Statystyczne metody przetwarzania danych

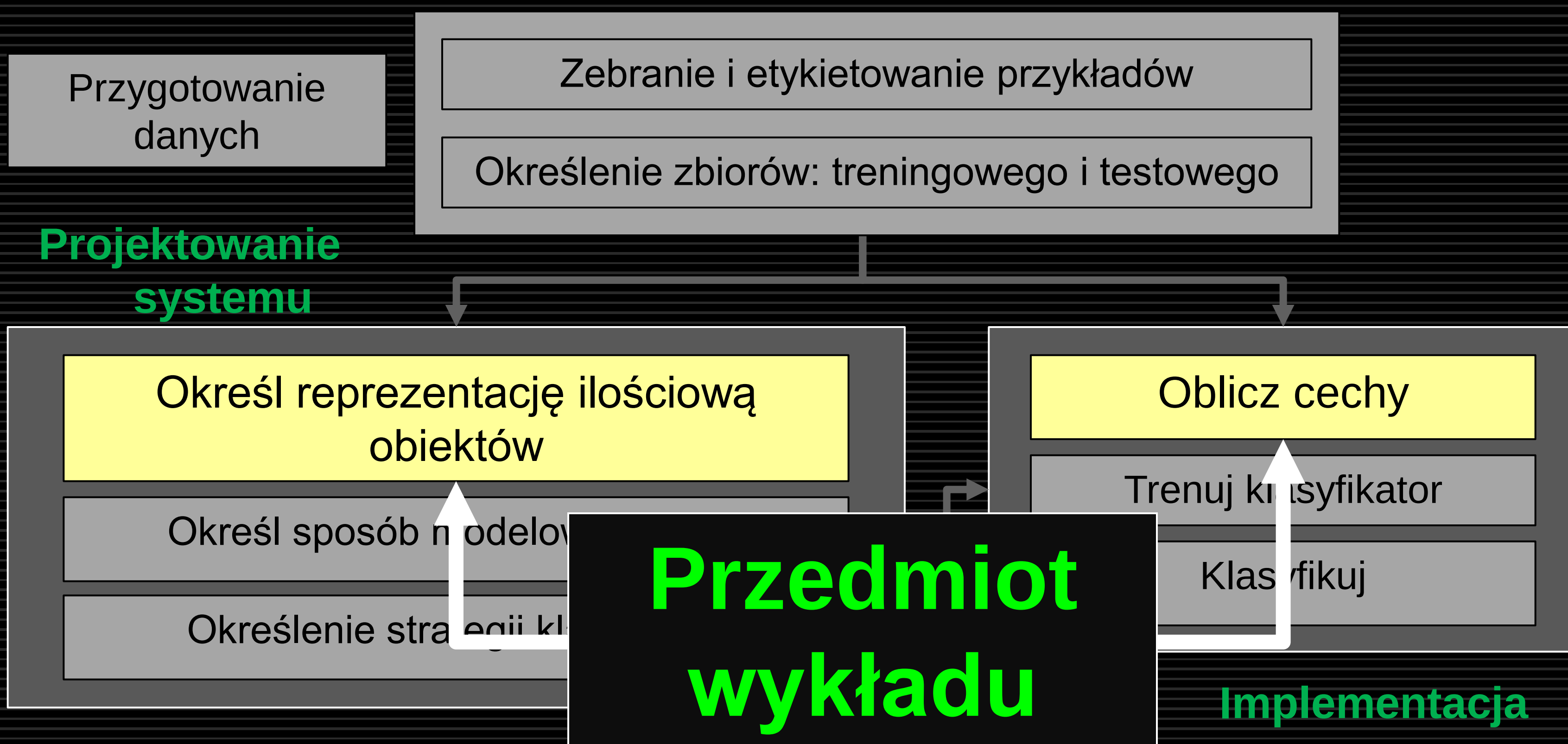
Cechy i przestrzenie cech

Krzysztof Ślot

Instytut Informatyki Stosowanej
Politechnika Łódzka

Wprowadzenie

- **Etapy procedury rozpoznawania**
  - Przetwarzanie wstępne, obejmuje ekstrakcję obiektu z tla
  - Projektowanie systemu rozpoznawania
  - Klasyfikacja, wykorzystująca wybrane cechy i zbudowane modele klas



Wprowadzenie

- **Unifikacja różnych typów zadań**
  - Analiza przynależności do klasy przeprowadzana w przestrzeniach cech

Akwizycja i przetwarzanie wstępne

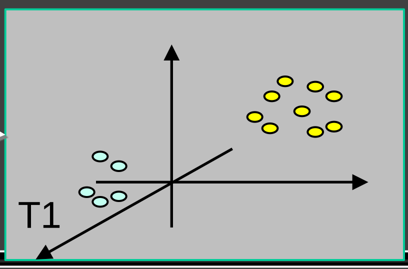
Cechy: ilościowa reprezentacja obiektów

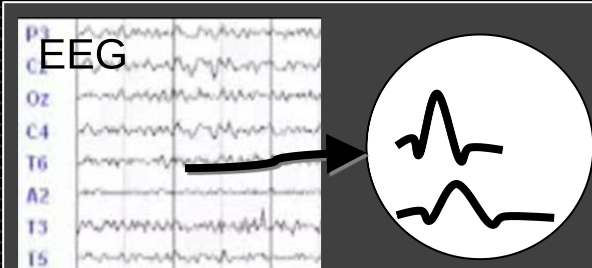
Reprezentacja w przestrzeni cech

$$\mathbf{F}_1^{\alpha} = \left[f_0^{\alpha}, f_1^{\alpha} \ldots f_{n-1}^{\alpha}\right]$$

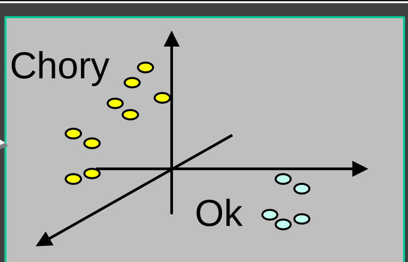
$$\mathbf{F}_2^{\alpha} = \left[f_0^{\alpha}, f_1^{\alpha} \ldots f_{n-1}^{\alpha}\right]$$

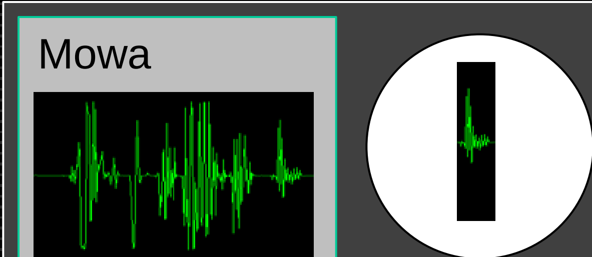
$$\mathbf{F}^{\alpha} = \left[f_0^{\alpha}, f_1^{\alpha} \ldots f_{n-1}^{\alpha}\right]$$



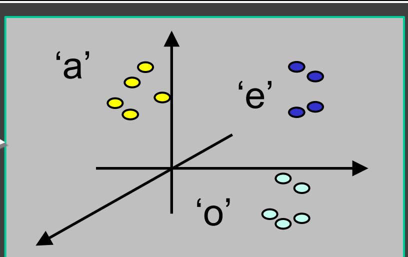


$$\mathbf{F}^{\beta} = \left[f_0^{\beta}, f_1^{\beta} \ldots f_{m-1}^{\beta}\right]$$

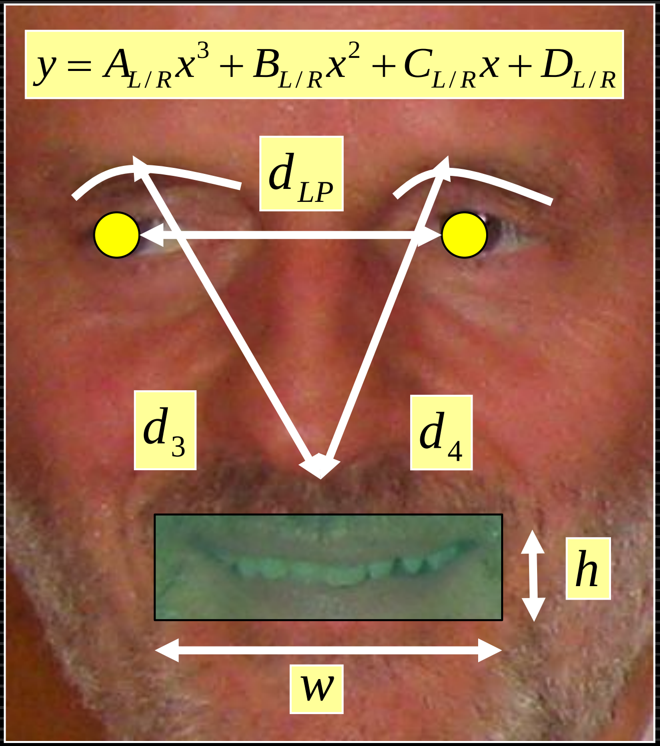




$$\mathbf{F}^{\gamma} = \left[f_0^{\gamma}, f_1^{\gamma} \ldots f_{o-1}^{\gamma}\right]$$



Cechy i przestrzenie cech



- **Definicje**
  - Cecha: charakterystyka ilościowa
  - Przestrzeń cech: przestrzeń metryczna wyznaczona przez cechy

- **Przestrzenie cech dla rozpoznawania**
  - Powinny być ‘najlepsze’ (‘**dyskryminatywne**’)
  - Typowo brak wskazówek wyboru cech (które, ile)

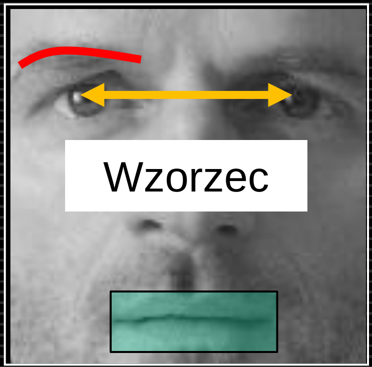
Wektor cech $[d_{LP}, A_L, B_L, C_L, D_L, A_R, B_R, C_R, D_R, w, h, d_1, d_2, d_{OBL}, d_{OBP} \ldots]$

Staje się coraz dłuższy … np. kilkaset tysięcy elementów (Eigenfaces)

- **Wymiar wektora – wymiar przestrzeni cech**
  - Wielowymiarowe przestrzenie cech

Opis obiektów = przestrzeń wielkowymiarowa?

- ~~**Hipoteza**~~
  - ~~Użyć wszystkich możliwych do wyznaczenia cech (ponieważ nie wiadomo, które są ważne)~~



$$\begin{bmatrix} d \\ A_L \\ B_l \\ \vdots \end{bmatrix}$$

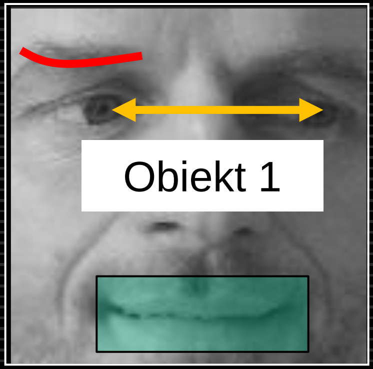
$$d : 12.35 \approx 12.2 \approx 12.4 \approx \cdots$$

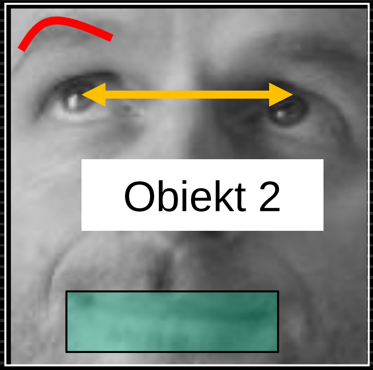
$$A_L : 2.3 \neq -1.9 \neq 4.3 \cdots$$

$$\vdots$$

$$w : 6 \neq 8.6 \neq 8.1 \cdots$$

$$\vdots$$





– Brak zgodności obiektów z wzorcem dla różnych cech

Podobieństwo

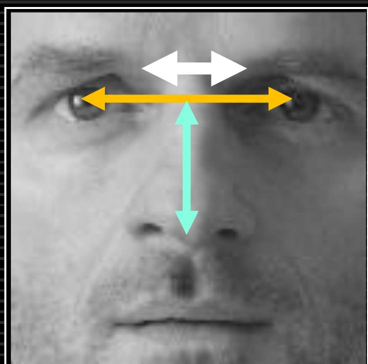
$$D = \sqrt{(d_w - d_x)^2 + (A_{Lw} - A_{Lx})^2 + \cdots}$$

– „Zła” cecha to ‘szum’ obliczania podobieństwa: losowe wartości miary podobieństwa (D)

99% fiasko

Opis obiektów = przestrzeń wielkowymiarowa?

- ~~**Hipoteza**~~
  - ~~Użyć wszystkich dobrych cech (np. 100)~~



$$\begin{bmatrix} d/d_N \\ d/d_{K\_LP} \\ \vdots \end{bmatrix}$$

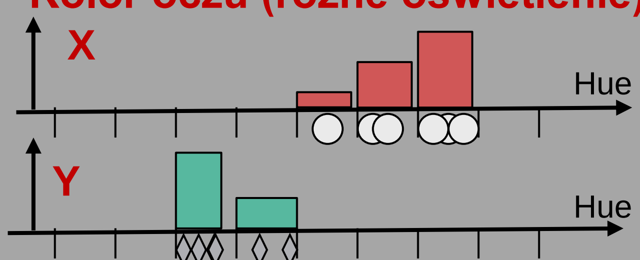
$d/d_N : 1.35 \approx 1.29 \approx 1.34 \approx \cdots$

$d/d_{K\_LP} : 3.2 \approx 3.4 \approx 3.7 \approx \cdots$

$\vdots$

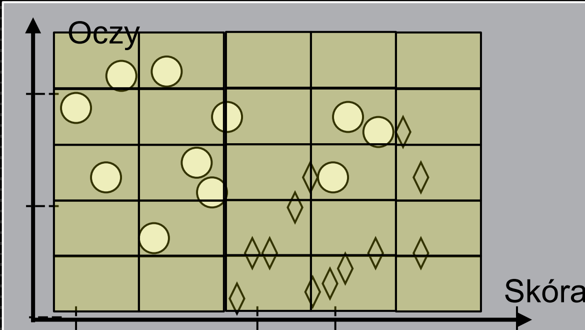
– Wyniki obliczania podobieństwa jest wiarygodny, ale …

Kolor oczu (różne oświetlenie)



'Modele' biometryczne 1D osób

(100 próbek/klasę, 10 przedz. -> 10p/p)



'Modele' 2D (100 p/k,10 prz.: 1p/p)

3D : : 0.1pr./komórkę
4D : : 0.01pr./komórkę

Złe modele

Klątwa wymiarowości

Redukcja wymiarowości problemu rozpoznawania

- **Wymiar przestrzeni cech powinien być jak najmniejszy**
  - Dostępny zbiór cech będzie typowo nadmiarowy, będzie zawierał cechy dobre, złe i średnie
  - Konieczne będzie dokonanie redukcji tego zbioru w oparciu o odpowiednio sformułowane kryteria

- **Redukcja wymiarowości**
  - Poszukiwanie niskowymiarowej reprezentacji obiektów

$$\Re^D \to \Re^d, \qquad d \ll D$$

$$\mathbf{F} = \left[f_0, f_1 \dots f_{D-1}\right] \to \boldsymbol{\xi} = \left[\xi_0, \xi_1 \dots \xi_{d-1}\right]$$

- **Strategie redukcji wymiarowości**
  - **Selekcja cech**
  - **Ekstrakcja cech**

Selekcja cech

- **Sformułowanie problemu**
  – Znajdź podzbiór oryginalnego, nadmiarowego zbioru cech, zapewniający największą skuteczność rozpoznawania (cechy najlepsze)

- **Zagadnienia cząstkowe**
  1. Ocena jakości zbioru cech
  2. Sposób generacji podzbiorów do sprawdzenia

1. Kryteria oceny jakości podzbioru
  – Ocena skuteczności klasyfikacji
  – Prognozowanie skuteczności (miary statystyczne)

- **Ocena skuteczności klasyfikacji**
  – Narzucające się kryterium
  – Ocena pośrednia: wikła ocenę jakości cech i jakości przyjętej metody klasyfikacji, jest złożona obliczeniowo (dla każdego zbioru konieczne tworzenie osobnego systemu klasyfikacji i ocena wyników jego działania)

Statystyczne kryteria selekcji cech

- **Właściwości**
  - Zalety: prostota koncepcyjna i efektywność obliczeniowa
  - Wady: używanie założeń ograniczających ogólność rozważań

- **Wskazówki wyboru przestrzeni cech**
  - Dobra przestrzeń zapewnia dobrą separację klas (próbek klas)
  - Dobra separacja: ‘chmury’ próbek danej klasy znajdują się w odrębnych (możliwych do (łatwego) rozdzielenia) regionach przestrzeni cech

- **Podstawowe statystyki rozkładów 1D**
  - Średnia (arytmetyczna, geometryczna)
  - Wariancja
  - Inne momenty zwykłe
  - Inne momenty centralne
  - ...

$$E(X)=\sum_{i=0}^{n-1} p(x_i)x_i$$

$$\mu=\frac{1}{n}\sum_{i=0}^{n-1} x_i \qquad \mu_G=\frac{1}{n}\prod_{i=0}^{n-1} x_i$$

$$\sigma^2=E(X-\mu)^2=E(X)^2-\mu^2$$

$$s=E(X-\mu)^3$$

Przykład: oblicz podstawowe statystyki zbioru

$$A=\{1,2,3,4,5\}$$

Selekcja cech w przestrzeni 1D

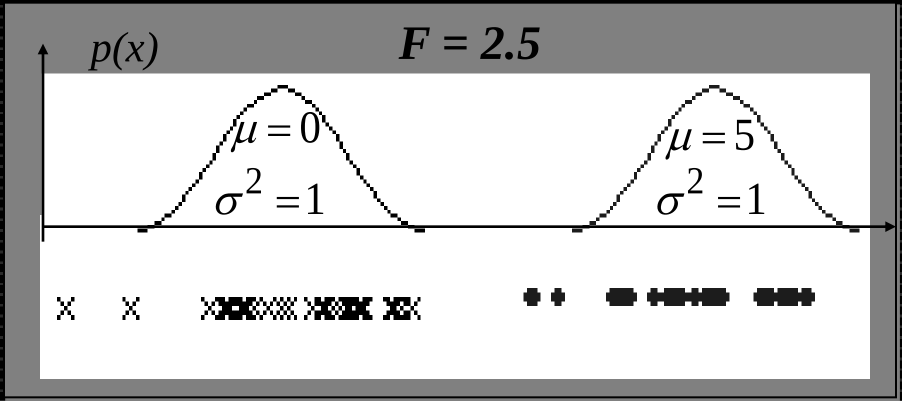
- **Założenia**
  - Obiekty są opisane za pomocą jednej cechy, rozważamy 2 klasy
  - Celem jest ocena 'separowalności' (możliwości separacji) klas

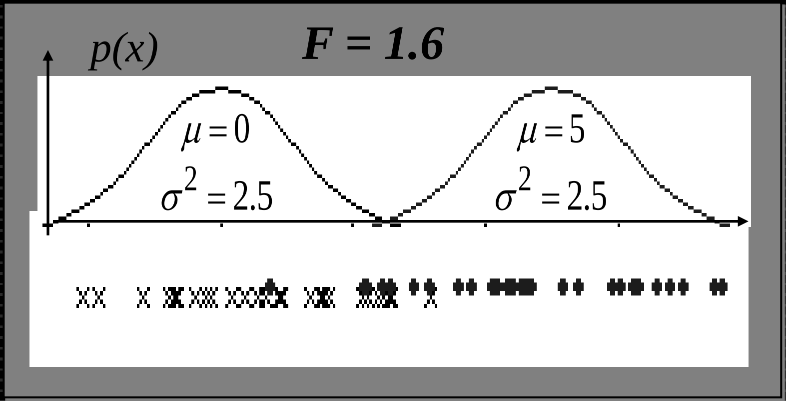
- **Kryterium statystyczne**
  - Dobra separacja: wartości średnie klas leżą daleko od siebie, wariancje w obrębie klas są jak najmniejsze

Współczynnik dyskryminacji liniowej Fishera:: 1D, 2 klasy

S_b – rozrzut między-klasowy, S_w rozrzut wewnątrz-klasowy

$$F = \frac{S_b}{S_w} = \frac{|m_A - m_B|}{\sigma_A + \sigma_B}$$





Przykład

$\mathbf{x}^A$ = { (-1, 0}, (1, 1), {0, 0)}, $\mathbf{x}^B$ = { (2, -1}, (0, 0), {-2, -2)}

Selekcja cech w przestrzeniach n-wymiarowych

- **Statystyczna reprezentacja danych w n-wymiarach**
  – Wartość średnia: wektor
  – Zmienność: macierz kowariancji

Przykłady klasy A

$$A = \left\{ \begin{bmatrix} x_0^0 \\ \vdots \\ x_{n-1}^0 \end{bmatrix}, \begin{bmatrix} x_0^1 \\ \vdots \\ x_{n-1}^1 \end{bmatrix}, \dots \begin{bmatrix} x_0^m \\ \vdots \\ x_{n-1}^m \end{bmatrix} \right\}$$

Średnia

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} E(x_0) \\ \vdots \\ E(x_{n-1}) \end{bmatrix}$$

Kowariancja

$$Cov(X) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = \sum_{i=0}^{n-1} p(\mathbf{x}_i)(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

Macierz kowariancji

$$Cov(X) = \begin{bmatrix} \sigma_{00}^2 = c_{00} & c_{01} & \dots \\ c_{10} & \sigma_{11}^2 = c_{11} & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

Współczynnik Fishera dla n-D

- **Założenia: 2 klasy, 2D**
  - Jak określić odległość między klasami (dla 1D: moduł różnicy średnich)?
  - Jak opisać zmienność w obrębie klasy(1D: suma odchyleń)?

- **Rozrzut międzyklasowy dla 2 klas w przestrzeni 2D**
  - Odległość między wartościami średnimi (długość wektora)

$$S_b = \| \boldsymbol{\mu}^A - \boldsymbol{\mu}^B \|$$

- **Rozrzut wewnątrzklasowy dla 2 klas, 2D**

$$S_w = \det(C^c) = \det\begin{pmatrix} \sigma_0^2 & \dots & c_{0n} \\ : & & : \\ c_{n0} & \dots & \sigma_n^2 \end{pmatrix}$$

Wyznacznik macierzy kowariancji

$$\det\begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \end{pmatrix} = a_0 b_1 - b_0 a_1 = \mathbf{a} \times \mathbf{b} = ab\sin(\mathbf{a},\mathbf{b})$$

Objętość równoległościanu wyznaczonego przez wektory (kolumny) macierzy kowariancji

$\mathbf{x}^A$ = { (-1, 0, -1}, (1, 1, 1), {0, 0, 0)}

Współczynnik Fishera dla n-D

- **Przypadek 2D, 2 klasy**
  - Kryterium separowalności (Fishera) – analogiczna zasada: relacja między rozrzutami: międzyklasowym i wewnątrzklasowym

Macierz kowariancji

$$C^c = E\left[\left(\mathbf{x}^c - \boldsymbol{\mu}^c\right)\left(\mathbf{x}^c - \boldsymbol{\mu}^c\right)^T\right]$$

Macierz rozrzutu

$$\mathbf{S}^c = \sum_{i=1}^{N}\left(\mathbf{X}_i^c - \boldsymbol{\mu}^c\right)\left(\mathbf{X}_i^c - \boldsymbol{\mu}^c\right)^T$$

- **Alternatywne formy współczynnika Fishera**

Dokładna estymacja rozrzutu

$$F = \frac{S_m}{S_w} = \frac{||\boldsymbol{\mu}^A - \boldsymbol{\mu}^B||}{\det(\mathbf{S}^A) + \det(\mathbf{S}^B)}$$

Zgrubna estymacja

$$F = \frac{S_m}{S_w} = \frac{||\boldsymbol{\mu}^A - \boldsymbol{\mu}^B||}{Tr(\mathbf{S}^A) + Tr(\mathbf{S}^B)}$$

- **Wyrażenia są ogólne i dotyczą zmiennych n-wymiarowych**

Współczynnik Fishera dla n-D

- **Przypadek najogólniejszy: wiele klas, n-wymiarów**
  - Idea kryterium bez zmian: relacja rozrzutu międzyklasowego do wewnątrz-klasowego

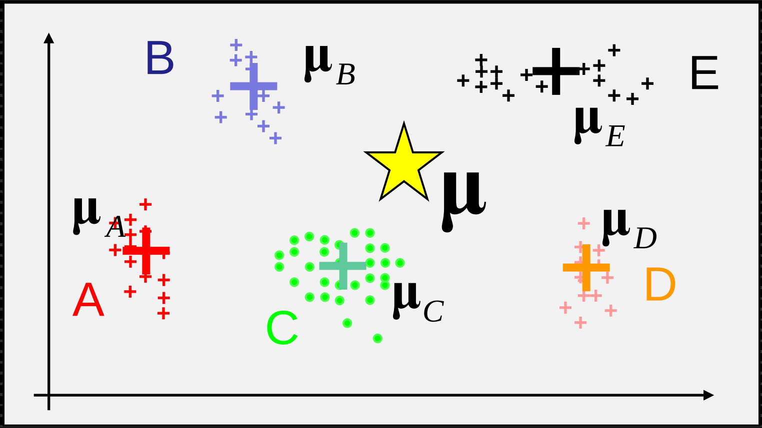
- **Rozrzut wewnątrzklasowy**
  - Miara identyczna jak poprzednio – suma wyznaczników lub śladów indywidualnych macierzy rozrzutu

$$S_w = \sum_{i=1}^{C} \det(\mathbf{S}^i) \qquad \text{lub} \qquad S_w = \sum_{i=1}^{C} Tr(\mathbf{S}^i)$$

- **Rozrzut międzyklasowy**
  - Konieczna redefinicja (mamy do czynienia z wieloma wartościami średnimi)

Współczynnik Fishera

- **Rozrzut międzyklasowy dla wielu klas i przestrzeni n-D**
  - Idea: potraktowanie wartości średnich klas jako próbek nowego, hipotetycznego rozkładu i określenie rozrzutu wewnątrzklasowego tego rozkładu



Rozrzut międzyklasowy

$$\tilde{\mathbf{S}}_B = \sum_{i=1}^{C} \left(\boldsymbol{\mu}^i - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}^i - \boldsymbol{\mu}\right)^T$$

$\boldsymbol{\mu}^i$ Średnia klasy ‘i’

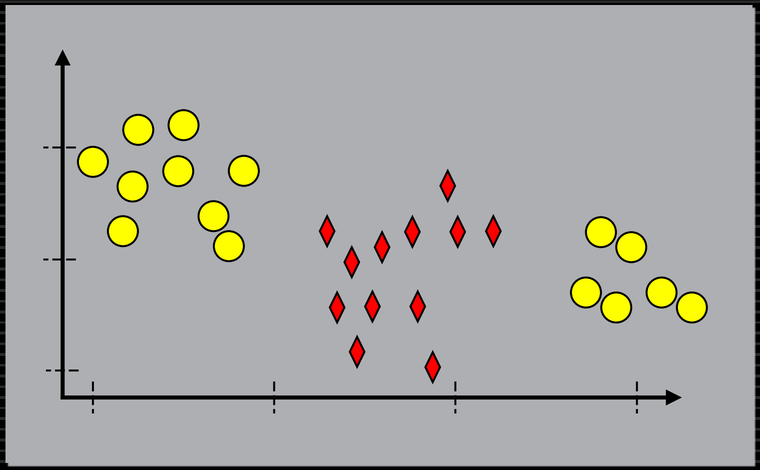
$\boldsymbol{\mu}$ Średnia rozkładu

Współczynnik Fishera C-klas, n-wymiarów

$$F = \frac{S_m}{S_w} = \frac{\det(\tilde{\mathbf{S}}_B)}{\sum_{i=1}^{C} \det(\mathbf{S}^i)} \quad , \quad F = \frac{S_m}{S_w} = \frac{Tr(\tilde{\mathbf{S}}_B)}{\sum_{i=1}^{C} Tr(\mathbf{S}^i)}$$

Selekcja cech z użyciem wsp. Fishera

- **Właściwości**
  - Prosta i szybka procedura
  - Zakłada, że rozkłady próbek klas mają charakter Gaussowski



Klasy separowalne nieliniowo

Wartość współczynnika Fishera sugeruje brak możliwości separacji klas

Selekcja cech

- **Zasadnicze zagadnienia problematyki selekcji cech**
  - Wybór optymalnej przestrzeni cech (omówiony)
  - Określenie systematycznej metody tworzenia podzbiorów cech

- **Generacja podzbiorów zbioru cech**
  - Selekcja wymaga sprawdzenia każdego podzbioru
  - Generacja wszystkich podzbiorów: zadanie o wykładniczej złożoności obliczeniowej

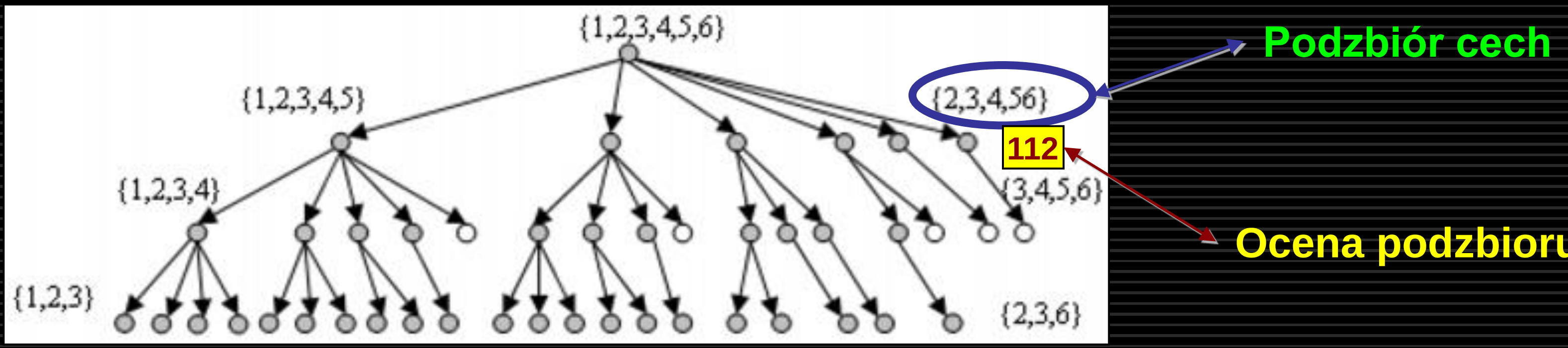
- **Przykład**
  - Liczność zbioru cech: 100
  - Cel: wybór optymalnej przestrzeni cech, co najwyżej 10-cio wymiarowej
  - Liczba wszystkich podzbiorów:

$$\binom{100}{1} + \binom{100}{2} + \ldots + \binom{100}{10}$$

$$\binom{100}{10} = \frac{100 \cdot 99 \cdot 98 \cdot \ldots 91}{1 \cdot 2 \cdot \ldots \cdot 10} \approx 17 \cdot 10^{12}$$

Algorytmy generacji podzbiorów

- **Metoda B-B**
  - Budowa drzewa poszukiwań
  - „Przycinanie” drzewa – usuwanie węzłów z pewnością prowadzących do gorszych podzbiorów

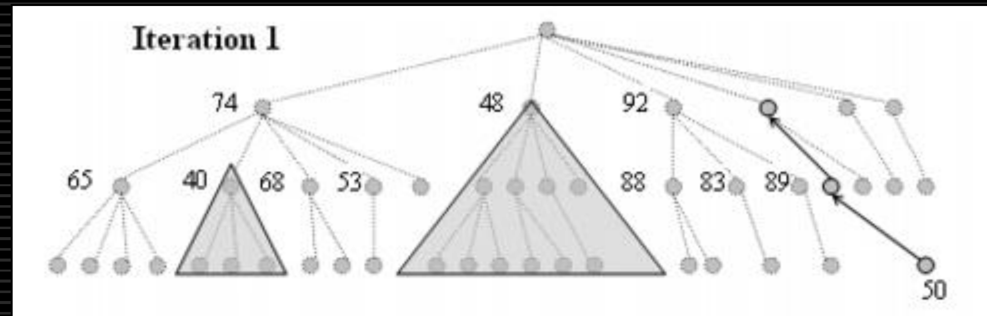


1. Sprawdzenie dowolnego podzbioru o założonej liczności S_k
2. ‘Wspięcie się’ do wierzchołka z rejestracją kosztu dla kolejnych warstw
3. Ekspansja w dół: jeżeli koszt dla podzbioru większy niż dowolny dotąd zarejestrowany, usuń dalsze poddrzewo,

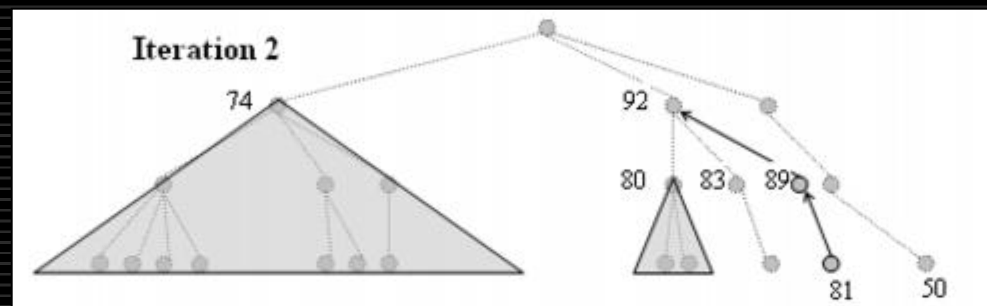
Metoda B-B generacji podzbiorów

- **Przykład**

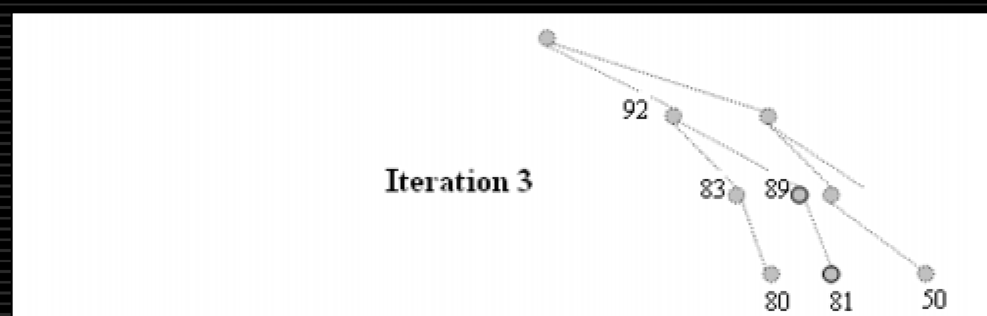
1. Jakość podzbioru startowego o docelowej liczności = 50



2. Ekspansja korzenia do poziomu, w którym znaleziono gorszy podzbiór (48), usunięcie poddrzewa (o ile nie przekroczono ustalonej głębokości



3. Wybór innego podzbioru o docelowej liczebności



4. Iteracyjne powtarzanie procedury

Metoda selekcji przyrostowej

- **Założenie**
  - Monotoniczność funkcji celu (liczniejszy podzbiór ma lepsze właściwości)

- **Selekcja przyrostowa (Sequential forward selection - SFS)**
  - Iteracyjne powiększanie najlepszego podzbioru o liczności n o kolejny element, dający najlepszy podzbiór n+1 - elementowy

$t=0:\mathbf{F}^0=\{\}$

$\forall i \in D : \mathbf{F}' = \{\mathbf{F}^t, f_i^{t+1}\}$

$\mathbf{F}^{t+1} = \mathbf{F}^t \cup f_i^{t+1}, i : FLD(\mathbf{F}^t \cup f_i^{t+1}) \rightarrow \max$

Koniec ?

STOP

$t=t+1$

- **Selekcja redukcyjna (Sequential Backwards selection - SBS)**
  - Iteracyjna redukcja zbioru o liczności n o kolejny element, dający najlepszy podzbiór n-1 - elementowy

- **Złożoność obliczeniowa**

$n+(n-1)+...(n-k+1) < k \cdot n$

$<< \frac{n!}{(n-k)!k!}$