania danych wejściowych w inną formę. Typowe funkcje jądra obejmują liniowe, nieliniowe, wielomianowe itp.



Maszyna wektorów nośnych klasyfikuje dane wykorzystując niejawnie przekształcenie zbioru treningowego do przestrzeni cech wyższego wymiaru. W nowej przestrzeni cech dopasowywana jest optymalna hiperpłaszczyzna rozdzielająca dwie klasy danych i jednocześnie maksymalizuje margines pomiędzy hiperpłaszczyzną, a punktami znajdującymi się najbliżej niej, nazywanymi wektorami nośnymi.



# C oraz gamma





# Zalety oraz wady

## Zalety

- Wszechstronny do funkcji jądra specyficznych dla użytkownika
- Wydajna pamięć
- Skuteczny w przypadkach, gdy liczba wymiarów jest większa niż liczba próbek (więcej zmiennych jak obserwacji)
- Rozwiązuje problemy liniowe i nieliniowe

## Wady

- Podatny na błędy i nadmierne dopasowanie w przypadku zaszumionych danych (np. nakładające się funkcje dla różnych etykiet) oraz outliery
- Długi czas obliczeń w przypadku bardzo dużych zbiorów danych
- Nie daje probabilistycznego wyjaśnienia wyników