

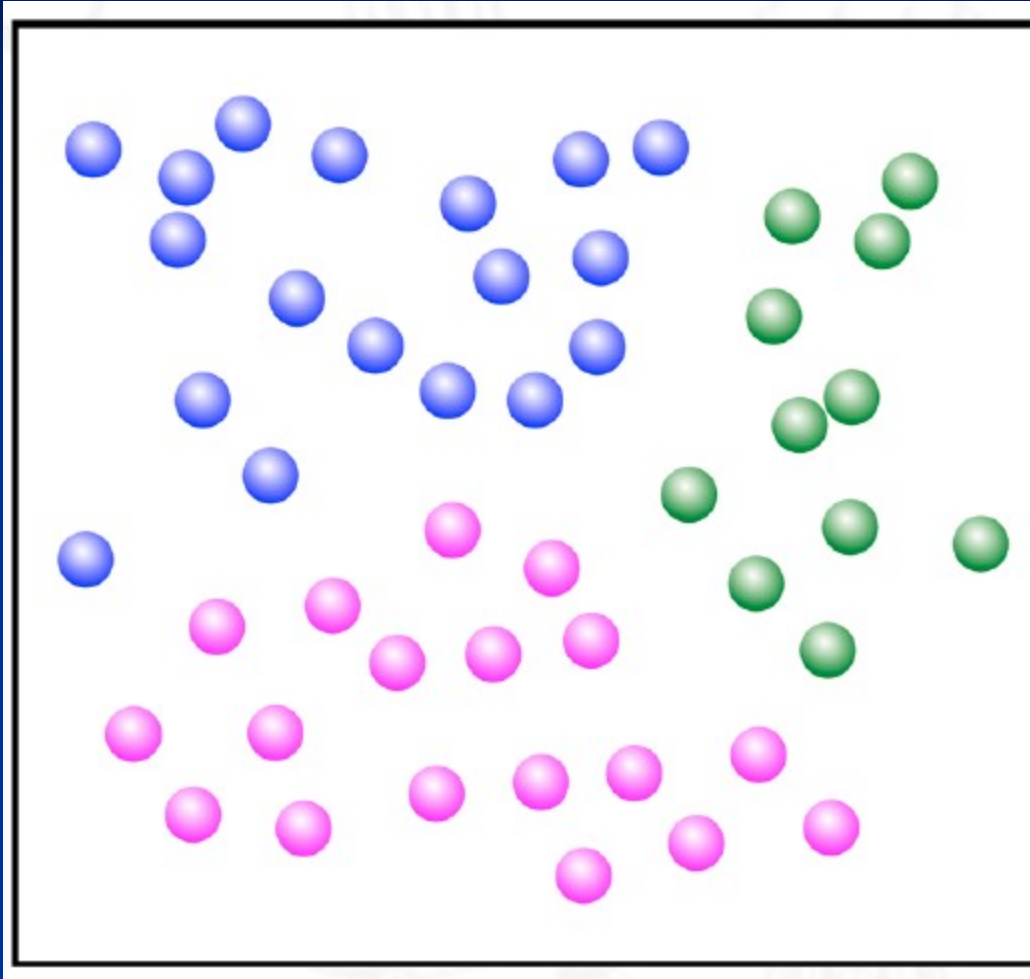
Drzewa decyzyjne, las losowy

Wykorzystane materiały:
wykłady W. Rudnickiego, PB

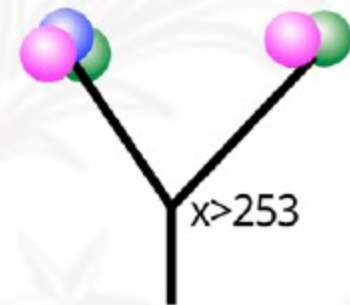
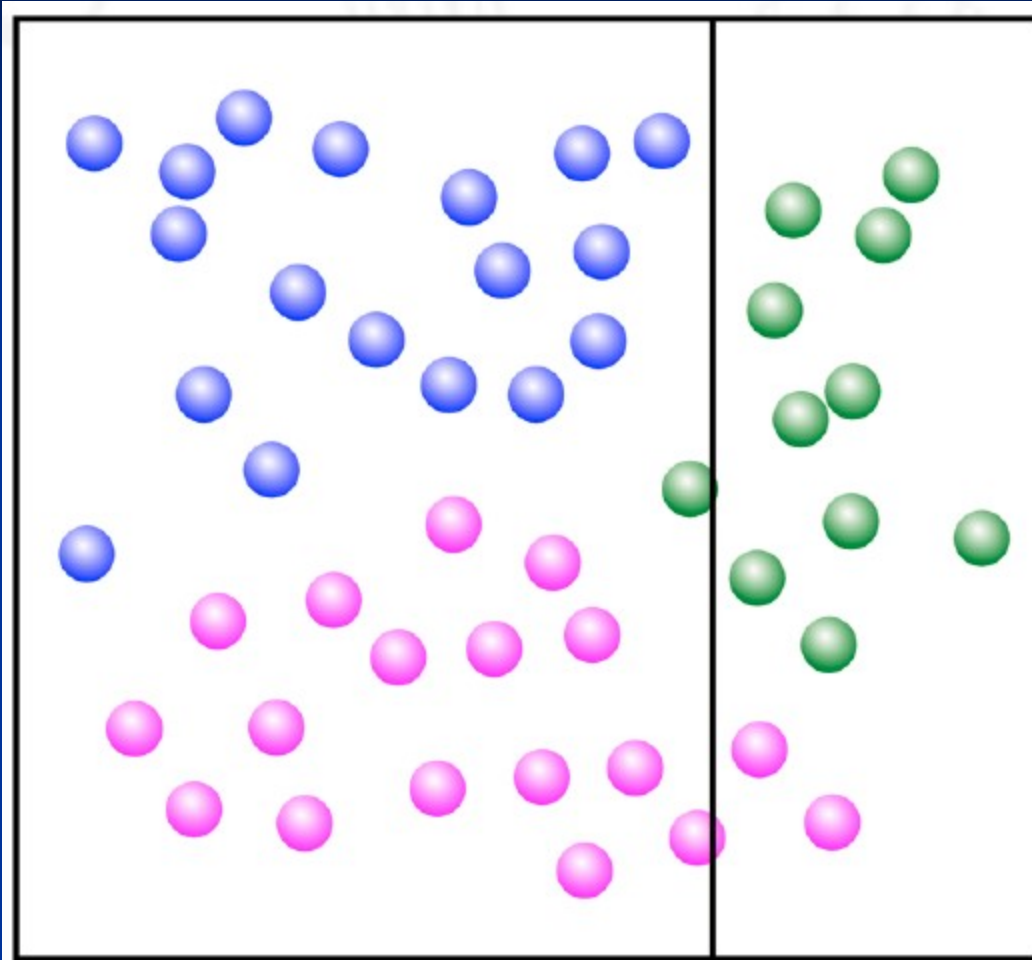
Drzewo decyzyjne (ang. *Decision tree*)

- Chcemy dokonać klasyfikacji danego wektora cech
- Mamy do dyspozycji przykłady uczące
- Metoda polega na **cyklicznym podziale wektorów uczących wg. wybranych kryteriów**

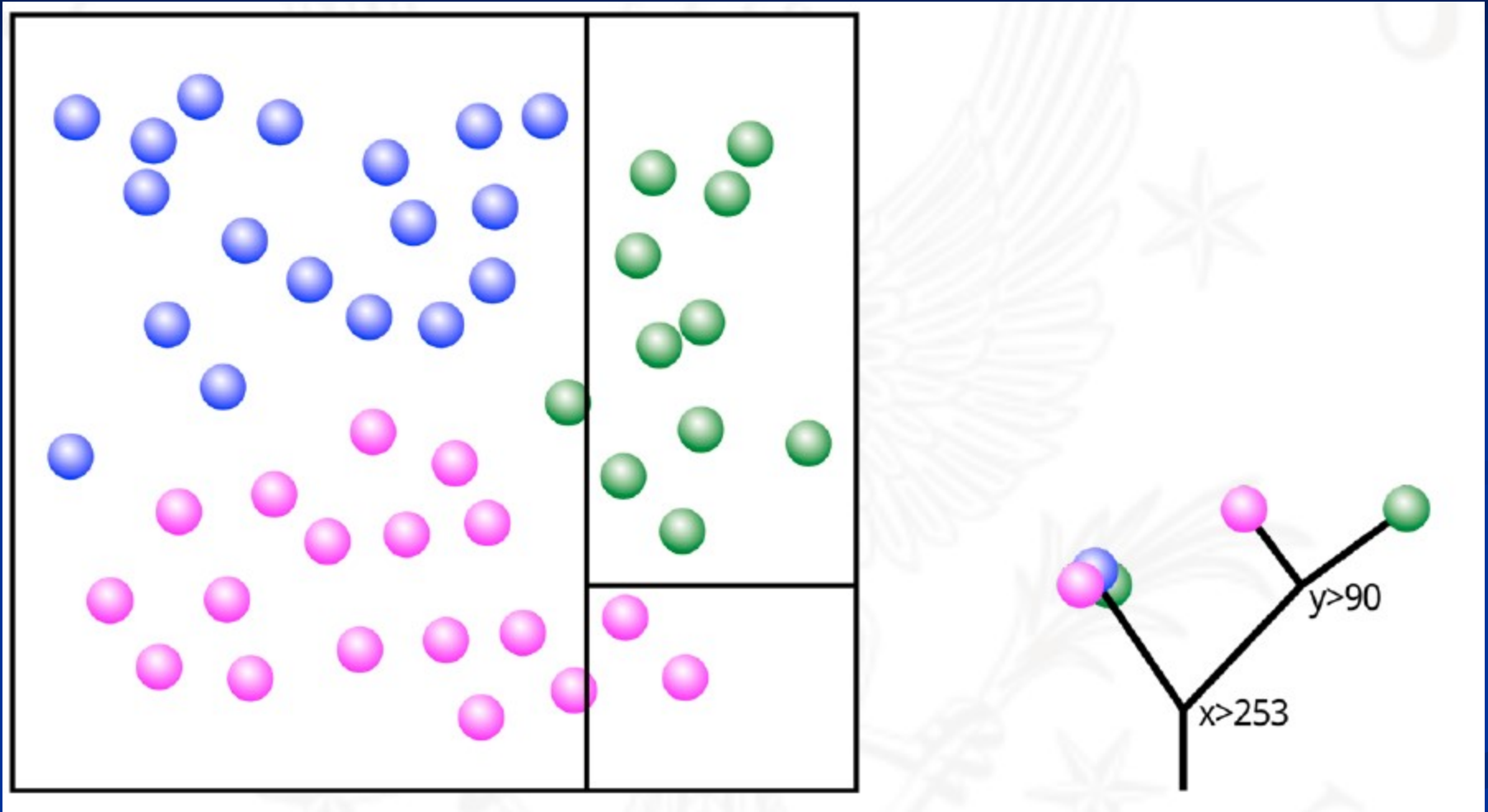
Drzewo decyzyjne (1)



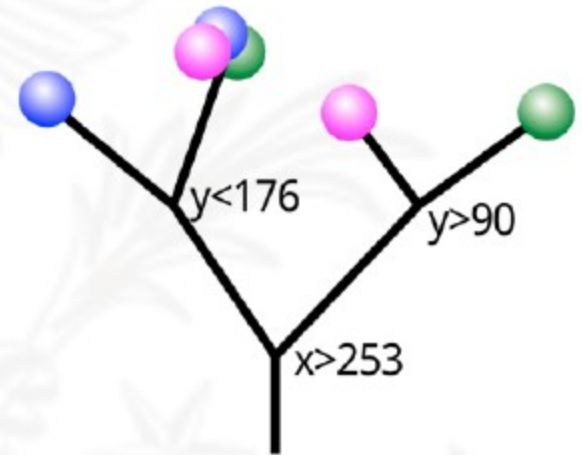
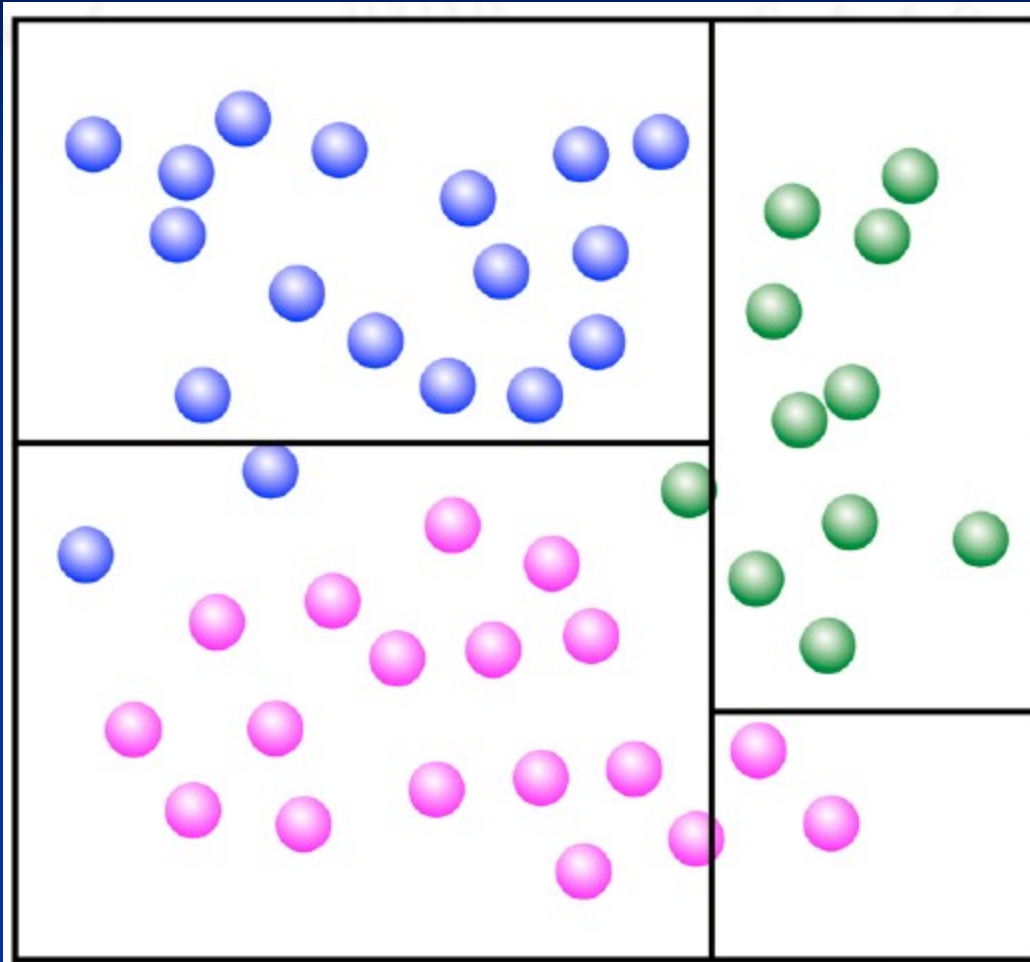
Drzewo decyzyjne (2)



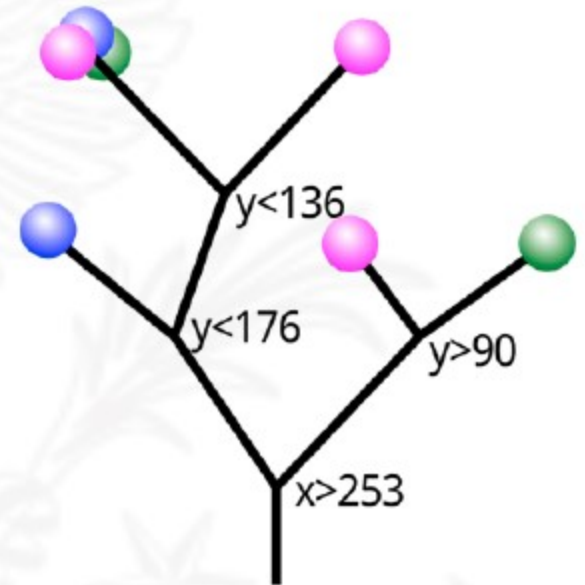
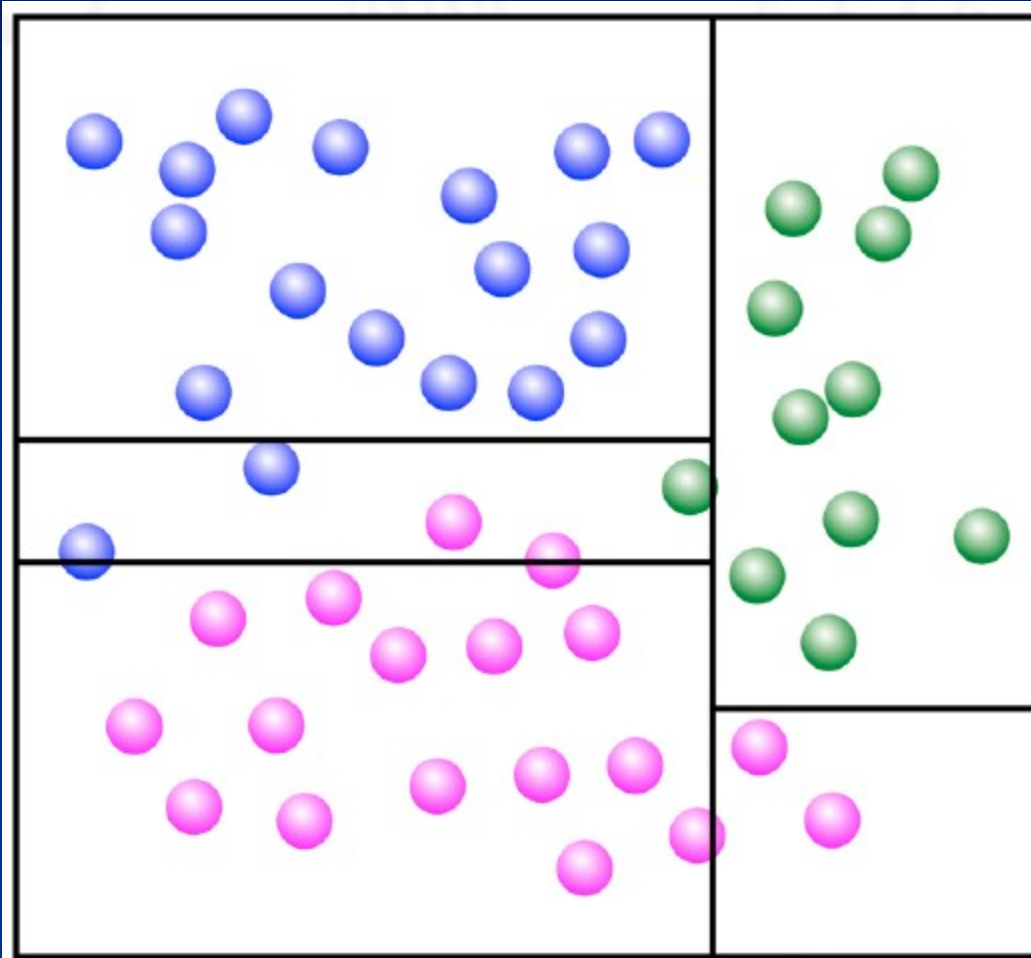
Drzewo decyzyjne (3)



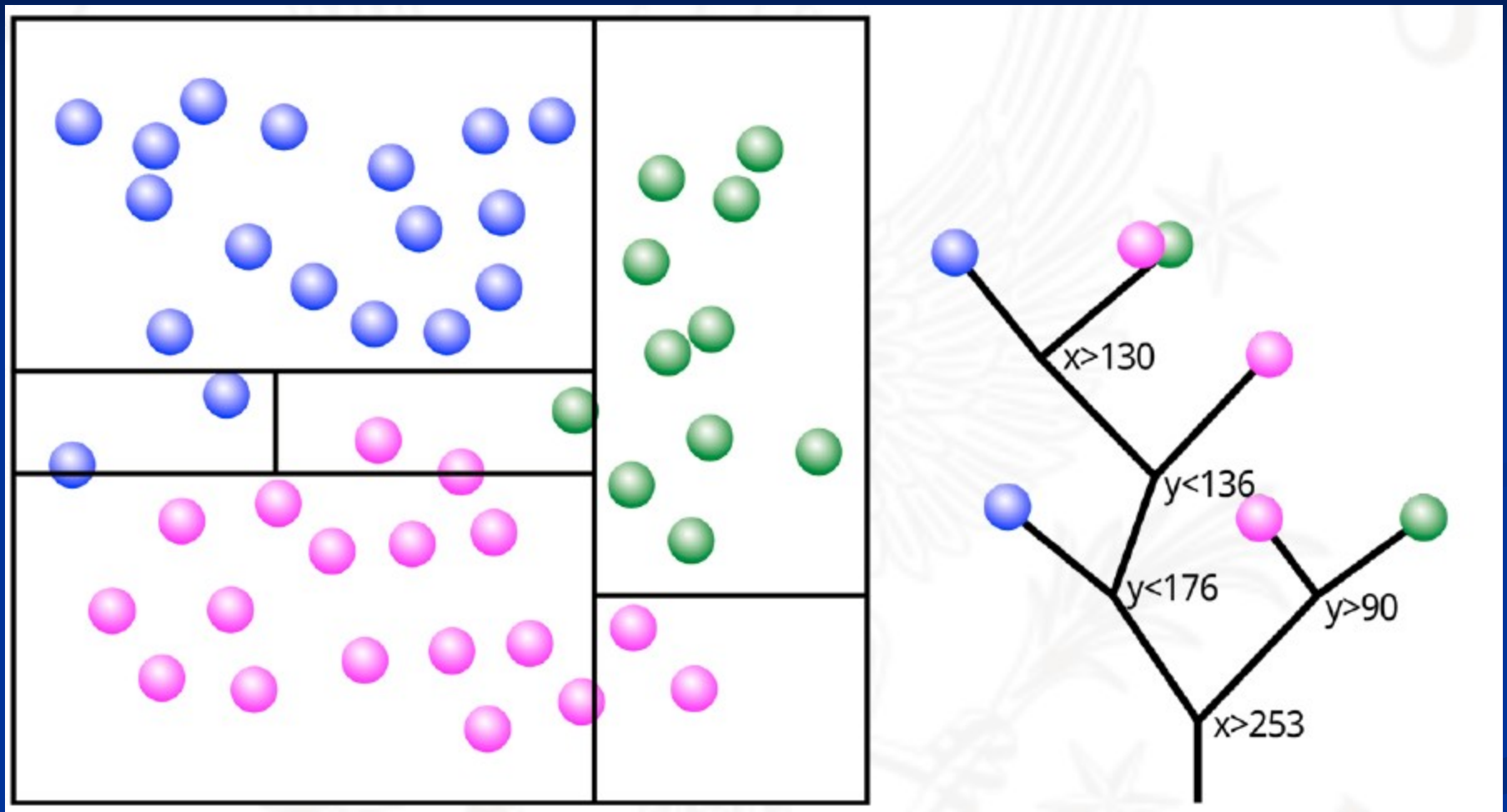
Drzewo decyzyjne (4)



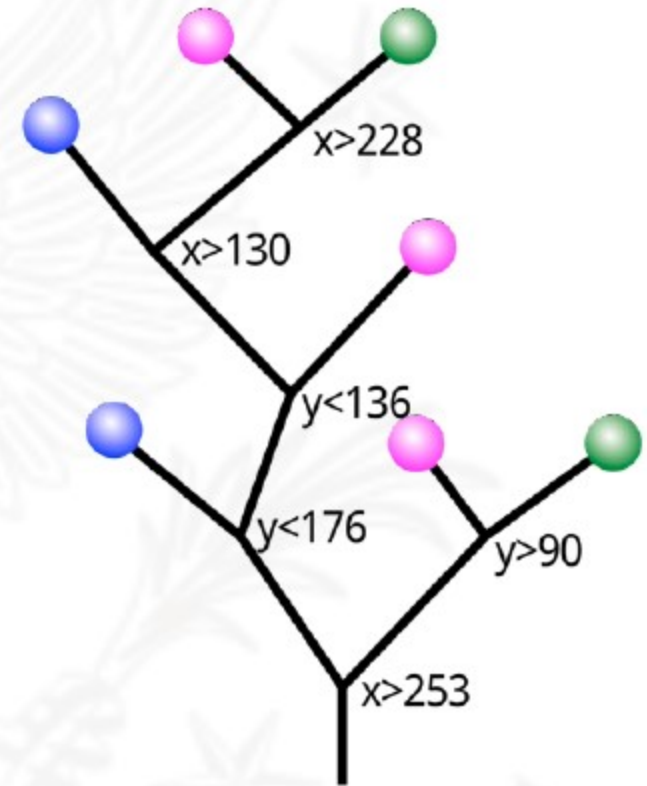
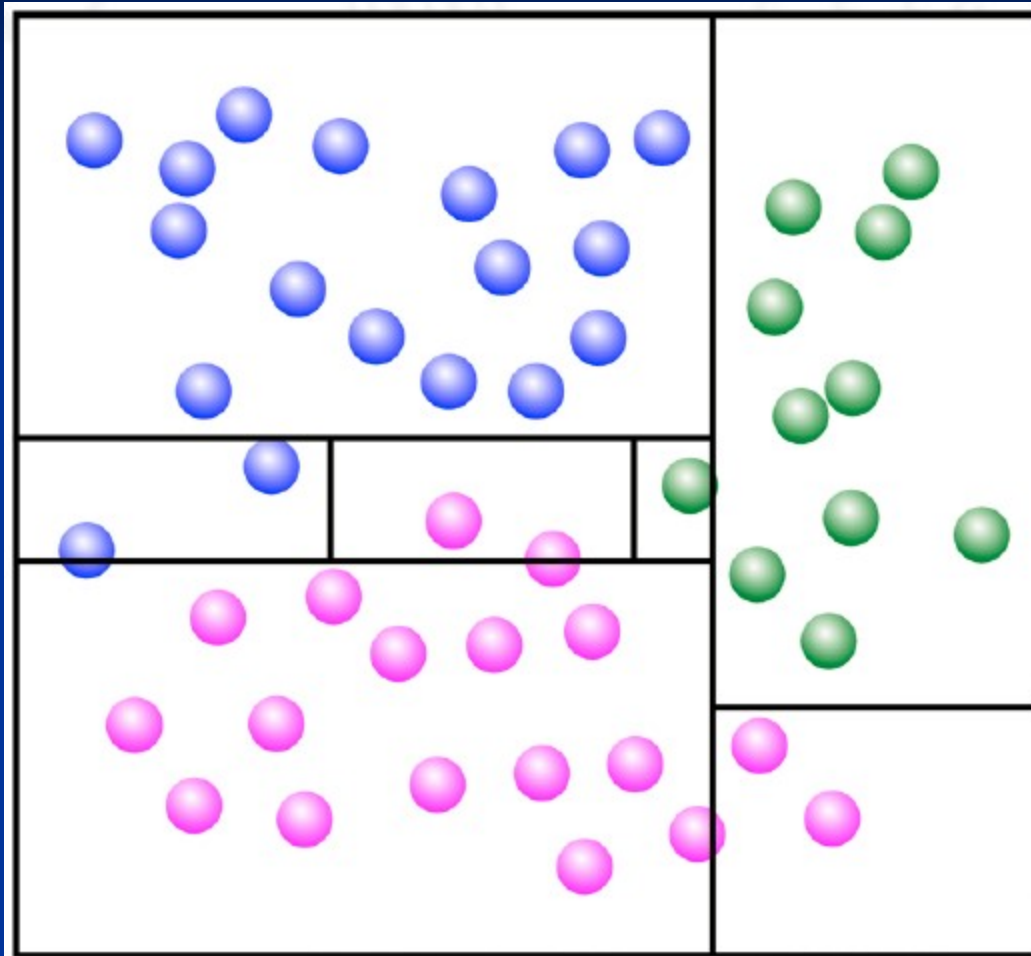
Drzewo decyzyjne (5)



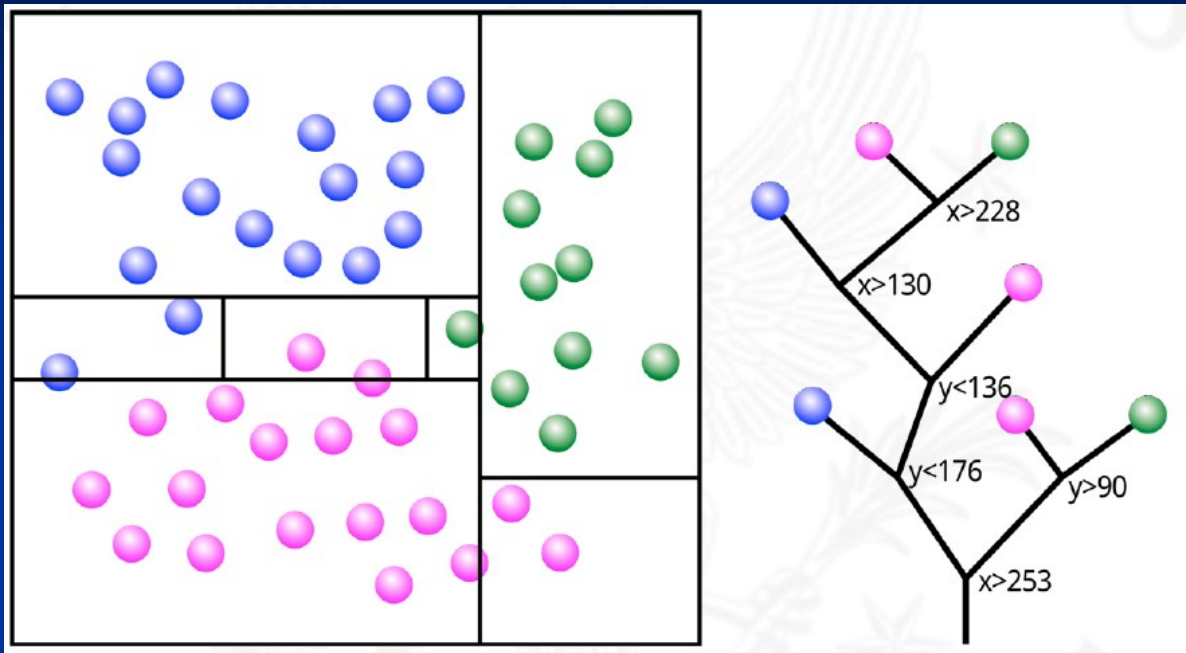
Drzewo decyzyjne (6)



Drzewo decyzyjne (7)



Drzewo decyzyjne – podsumowanie



Zalety:

- + proste
- + szybkie
- + elastyczne

Wady:

- przeczają się
- niestabilne

Las losowy (ang. *Random forest*)

- Zbudowany z **wielu klasyfikatorów** (drzew decyzyjnych) – stąd nazwa *Las*
- Każde drzewo budowane na **innym, losowo wybranym** podzbiorze zbioru danych uczących
- Każdy podział w drzewie jest wybierany jako najlepszy możliwy podział dla losowo wybranego małego podzbioru zmiennych (stąd nazwa *Losowy*).
- Błąd klasyfikacji szacowany na podstawie obiektów **nie włączonych do próby**
- Decyzja „**przez głosowanie**” – dla danego zestawu cech wygrywa klasa, do której prowadziło najwięcej drzew

Jedna ze skuteczniejszych metod klasyfikacji

Las losowy – same zalety 😊

- + prosty algorytm
- + mało wolnych parametrów:
 1. głębokość drzewa
 2. liczba zmiennych rozpatrywanych przy budowie drzewa
 3. liczba drzew
- + wewnętrzna nieobciążona ocena jakości modelu
 - Dla każdego obiektu węź przewidywania tylko tych drzew, które go nie widziały
- + ocena ważności zmiennych
 1. Dla każdej zmiennej, weź drzewa, które jej używały
 2. Dla każdego drzewa sprawdź jak bardzo wynik klasyfikacji zależy od tej zmiennej
 3. Uśrednij po wszystkich drzewach

