
Klasyfikacja sygnału i tła metodą wyboru obszaru granicznego

Cwiczenia zestaw 1

Termin oddania: 30.10 2009

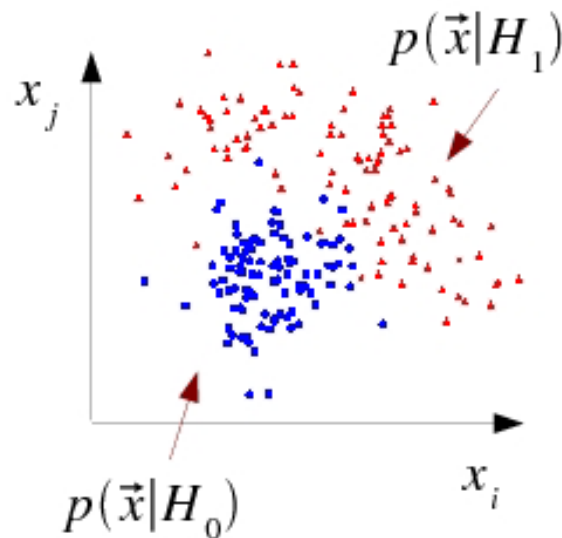
Problem klasyfikacji sygnału i tła

Przypuśćmy że chcemy dokonać klasyfikacji zdarzeń losowych w przestrzeni wielowymiarowej.

Zdarzenie losowe odpowiada wynikowi pomiaru

$$X = (x_1, x_2, \dots, x_n)$$

rozkład wg. pewnej funkcji prawdopodobieństwa



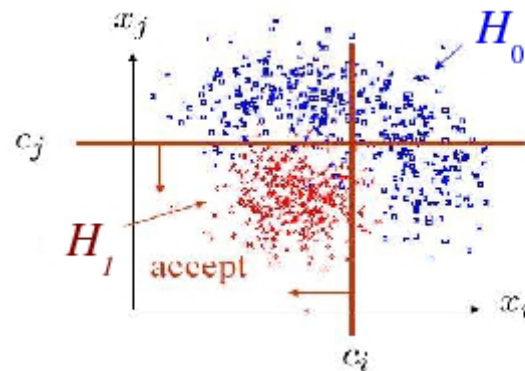
Potrafimy określić ten rozkład
Dla hipotezy klasy zdarzeń, np.
Sygnał (H_0) i tło (H_1)

Znajdowanie optymalnej decyzji

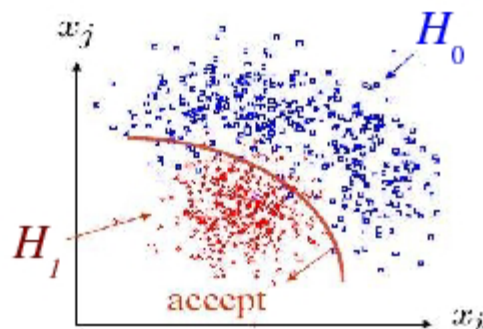
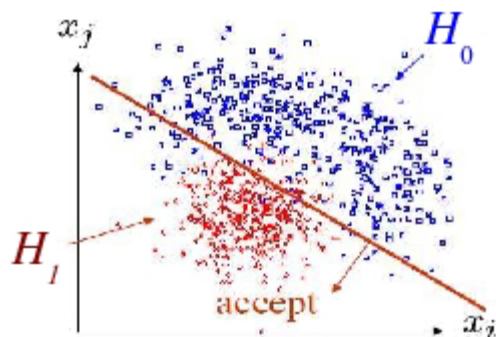
Może selekcja przy pomocy „cięć”

$$x_i < c_i$$

$$x_j < c_j$$



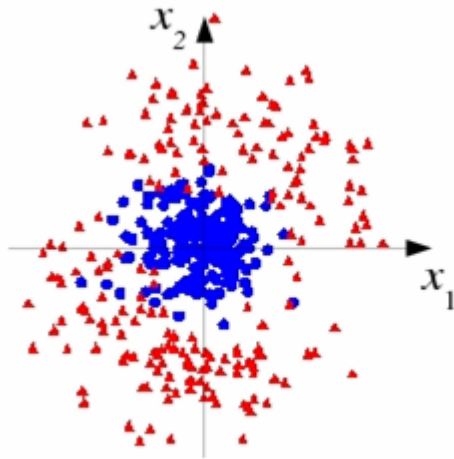
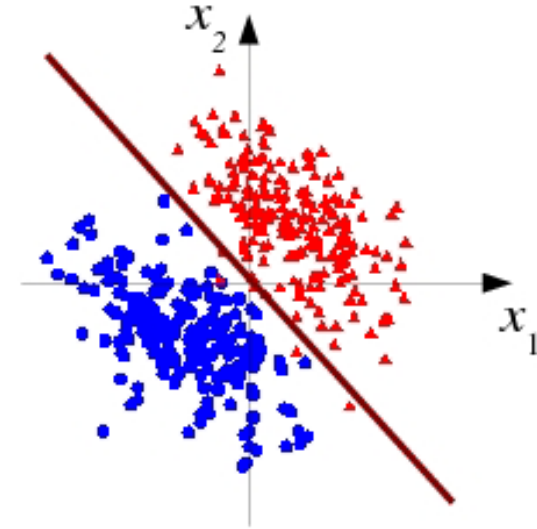
Może inny typ „obszaru granicznego”



Metody typu „multi-variate” zajmują się znajdowaniem optymalnej selekcji. Jedną z takich metod to sieci neuronowe.

Liniowe rozgraniczenie obszarów

Klasyfikacja przy pomocy liniowego rozgraniczenia jest optymalna tylko jeżeli obie klasy (sygnału i tła) mają charakterystykę rozkładu Gausa z tym samym rozmiarem i różną średnią.

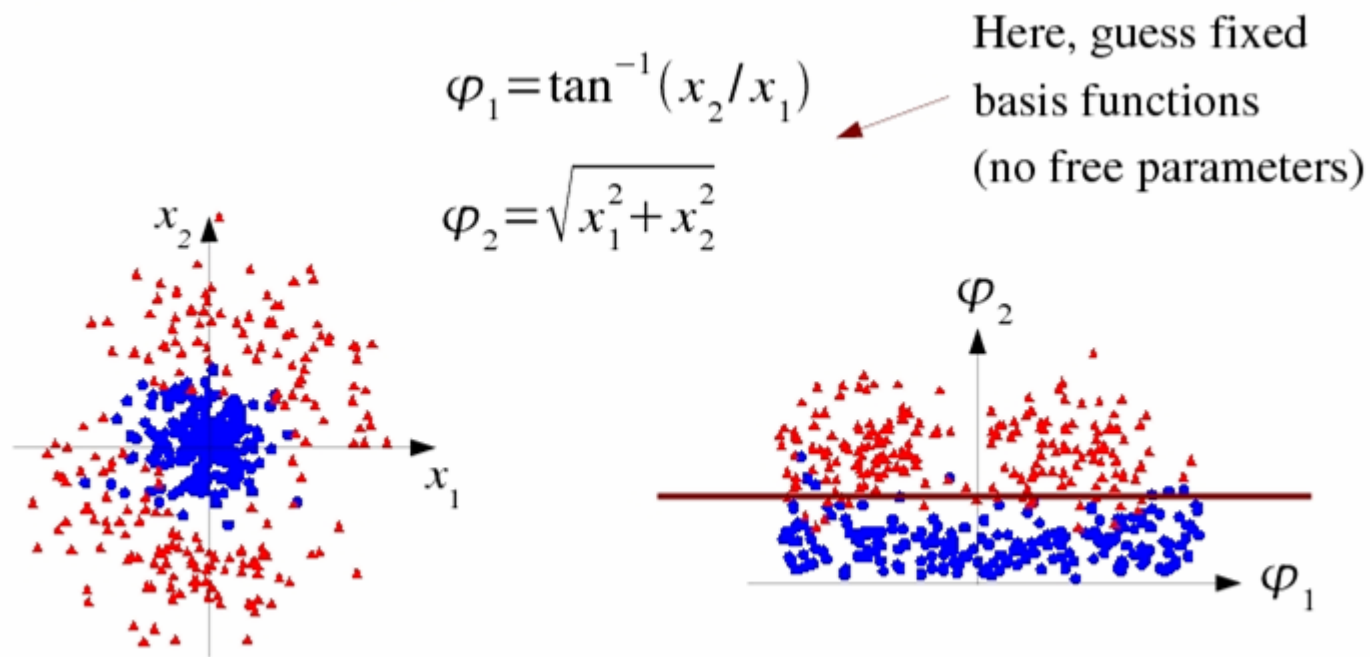


Multivariate Statistical Me

Dla pewnych klas problemów
Rozgraniczenie liniowe nie ma sensu.

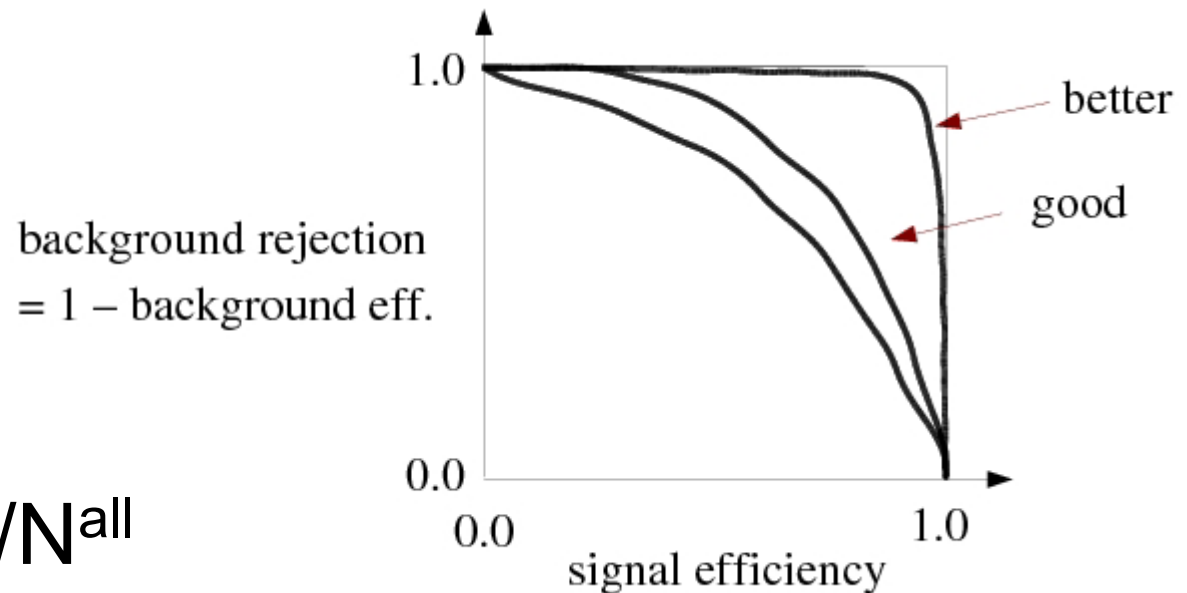
Nieliniowa transformacja

Możemy próbować znaleźć transformację zmiennych dla których separacja liniowa będzie miała sens.



Krzywe ROC

Możemy scharakteryzować jakość procedury klasyfikacji używając tzw. krzywej ROC (receiver operating characteristic).



$$\text{effic} = N^{\text{accept}} / N^{\text{all}}$$

Niezależne od prawdopodobieństwa rozkładu, powierzchnia pod krzywą może być miarą jakości procesu klasyfikacji.

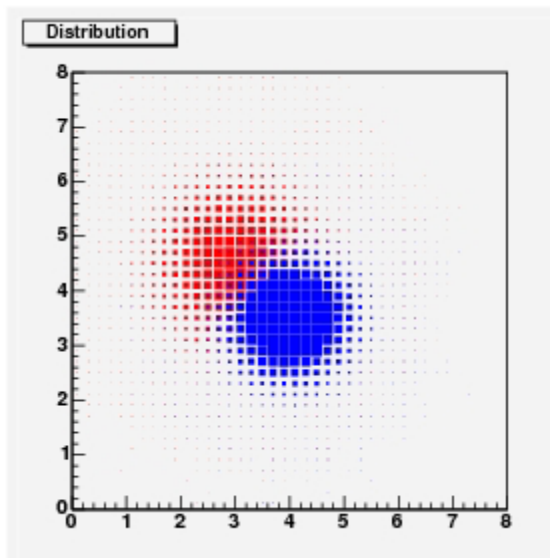
Zestaw zadań 1

- Wygenerować serie zdarzeń sygnału i tła
 - histogram 2D rozkładu gęstości
 - test poprawności generacji poprzez histogramowanie 1D oraz fit do histogramu, czy parametry fitu są zgodne z przyjętymi parametrami rozkładu zastosowanego do generacji serii.
 - Zapamiętać serię generowanych zdarzeń w formacie „rootupla”
- Dokonać klasyfikacji generowanej serii metodą „cieć” czyli wyboru obszaru granicznego sygnału i tła. Narysować histogram 2D rozkładu gęstości przed i po selekcji.
- Narysować krzywą ROC dla wybranego kształtu obszaru granicznego (systematyczne przeszukanie zakresu parametrów dla wybranych warunków cięcia).
- Zaliczenie:
 - ROOT macro dla każdego przykładu.
 - Krotki opis metody selekcji, uzyskanych wyników + rysunki (plik PDF lub PS).

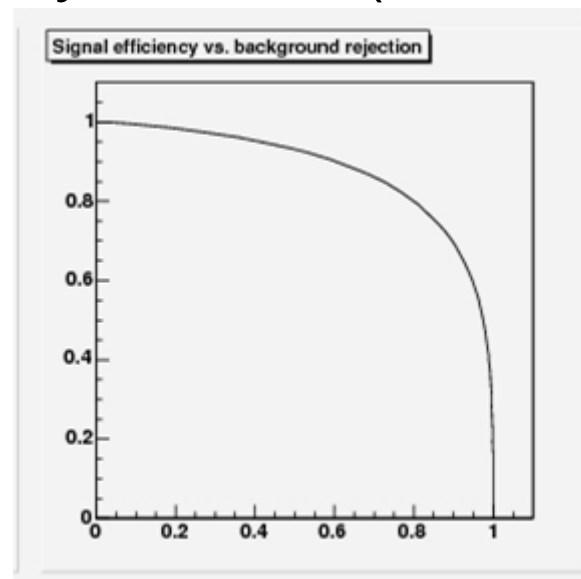
Problem 1

Zmienne nieskorelowane, sygnał i tło wg. rozkładu Gaussa dwu-wymiarowego o przesuniętych względem siebie środkach i różnej szerokości

Rozkład gęstości



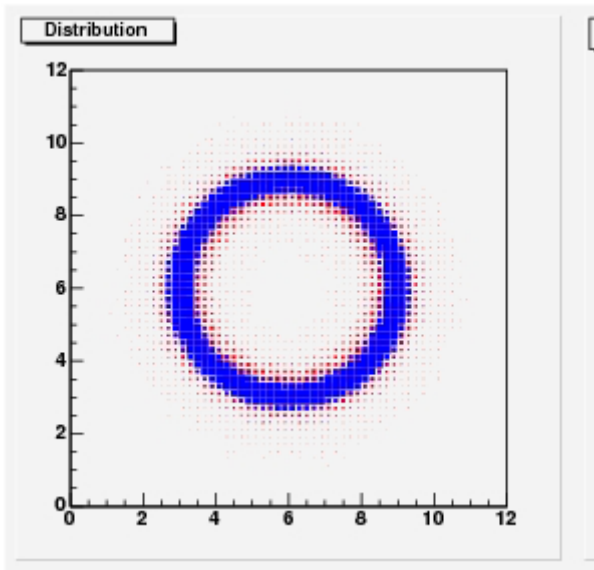
Krzywa ROC (metoda PDRS)



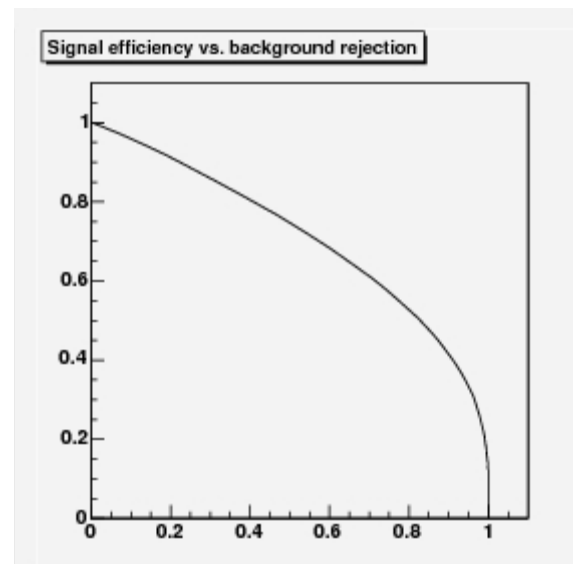
Problem 2

Zmienne silnie skorelowane, sygnał i tło wg. rozkładu na pierścieniu, o szerokości zgodnie z rozmyciem Gaussa

Rozkład gęstości



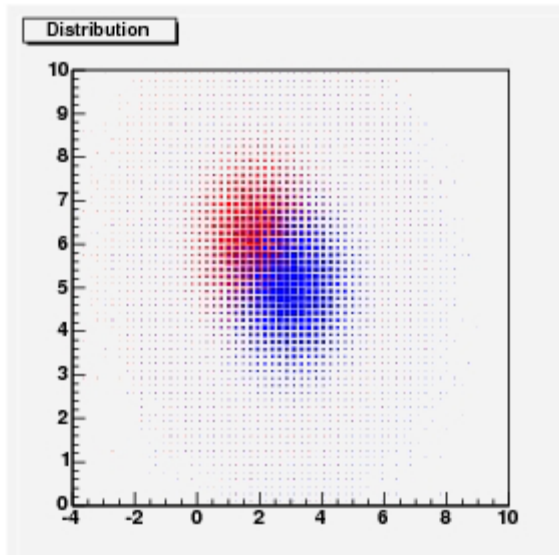
Krzywa ROC (metoda PDRS)



Problem 3

Zmienne słabo skorelowane, sygnał i tło w przestrzeni 5-cio wymiarowej. Generacja wektora X przy pomocy jednowymiarowych rozkładów Gaussa, wprowadzenie słabej korelacji poprzez macierz obrotu.

Rozkład gęstości (2-wym)



Krzywa ROC (metoda PDRS)

