...

Robert Susmaga
Instytut Informatyki
ul. Piotrowo 2
Poznań

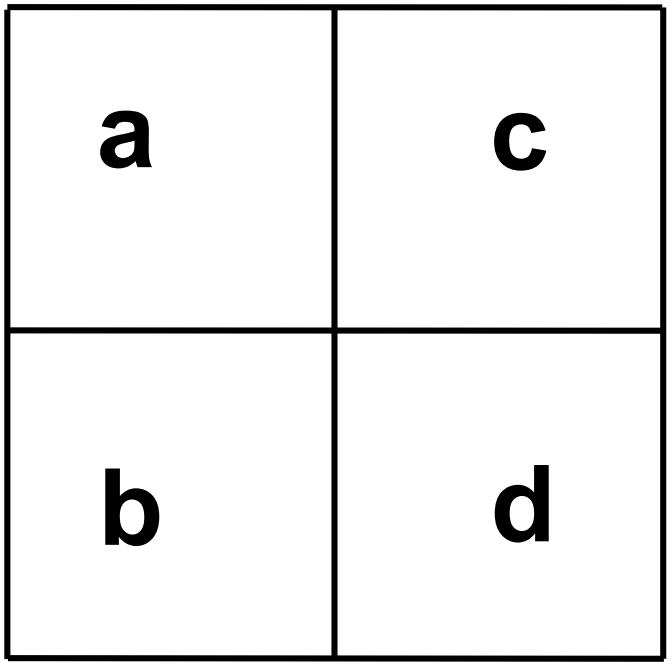
kontakt mail'owy
Robert.Susmaga@CS.PUT.Poznan.PL
kontakt osobisty
Centrum Wykładowe, „blok informatyki", pok. 7

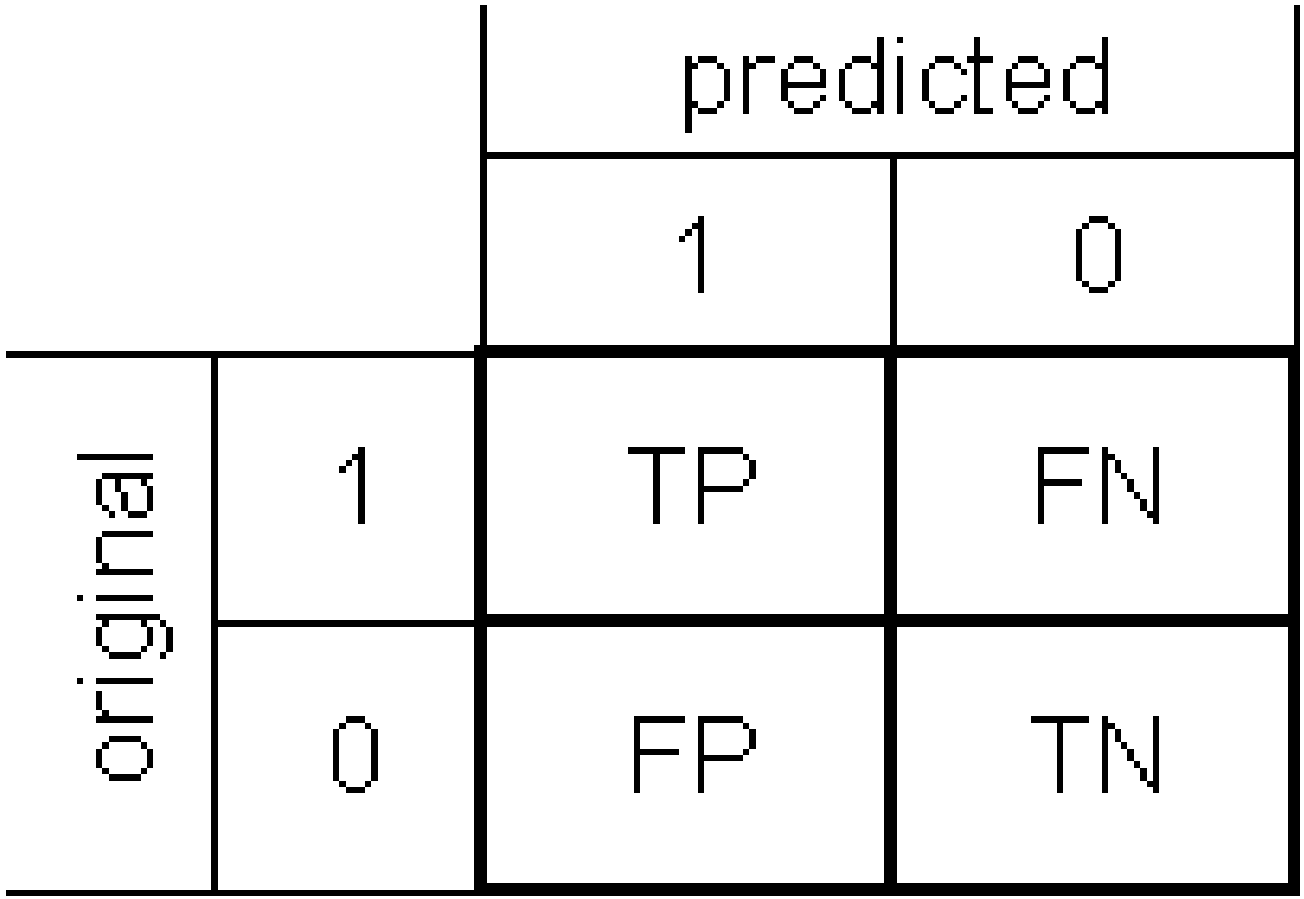
Wyłączenie odpowiedzialności

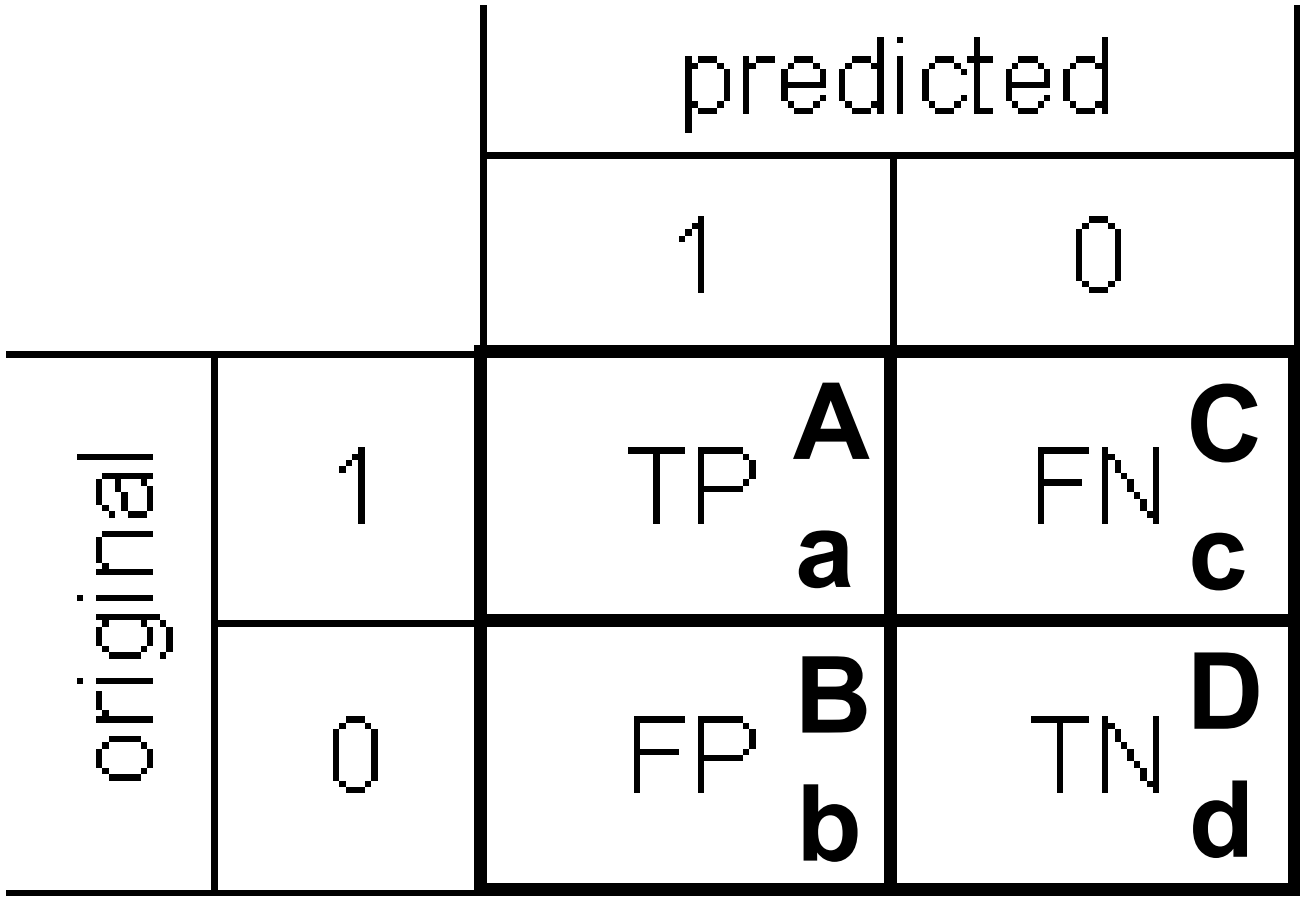
Prezentowane materiały, będące dodatkiem pomocniczym do wykładów, z konieczności fragmentarycznym i niedopracowanym, należy wykorzystywać z pełną świadomością faktu, że mogą nie być pozbawione przypadkowych błędów, braków, wypaczeń i przeinaczeń :-)

Autor

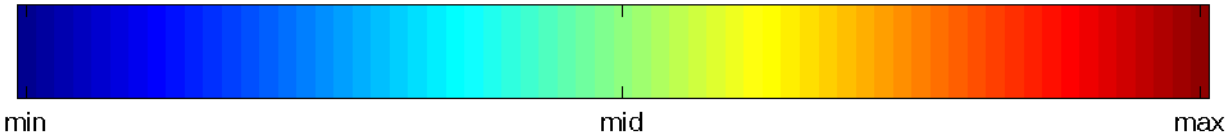
...

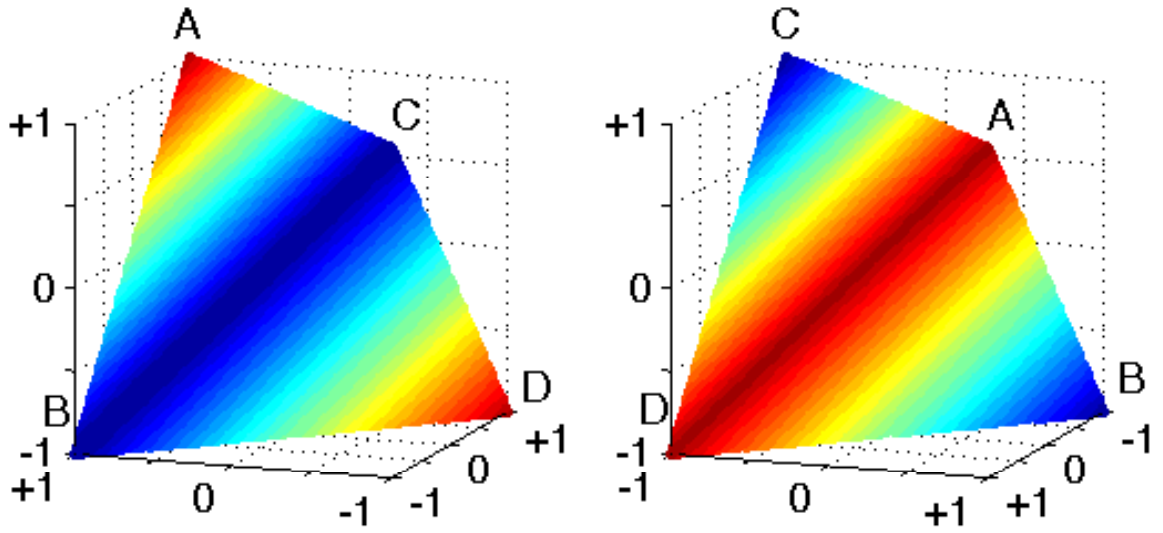

a
c
b
d


predicted
1
0
original
1
0
TP
FN
FP
TN

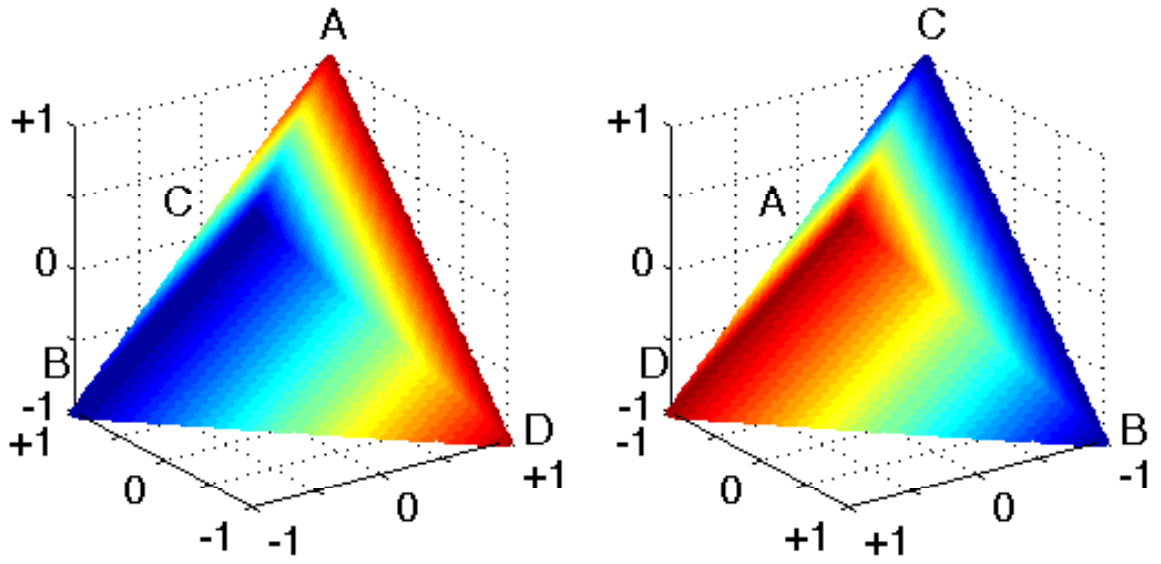

predicted
1
0
original
1
0
TP
A
a
FN
C
c
FP
B
b
TN
D
d

- Zakres wartości [min, max] = [0, 1]
  - (+ ewentualne NaN’y /fioletowe/)

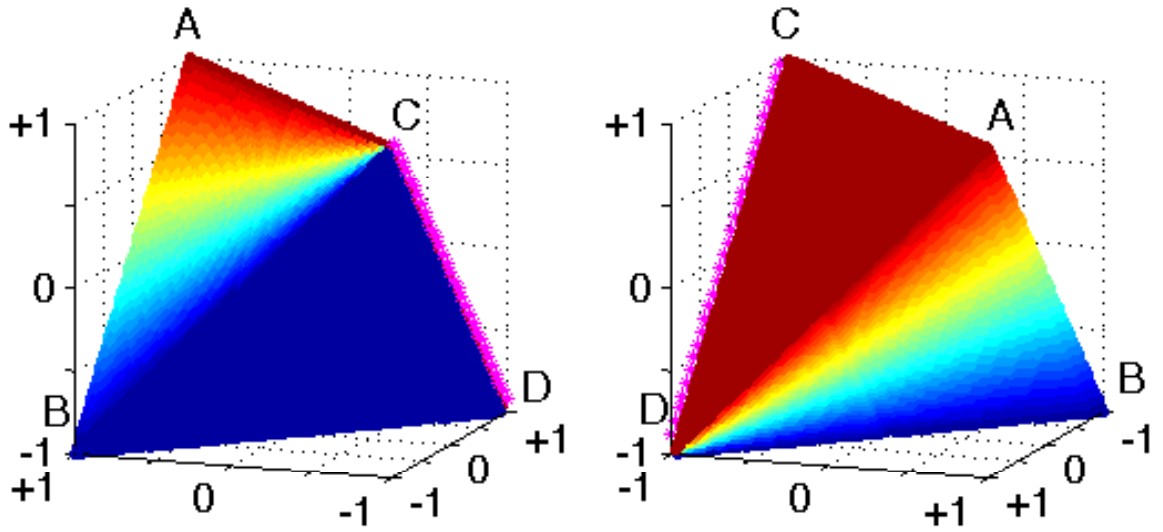

min
mid
max



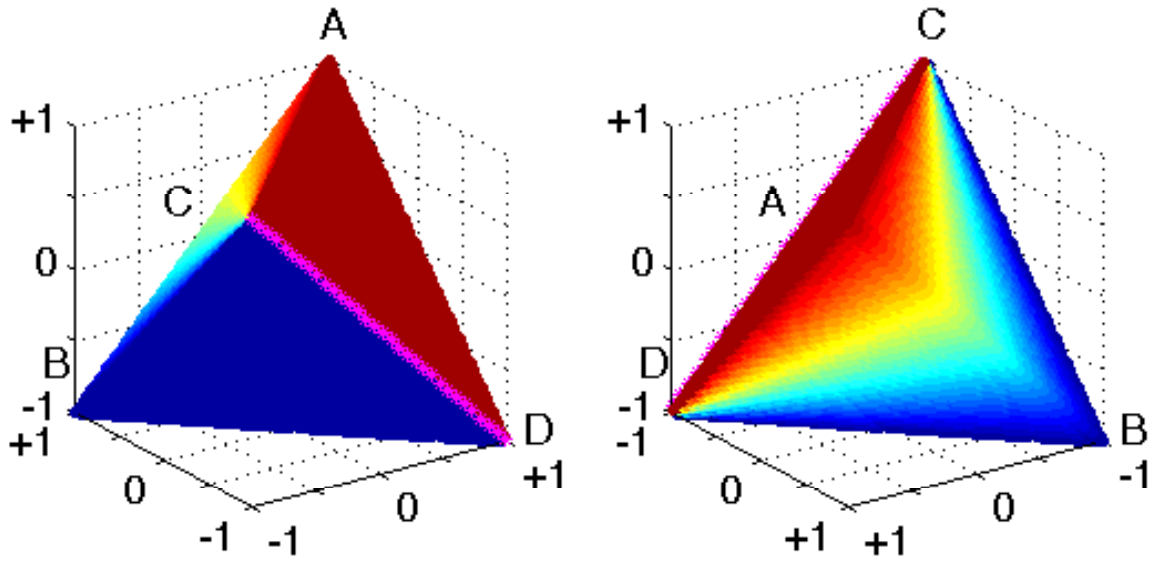
f(a,b,c,d) = (a+d)/(a+b+c+d) (accuracy)



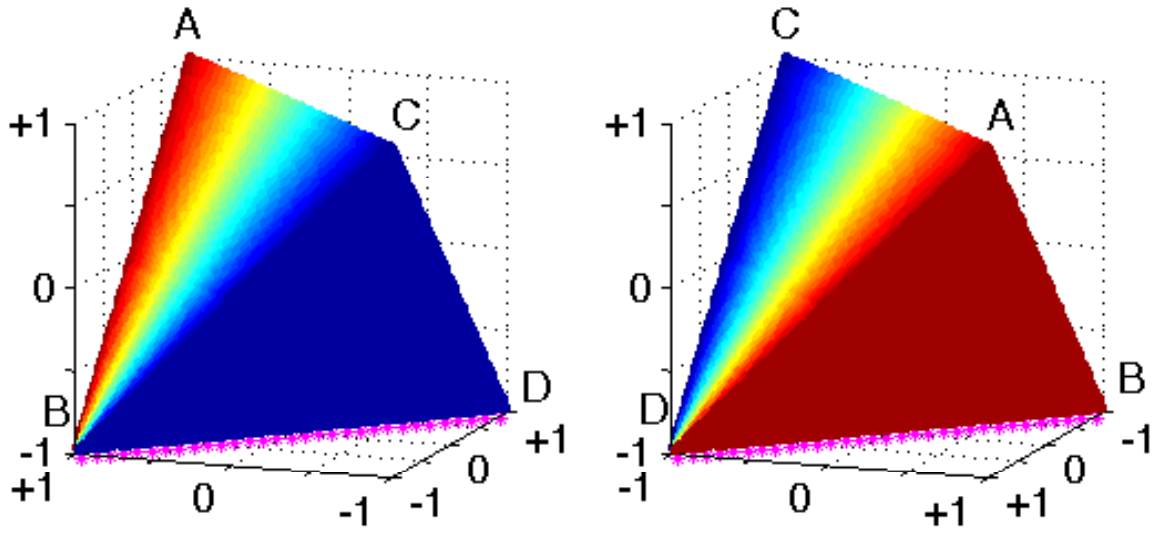
f(a,b,c,d) = (a+d)/(a+b+c+d) (accuracy)



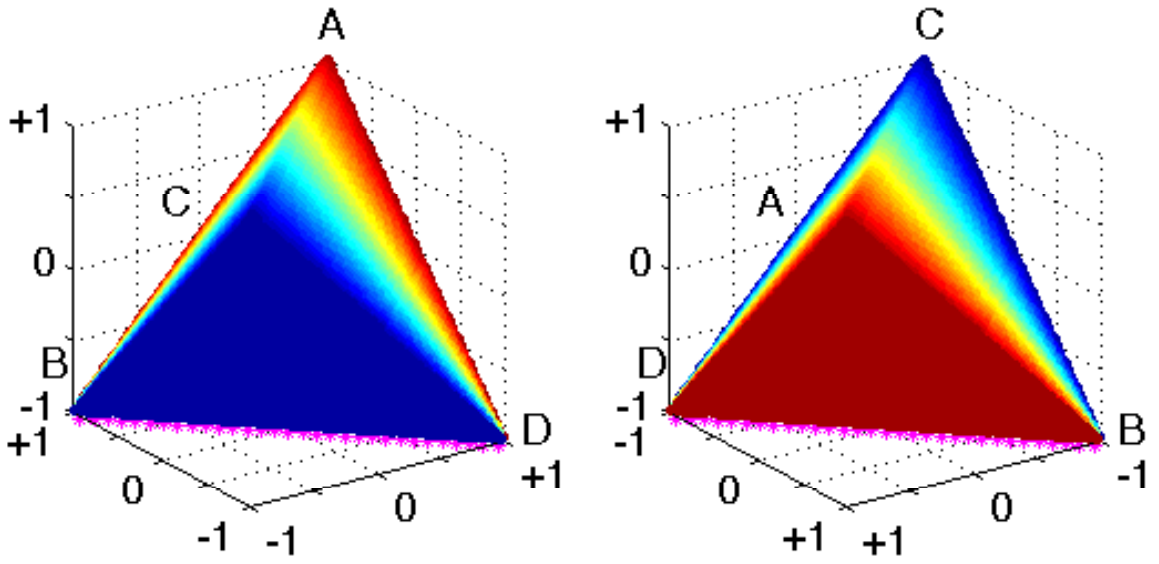
f(a,b,c,d) = a/(a+b) (precision)



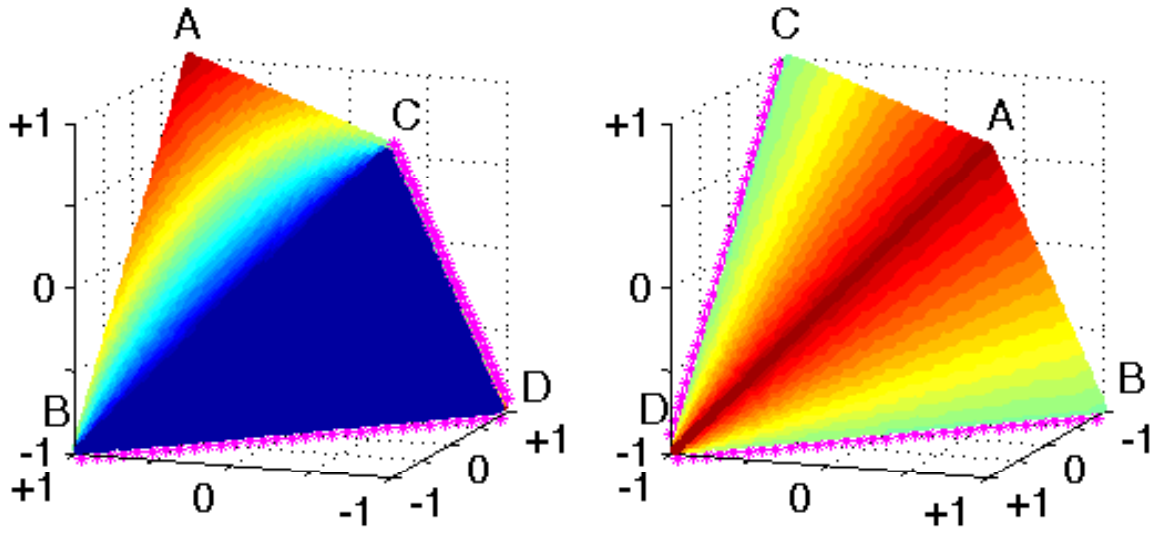
f(a,b,c,d) = a/(a+b) (precision)



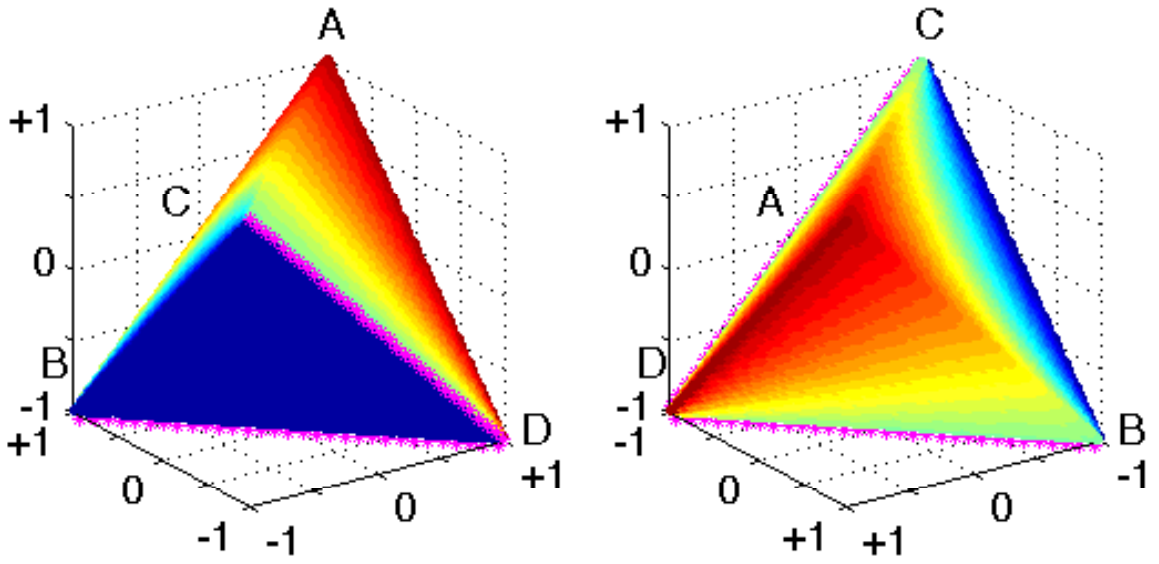
f(a,b,c,d) = a/(a+c) (recall)



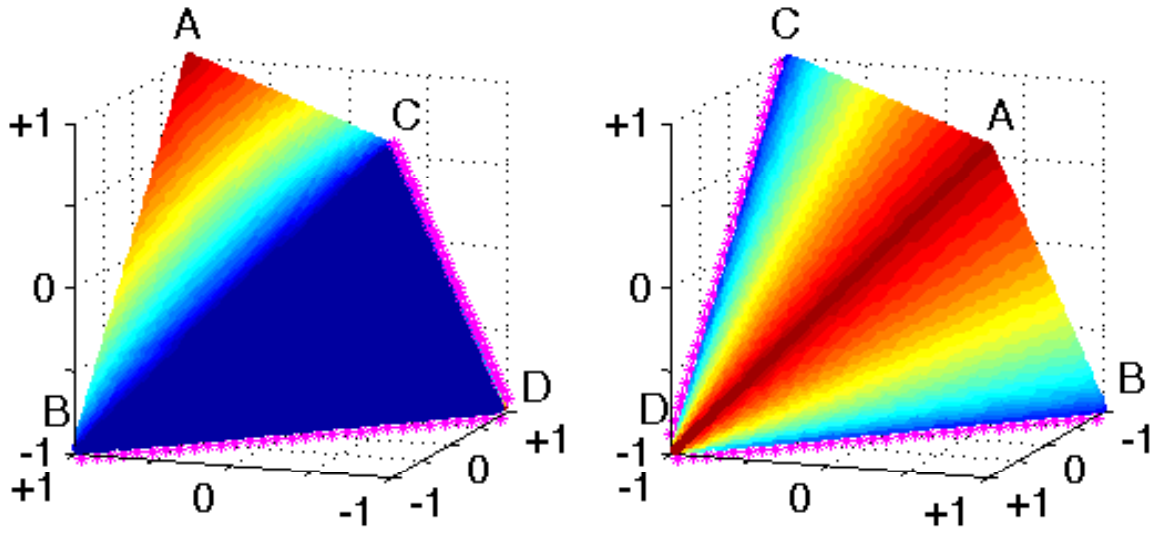
f(a,b,c,d) = a/(a+c) (recall)



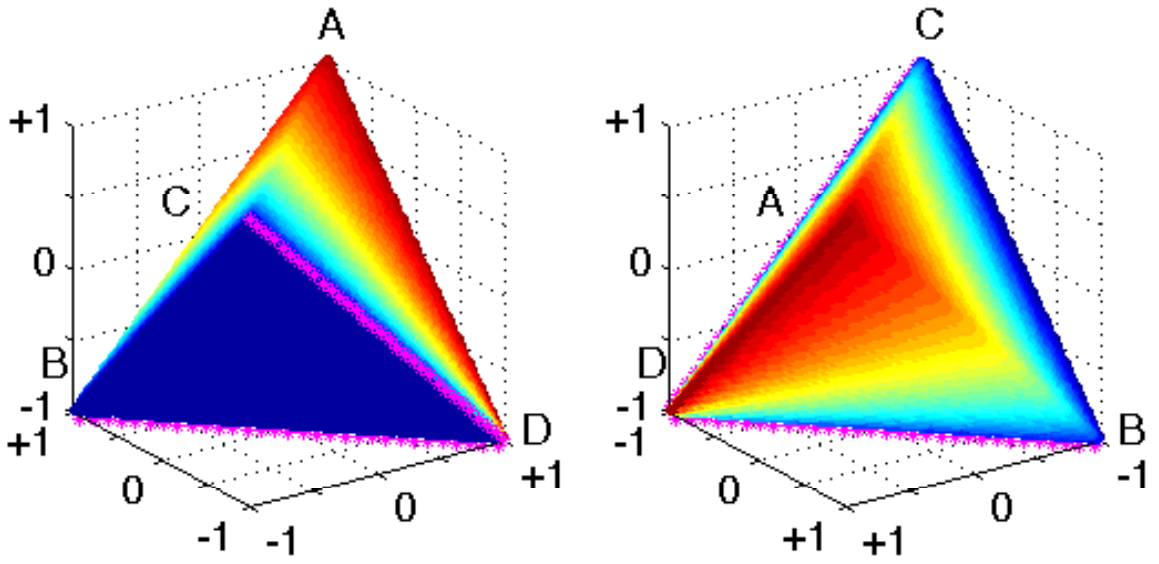
f(a,b,c,d) = amean(precision,recall))



f(a,b,c,d) = amean(precision,recall))



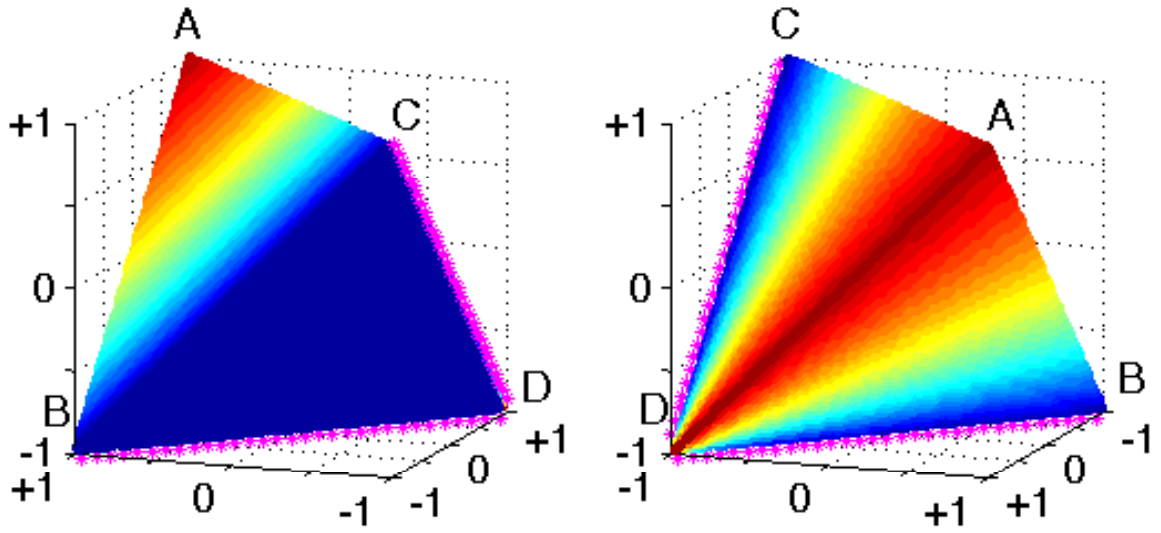
f(a,b,c,d) = gmean(precision,recall))



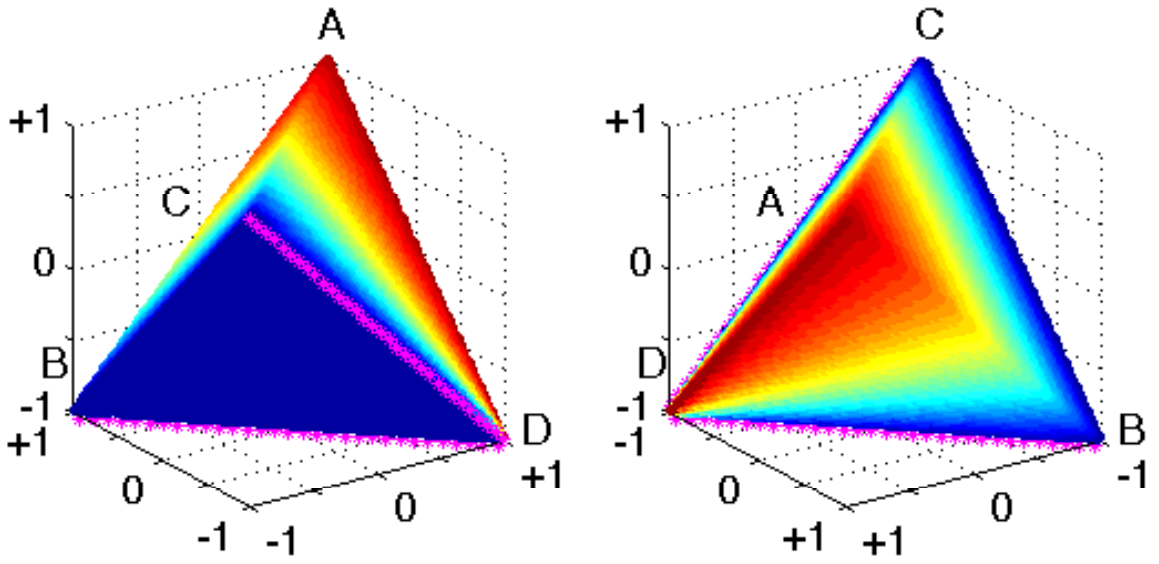
f(a,b,c,d) = gmean(precision,recall))

gmean(precision,recall)) = G-measure

$$A \geq G \geq H$$



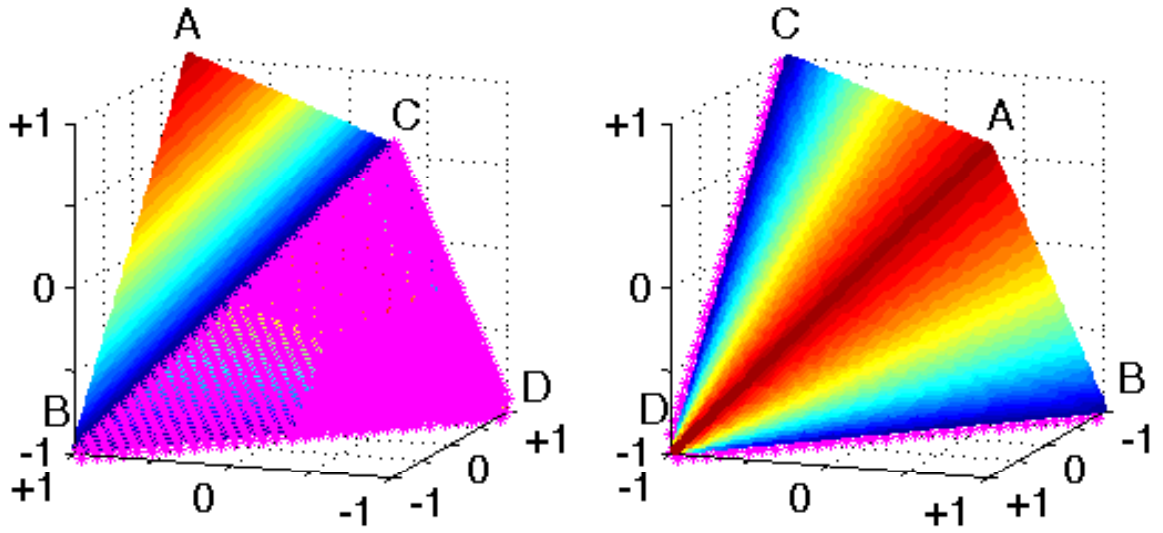
f(a,b,c,d) = hmean(precision,recall))



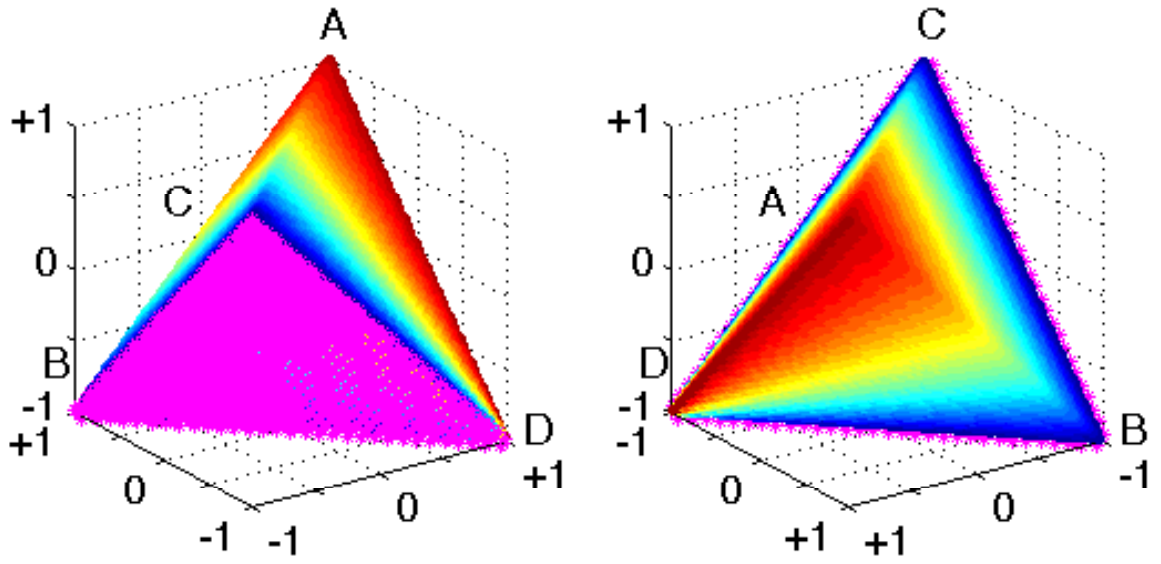
f(a,b,c,d) = hmean(precision,recall))

hmean(precision,recall)) = F_1

- wyjątki: wartości specjalne

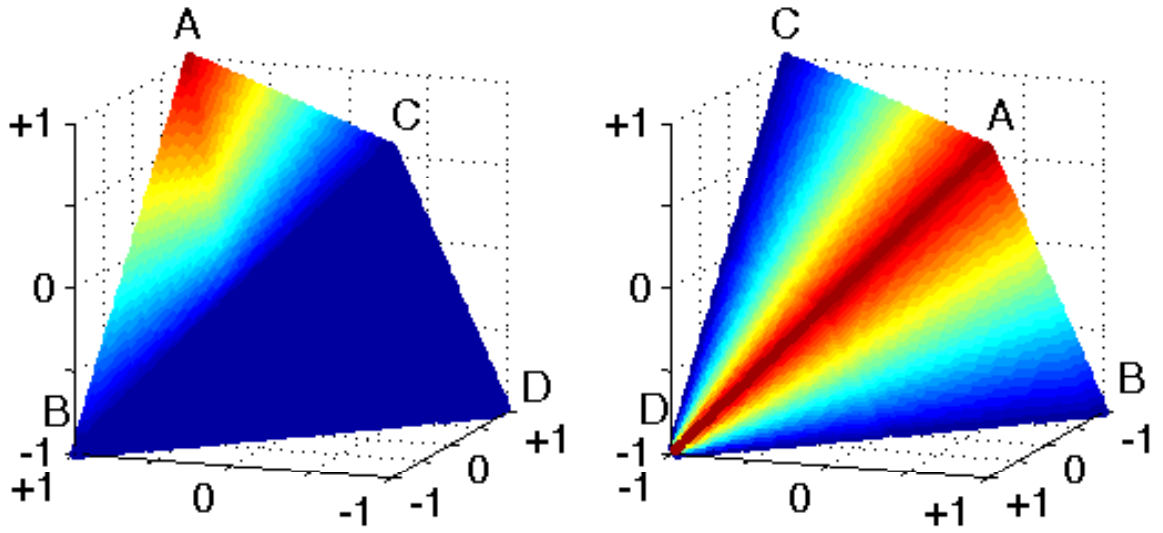


$$f(a,b,c,d) = F_1$$

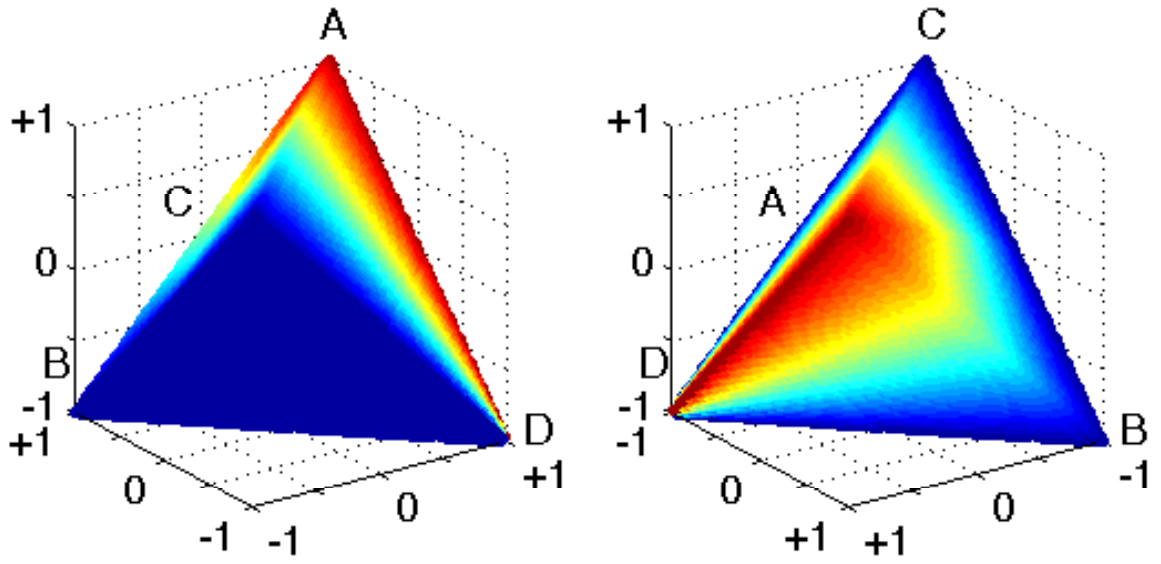


$f(a,b,c,d) = F_1$

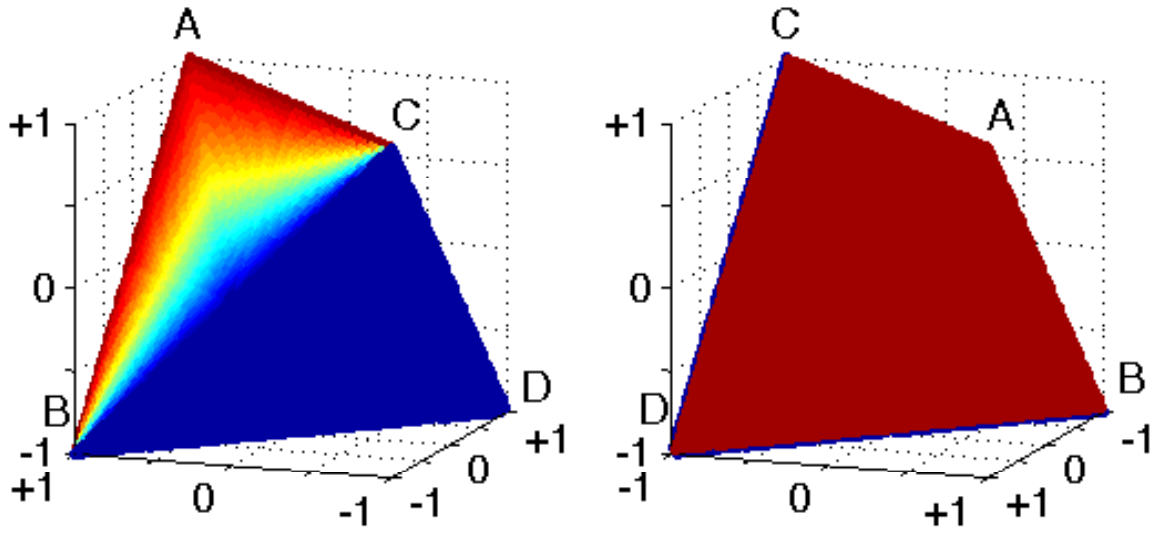
$$\max \geq A \geq G \geq H \geq \min$$



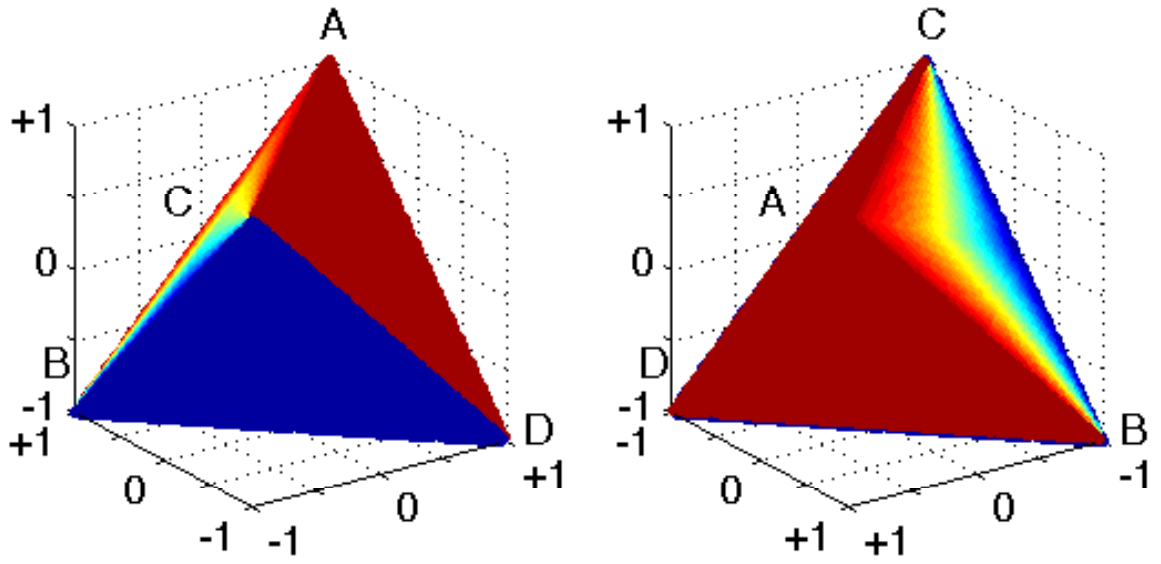
f(a,b,c,d) = min(precision,recall))



f(a,b,c,d) = min(precision,recall))

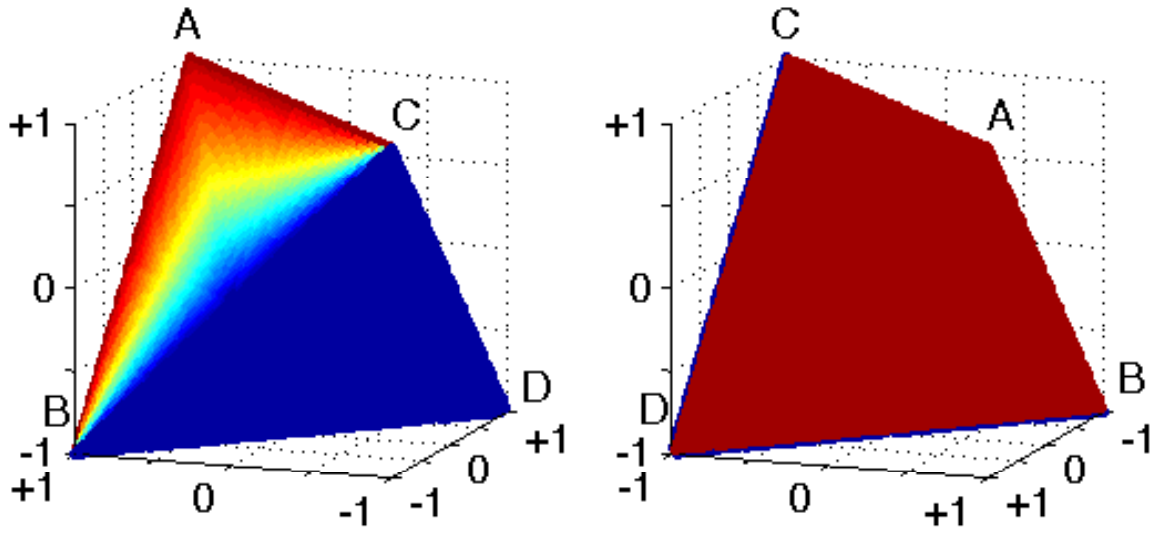


f(a,b,c,d) = max(precision,recall))

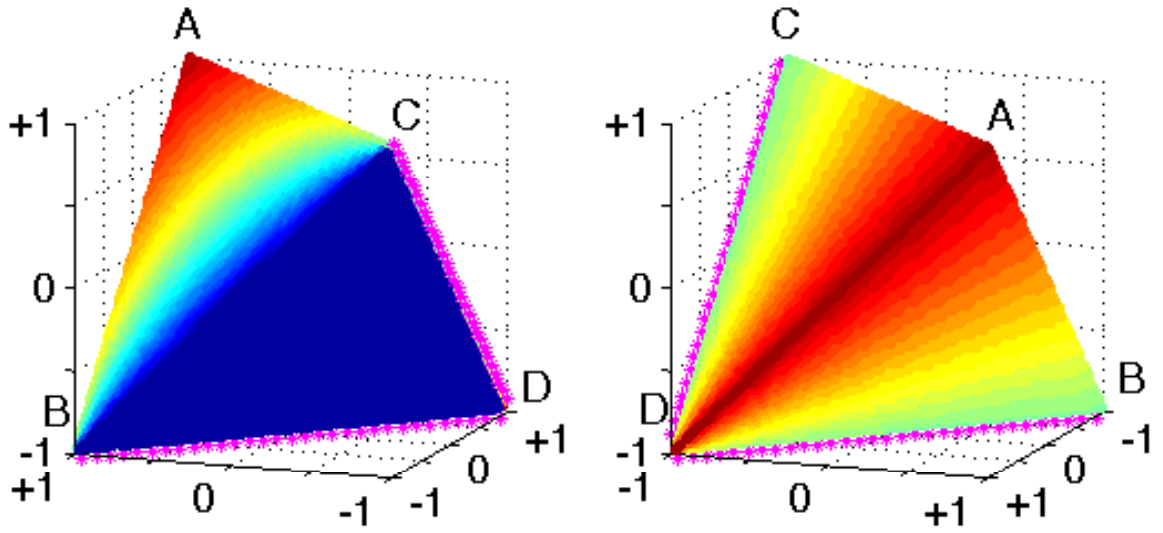


f(a,b,c,d) = max(precision,recall))

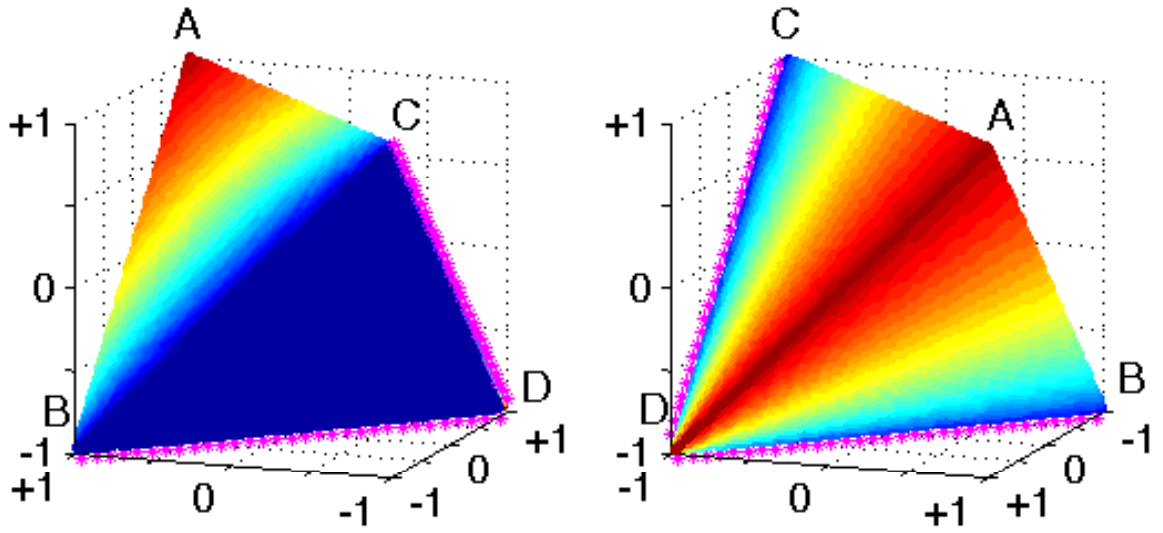
- sukcesja: max $\rightarrow$ A $\rightarrow$ G $\rightarrow$ H $\rightarrow$ min
  - wizualizacja a c / b d (+ rewers)



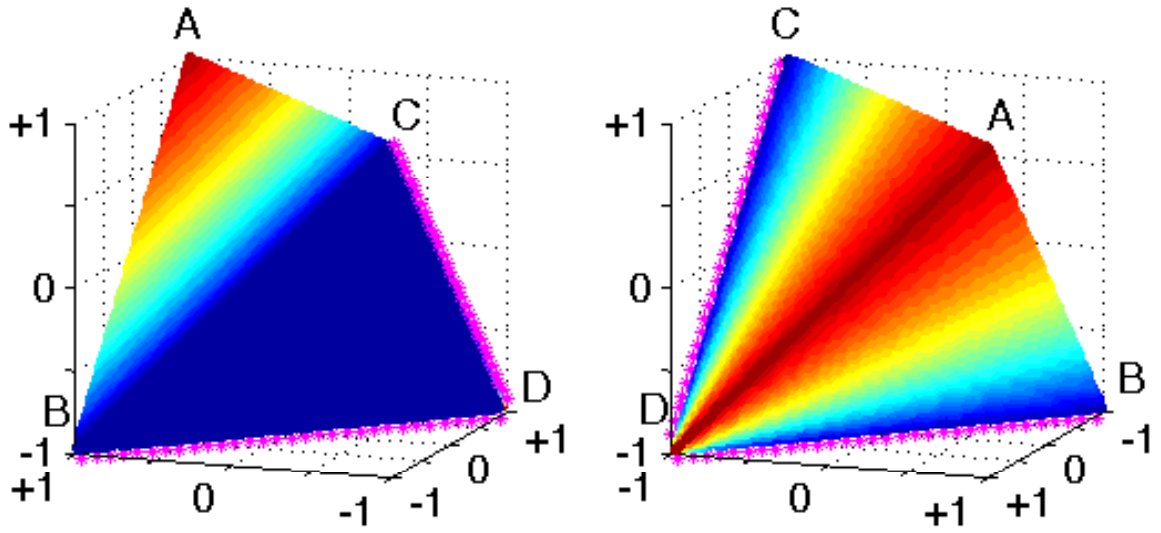
f(a,b,c,d) = max(precision,recall))



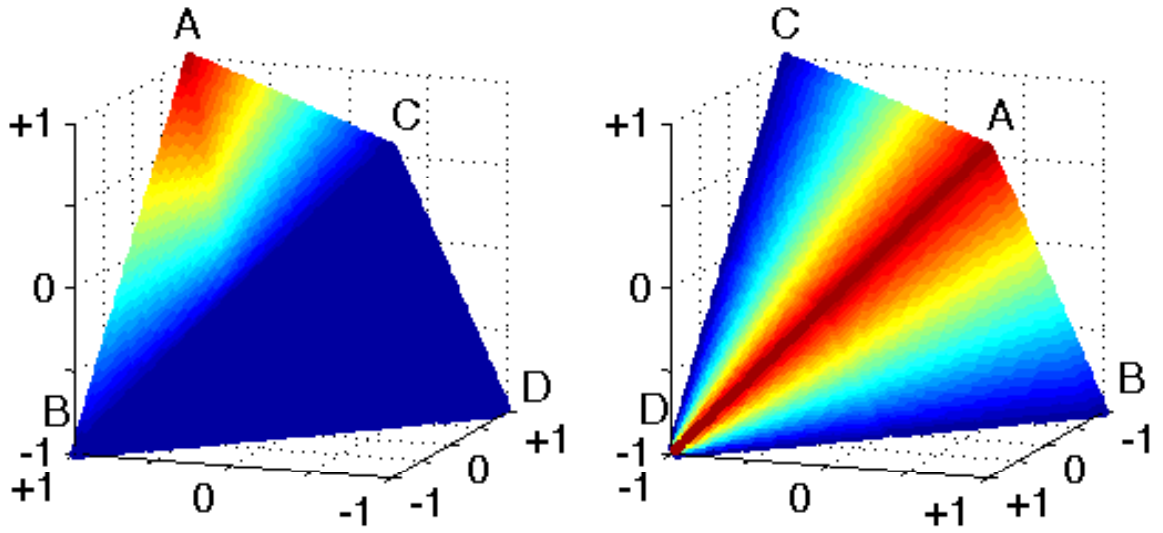
f(a,b,c,d) = amean(precision,recall))



f(a,b,c,d) = gmean(precision,recall))

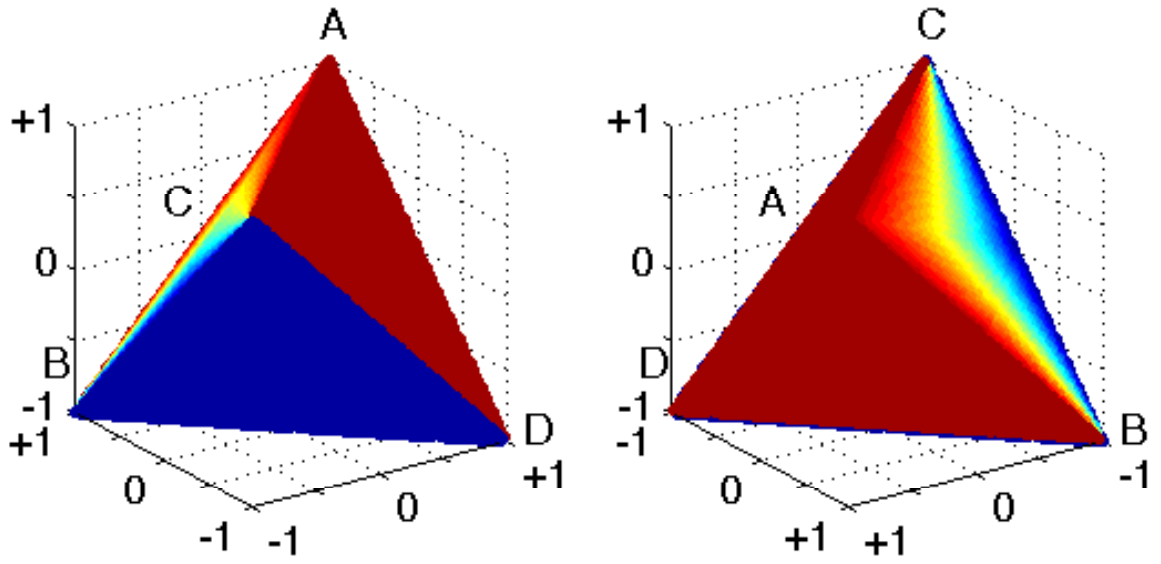


f(a,b,c,d) = hmean(precision,recall))

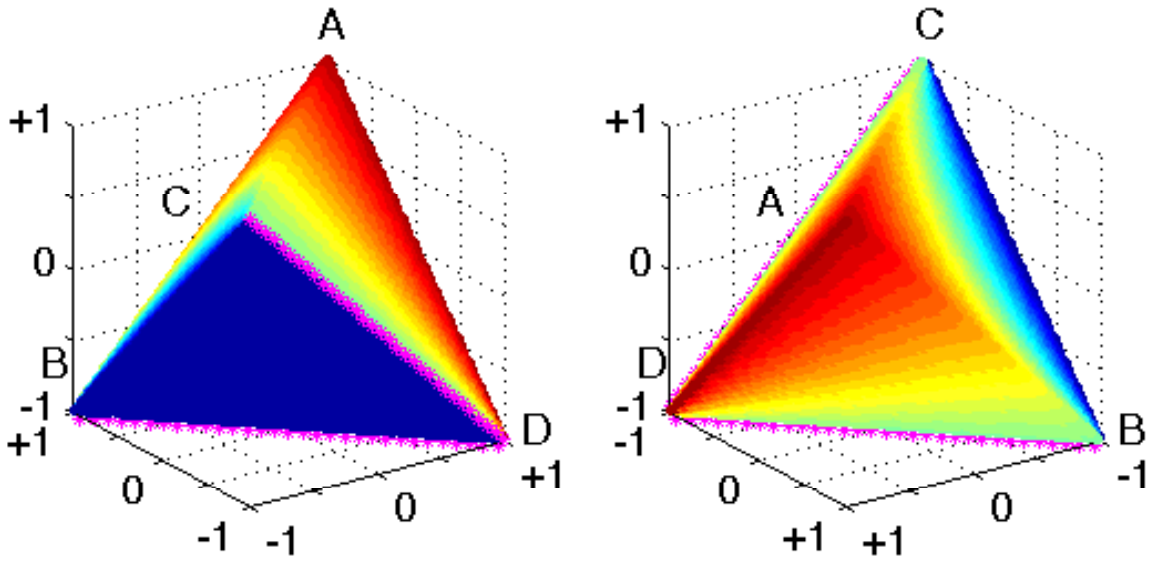


f(a,b,c,d) = min(precision,recall))

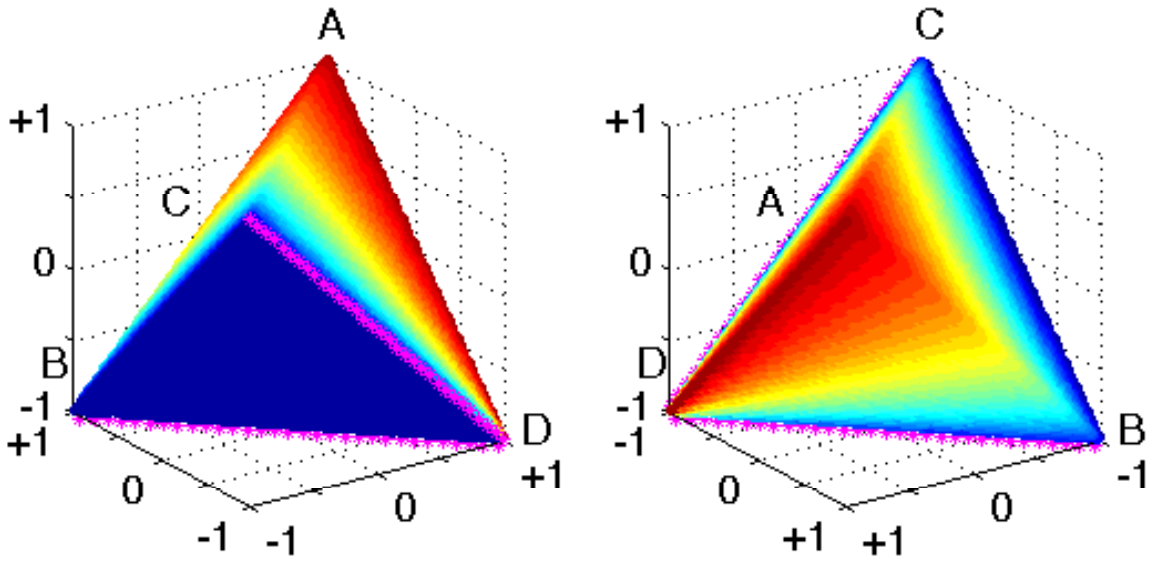
- sukcesja: max $\rightarrow$ A $\rightarrow$ G $\rightarrow$ H $\rightarrow$ min
  - wizualizacja c a / b d (+ rewers)



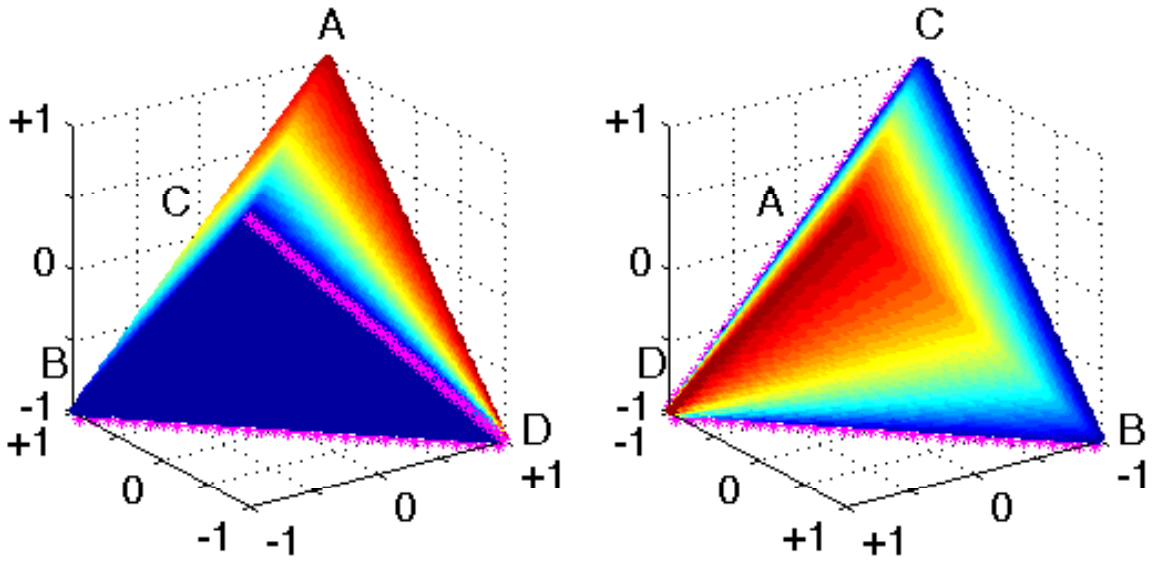
f(a,b,c,d) = max(precision,recall))



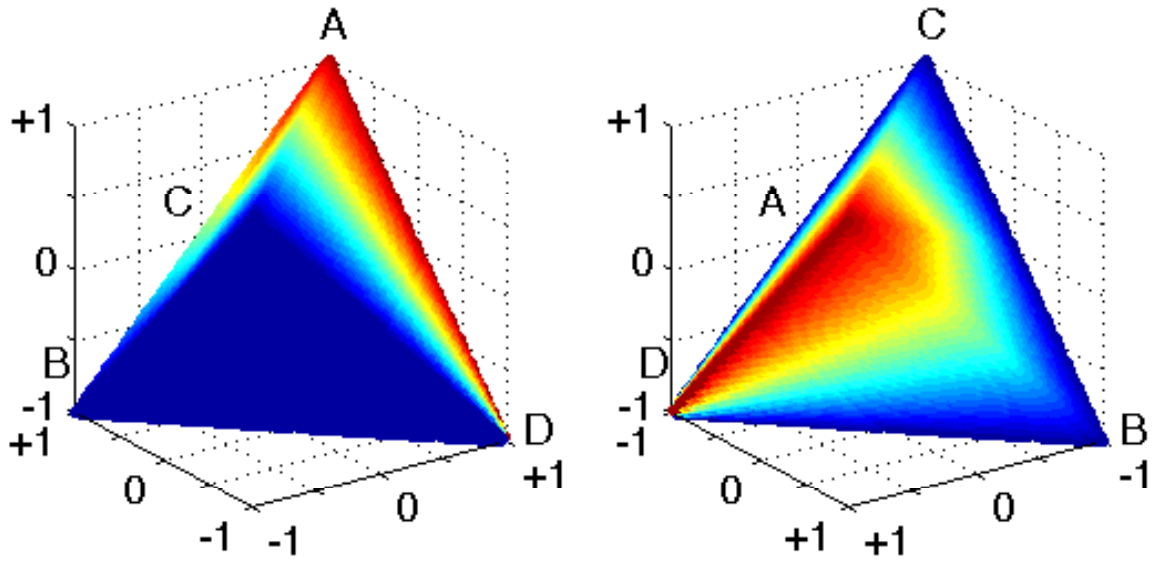
f(a,b,c,d) = amean(precision,recall))



f(a,b,c,d) = gmean(precision,recall))

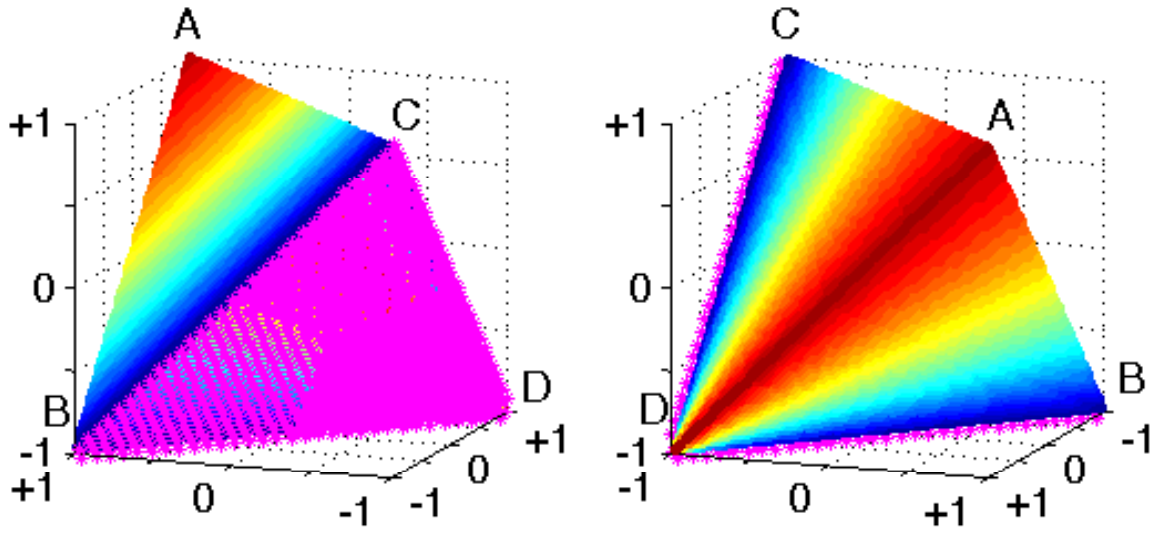


f(a,b,c,d) = hmean(precision,recall))

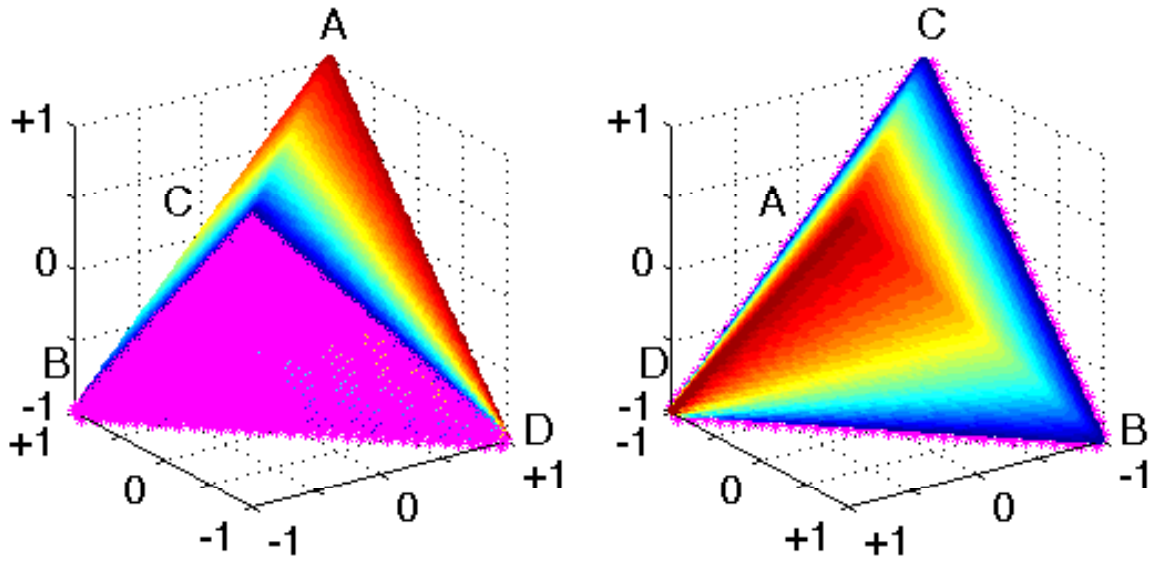


f(a,b,c,d) = min(precision,recall))

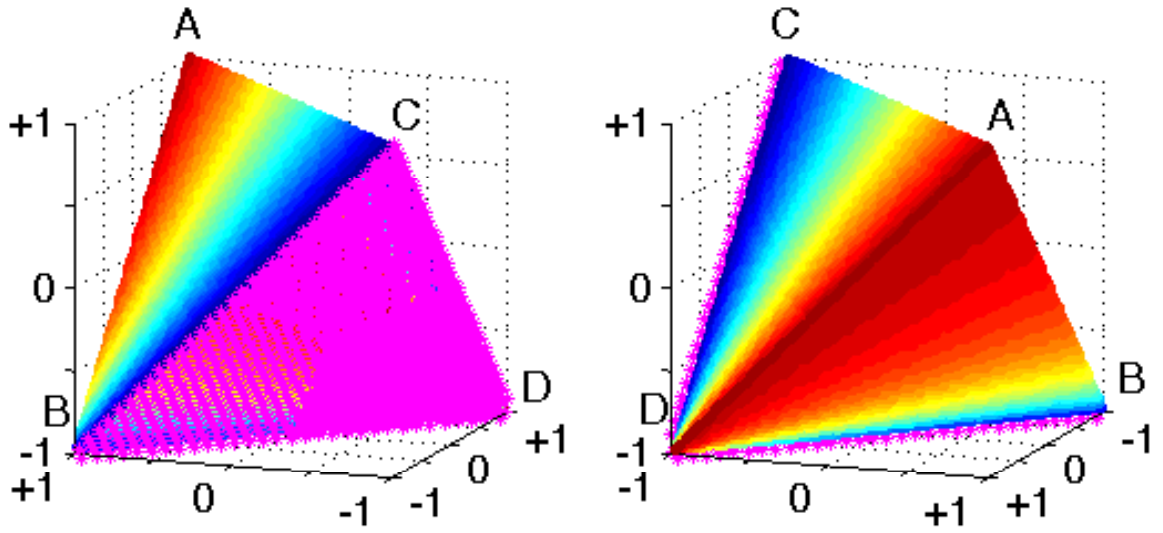
- Uogólnieniem F_1 (hmean) jest F_β (dla $\beta \neq 1$)



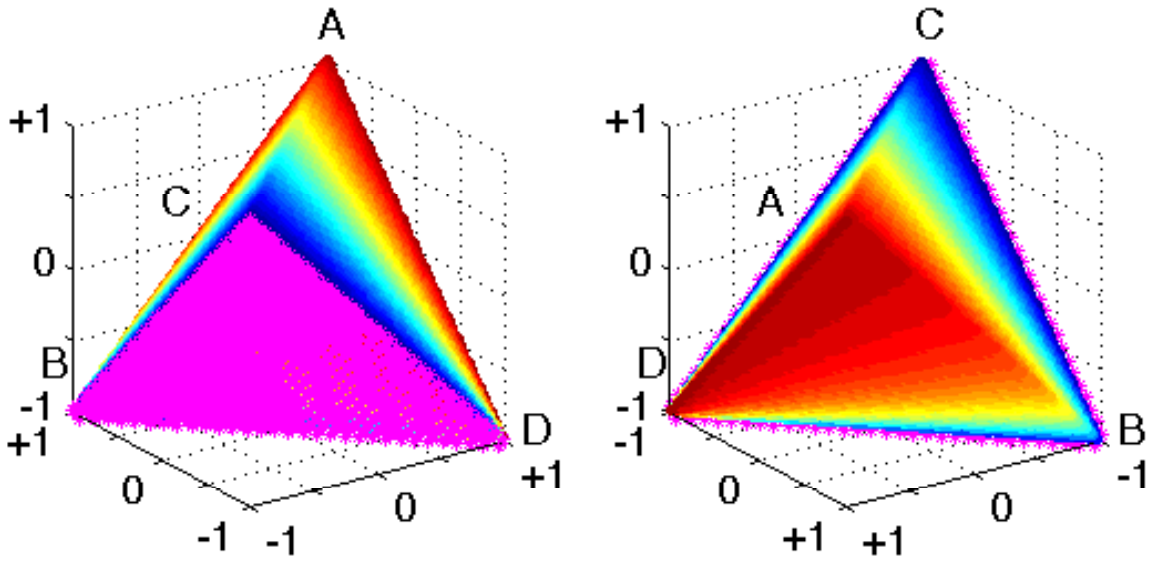
$$f(a,b,c,d) = F_1$$



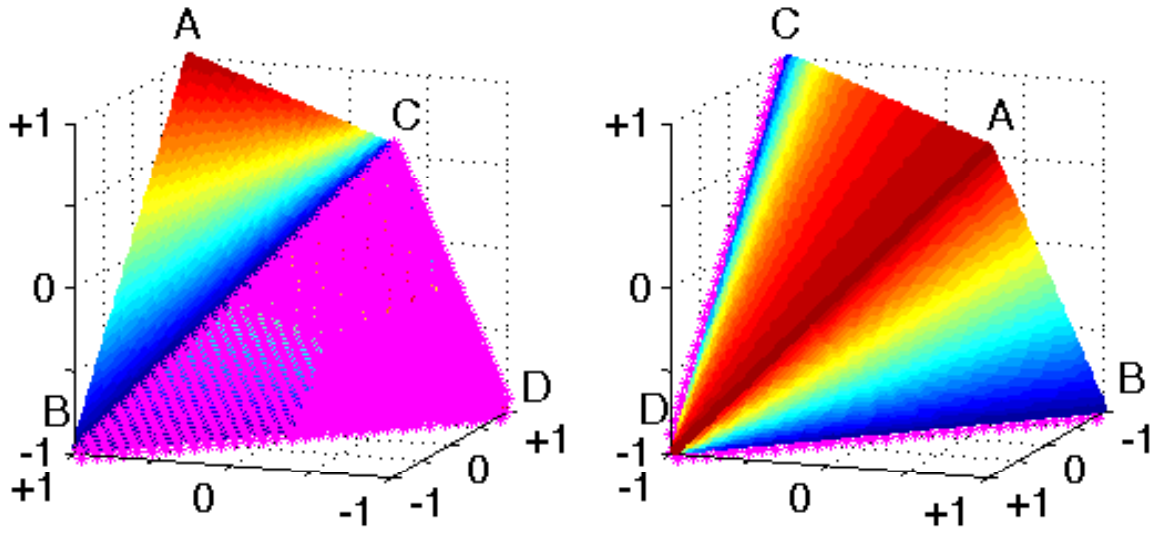
$f(a,b,c,d) = F_1$



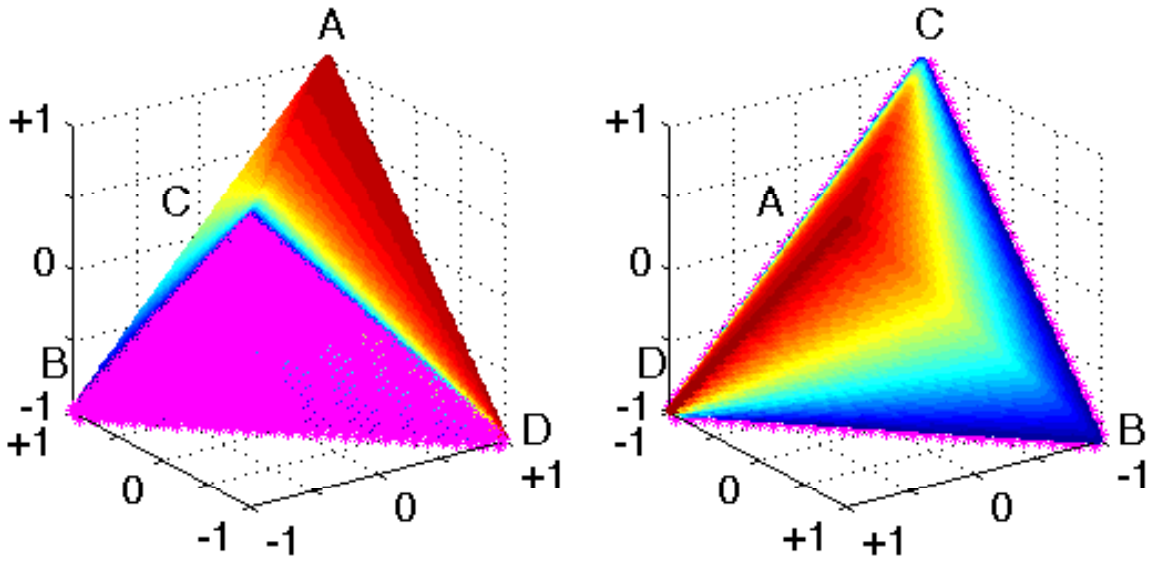
$f(a,b,c,d) = F_3$



$f(a,b,c,d) = F_3$

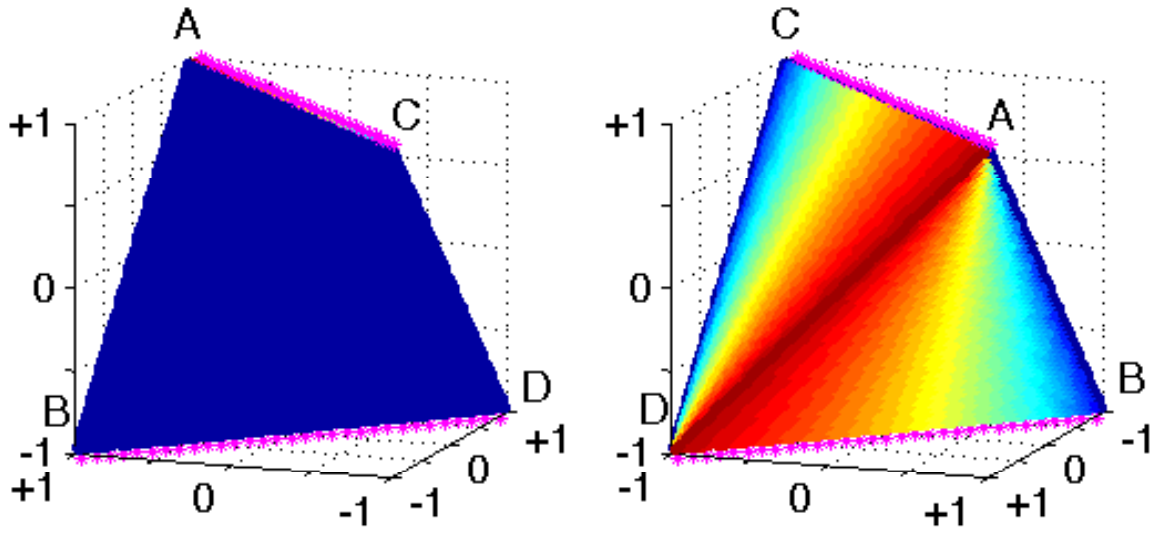


$$f(a,b,c,d) = F_{1/3}$$

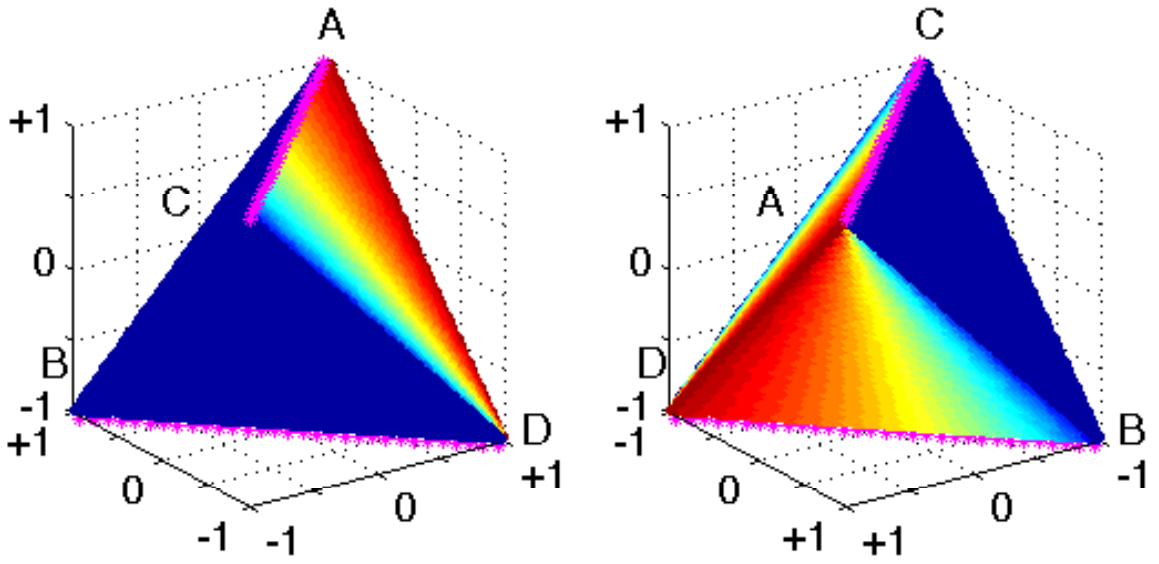


$$f(a,b,c,d) = F_{1/3}$$

- G-mean:
  - gmean(recall,specificity)
    - gmean(TP/(TP+FN),TN/(TN+FP)) = gmean(a/(a+c),d/(d+b))

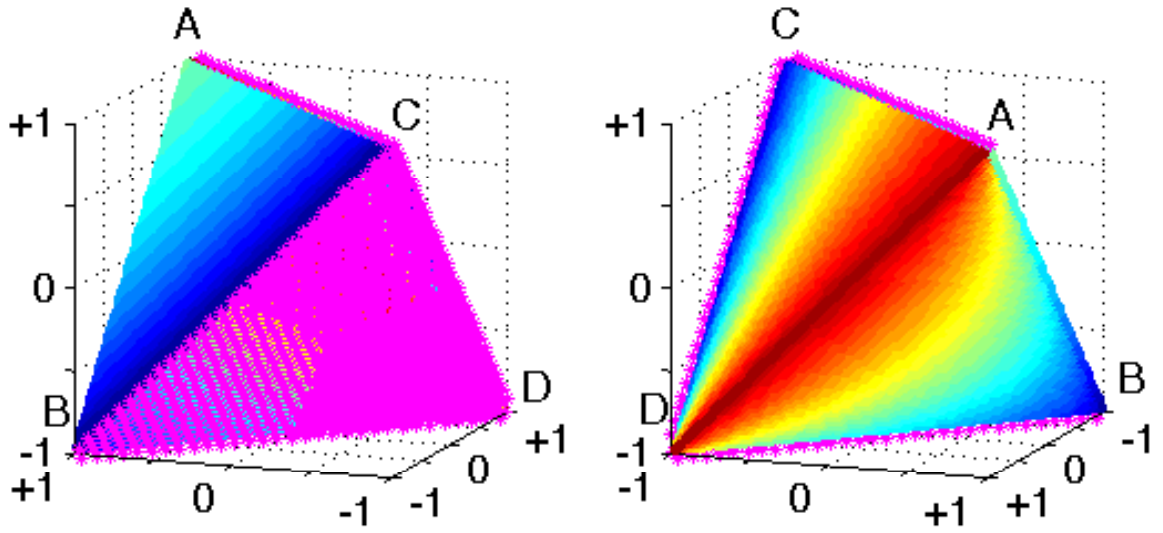


f(a,b,c,d) = G-mean

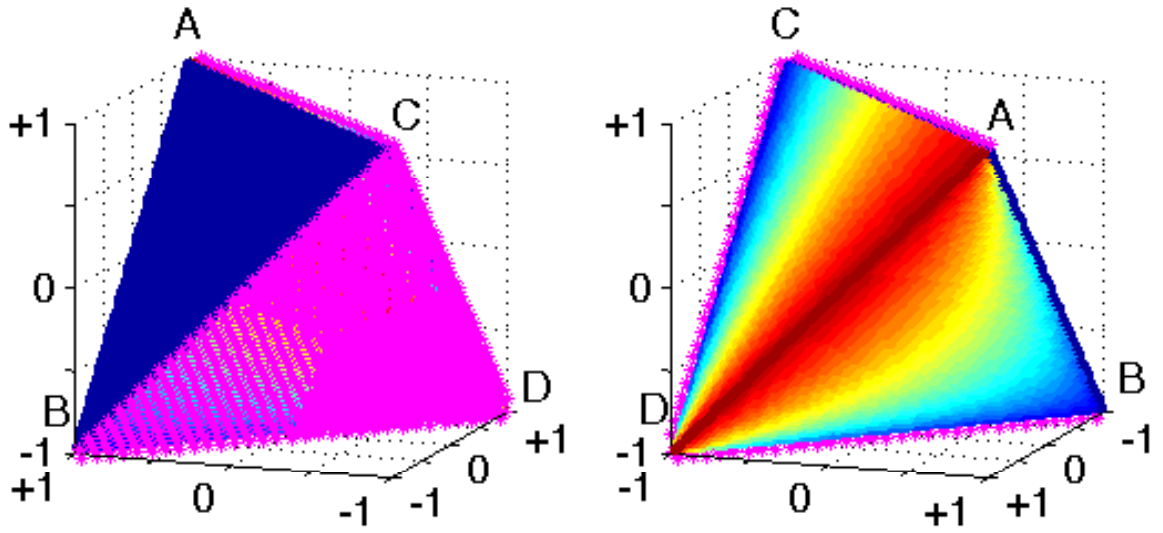


f(a,b,c,d) = G-mean

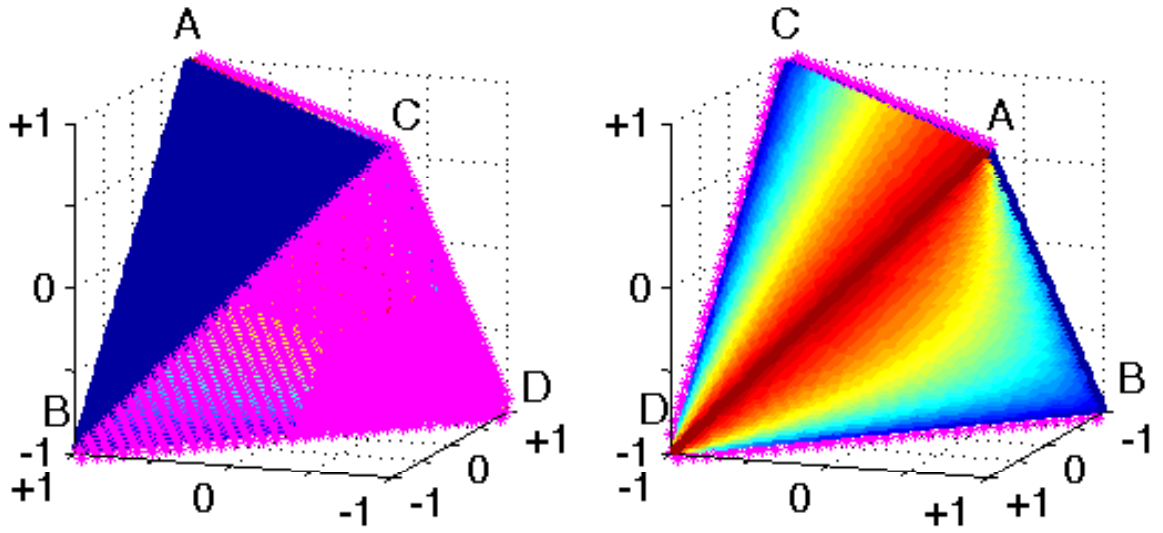
- kombinacje F_β i G-mean?



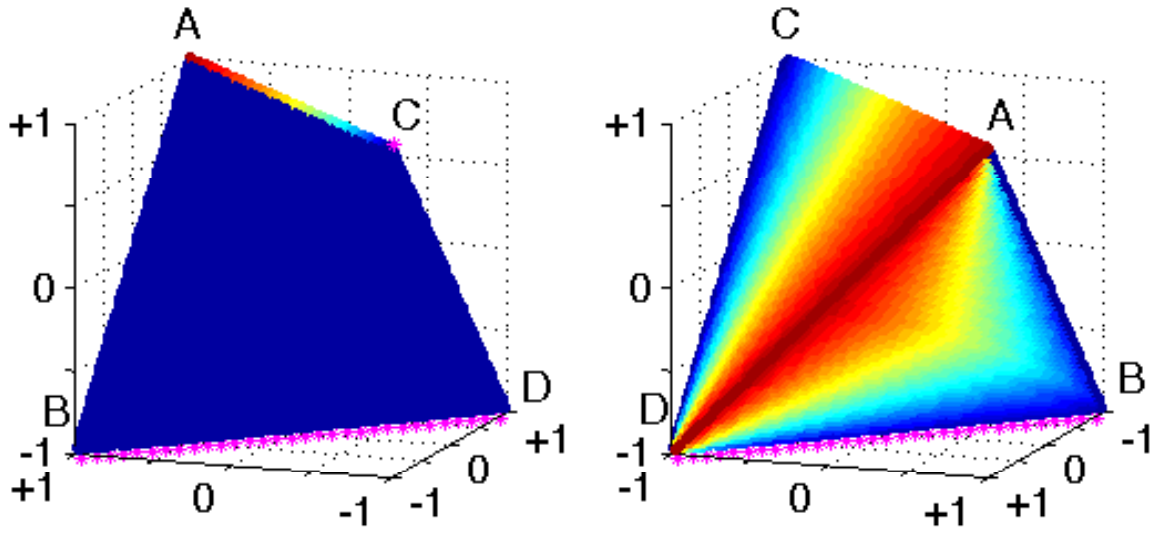
$f(a,b,c,d) = \text{amean}(F_\beta, \text{G-mean})$



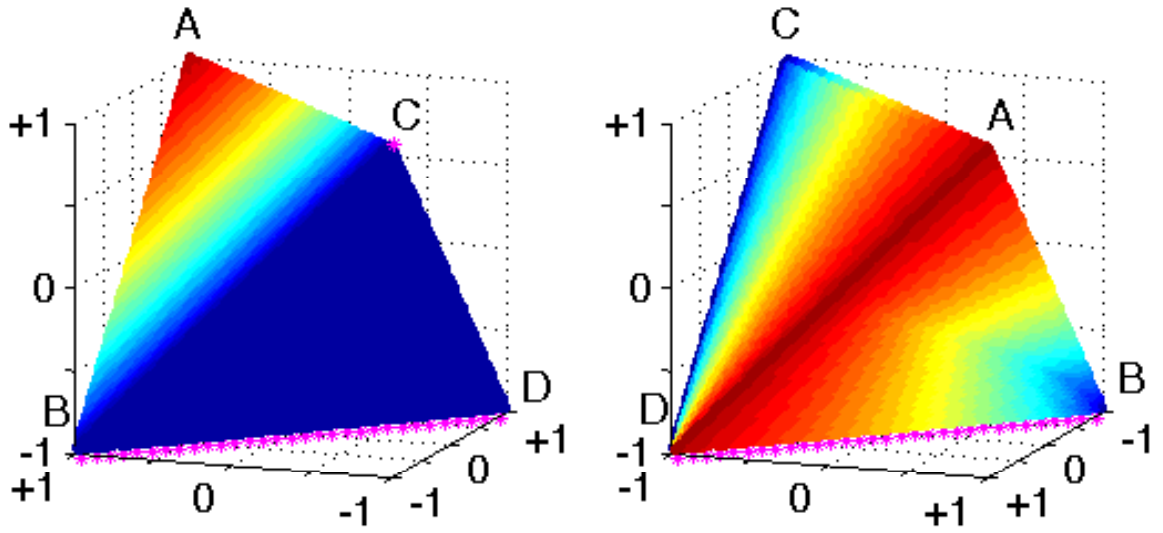
f(a,b,c,d) = gmean(F_β,G-mean)



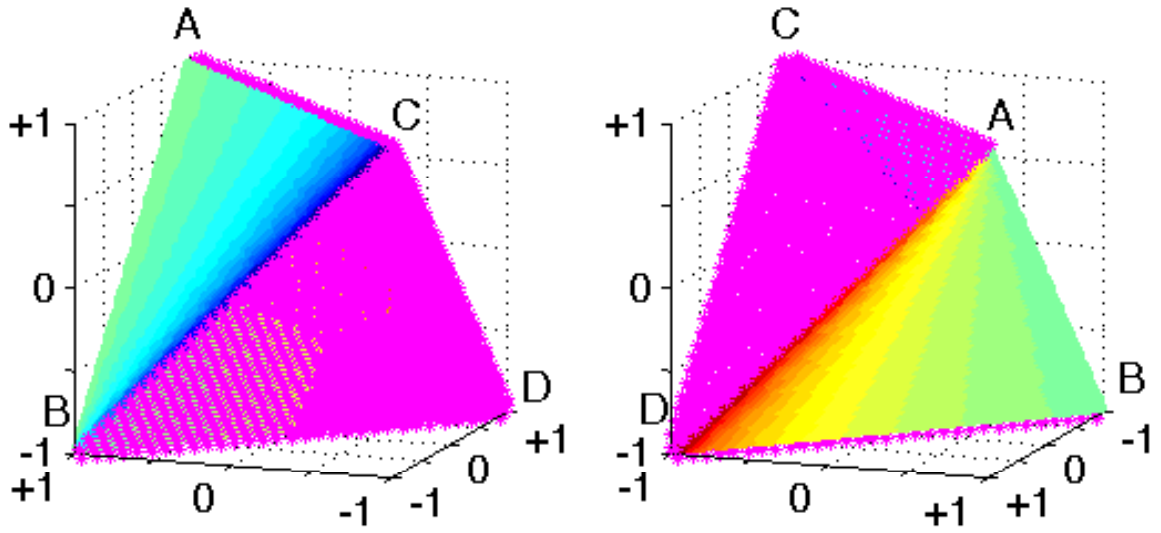
$f(a,b,c,d) = \text{hmean}(F_\beta, \text{G-mean})$



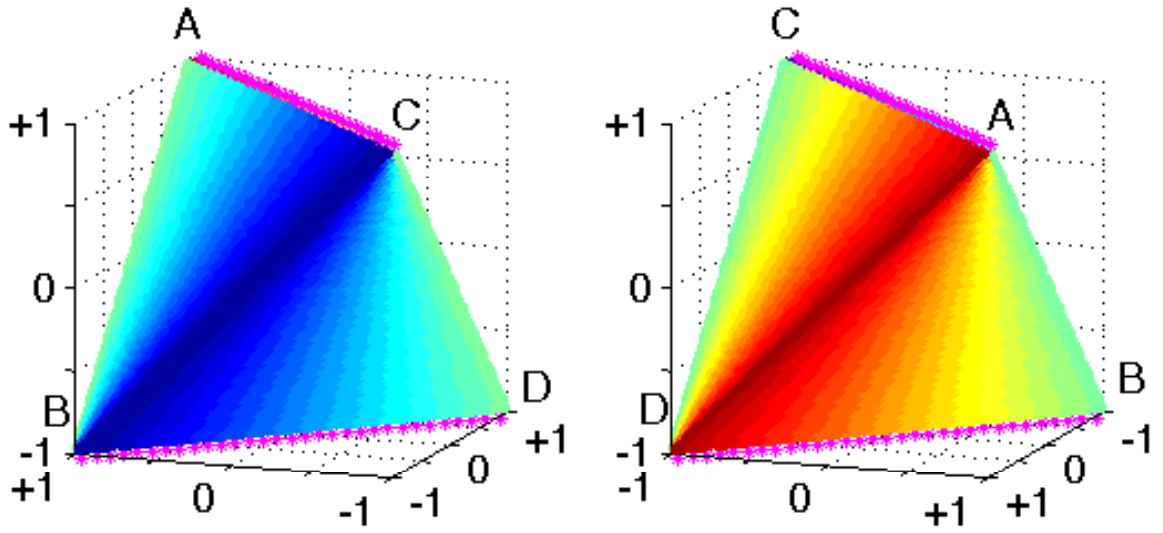
$$f(a,b,c,d) = \min(F_\beta, \text{G-mean})$$



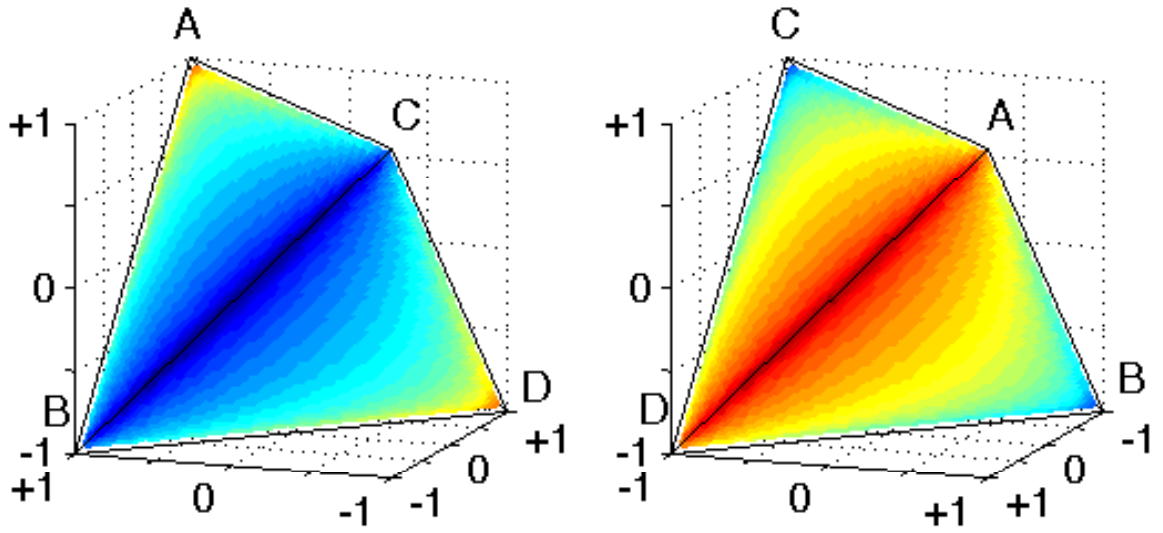
$f(a,b,c,d) = \max(F_{\beta}, \text{G-mean})$



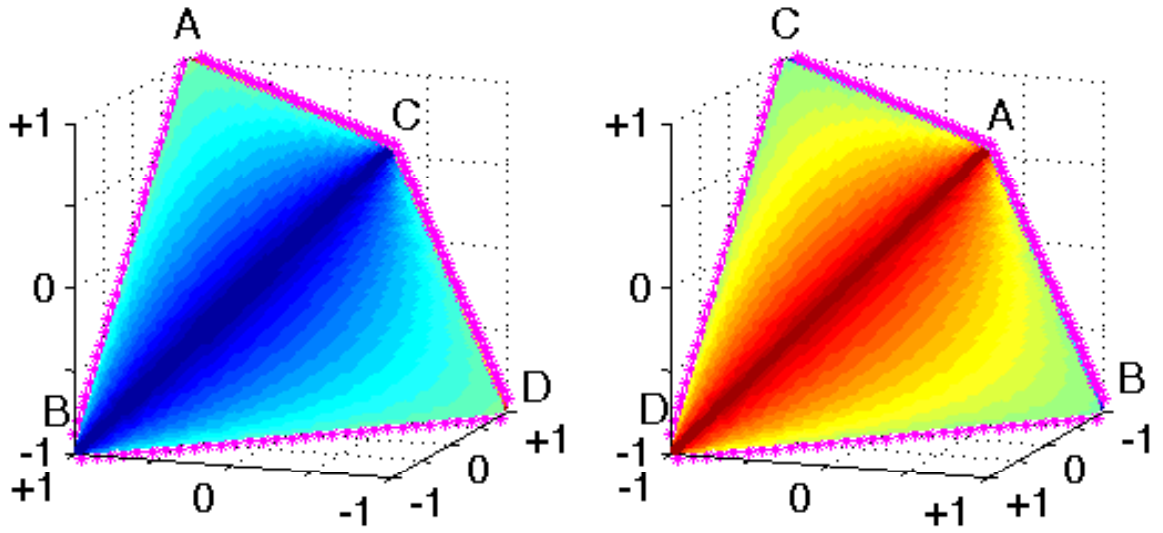
f(a,b,c,d) = Positive likelihood ratio



f(a,b,c,d) = Balanced accuracy

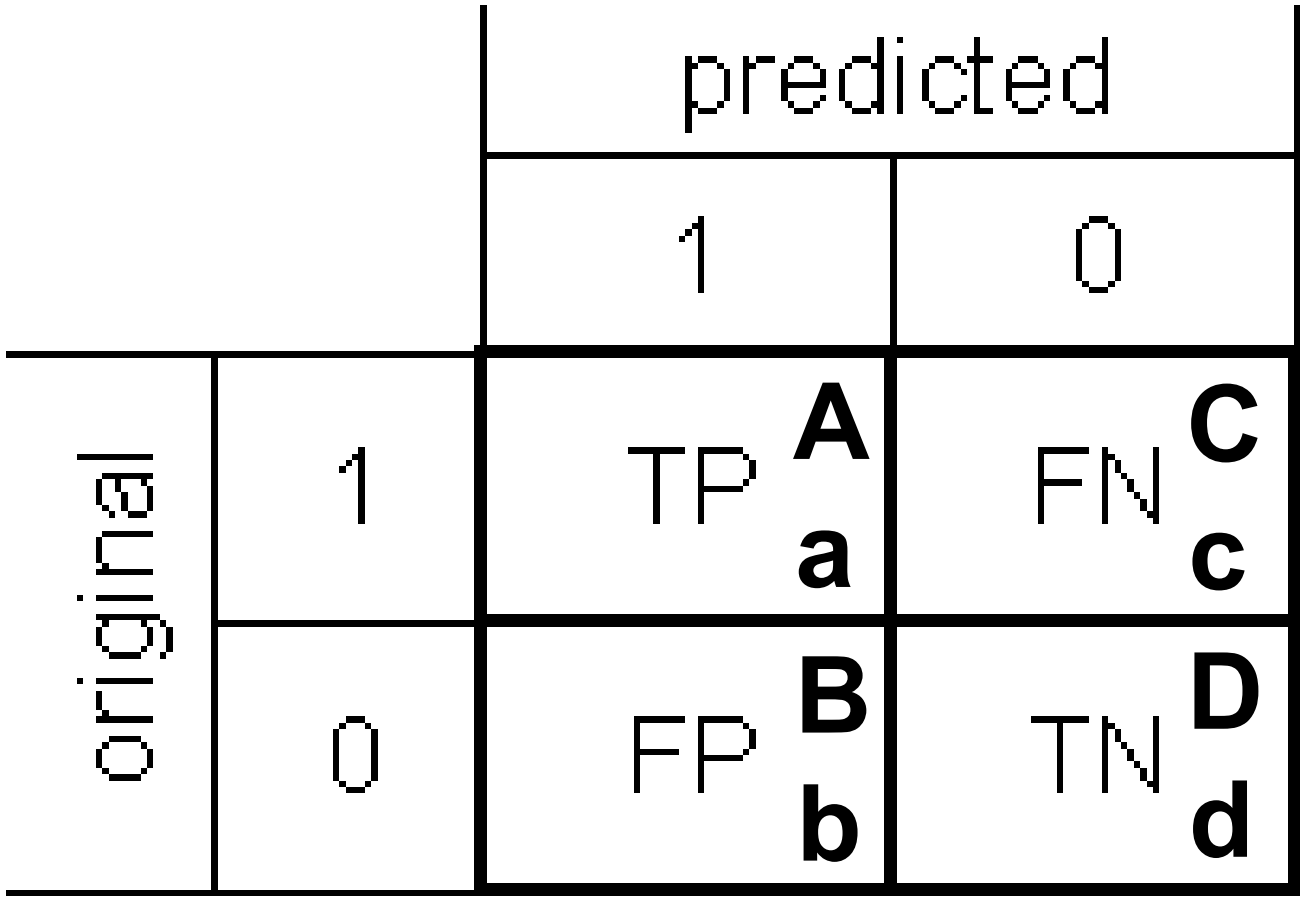


f(a,b,c,d) = Discriminant power



f(a,b,c,d) = MCC

...

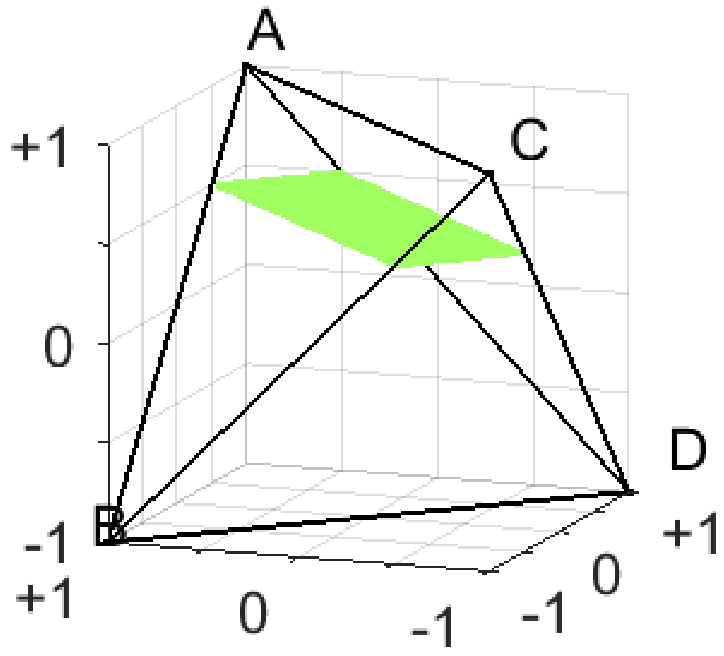

predicted
1
0
original
1
0
TP
A
a
FN
C
c
FP
B
b
TN
D
d

- Cel rozważań: relacje pomiędzy licznościami klas

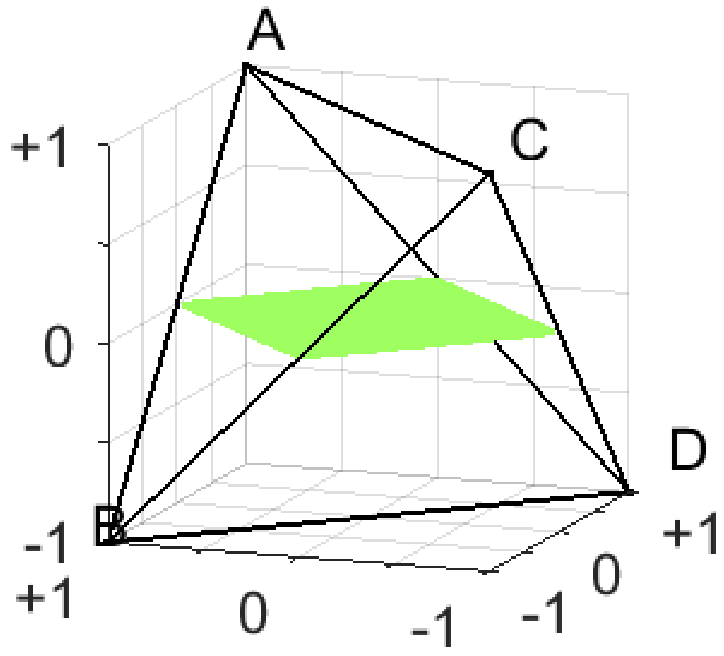
- liczność:
  - klasy 1: a+c
  - klasy 0: b+d

- wizualizacja ustalonej proporcji klas, np. p : q (przy założeniu, że a+b+c+d = 1)
  - przecięcie czworościanu z pozioma płaszczyzną (przebiegająca na poziomie wyznaczonym przez p i q)

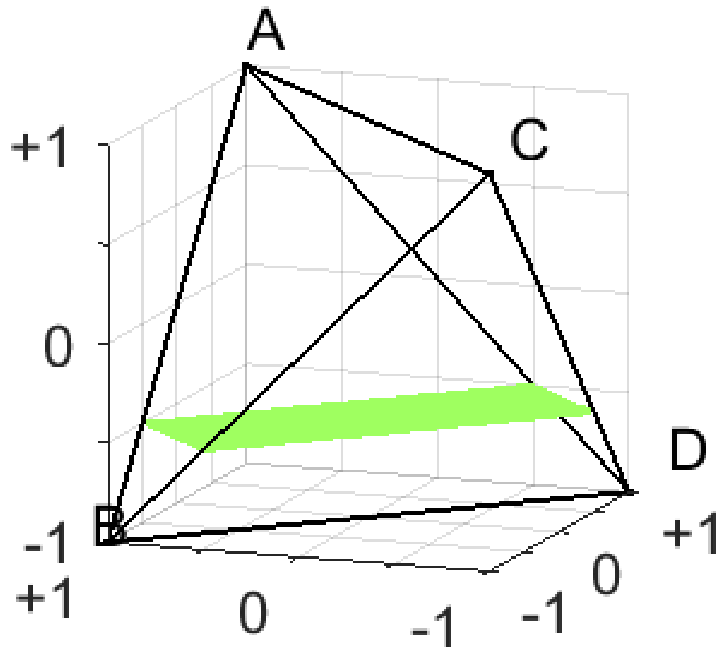
przewaga klasy 1



równowaga klas

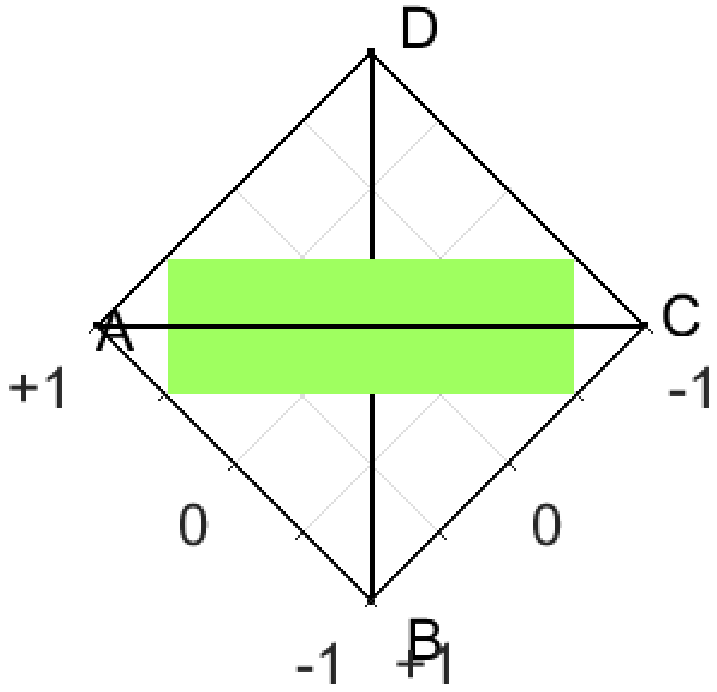


przewaga klasy 0

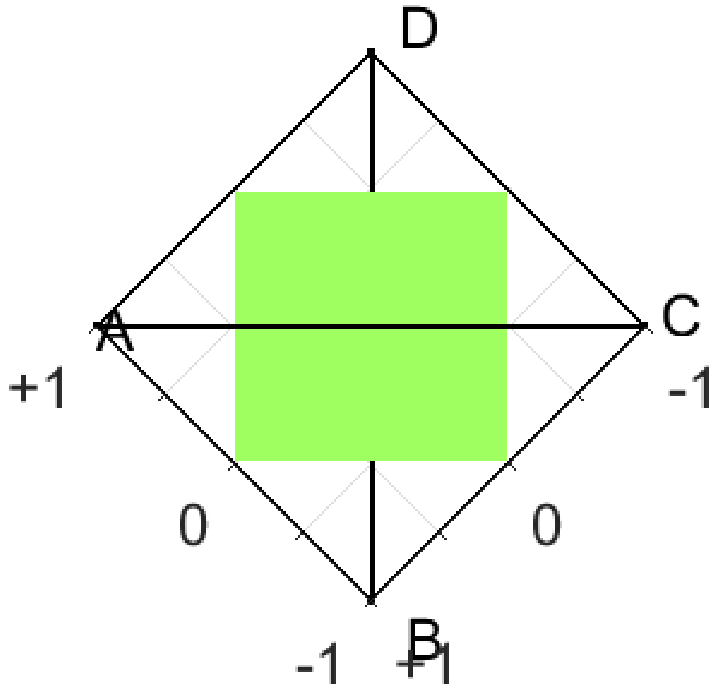


- dla ułatwienia: widok z góry
  (figura dodatkowo lekko obrócona w lewo)

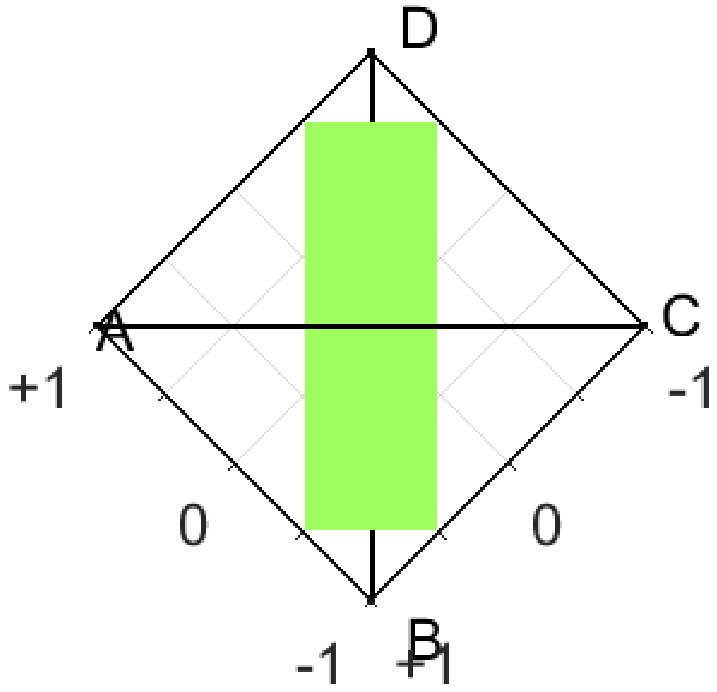
przewaga klasy 1



równowaga klas

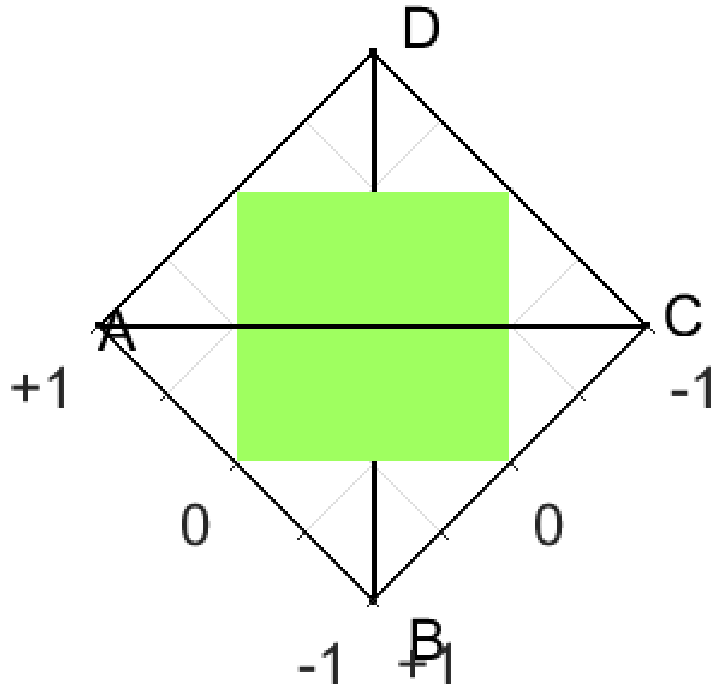


przewaga klasy 0



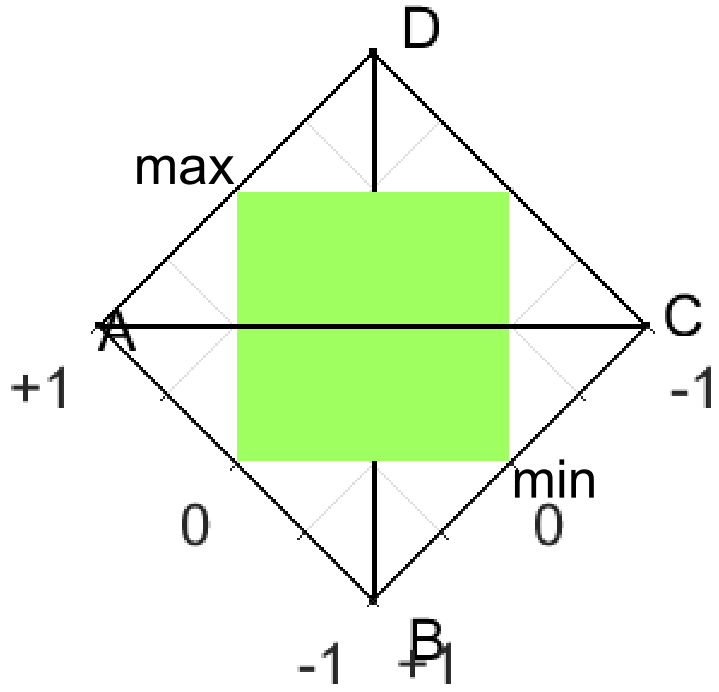
- na początek
  - równość klas: a+c = b+d

równowaga klas



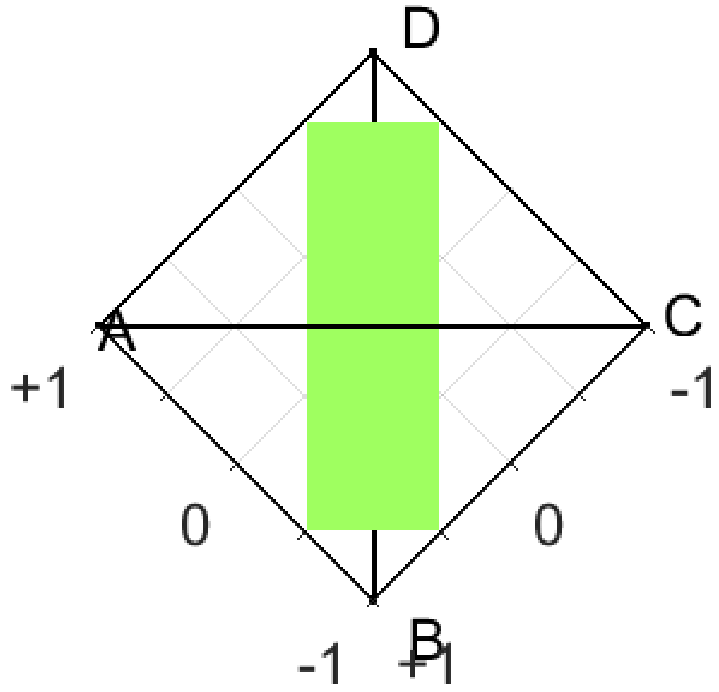
pożądany przebieg/wygląd miar?

równowaga klas

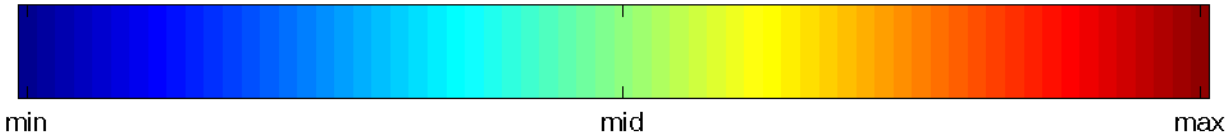


pożądany przebieg/wygląd miar

przewaga klasy 0

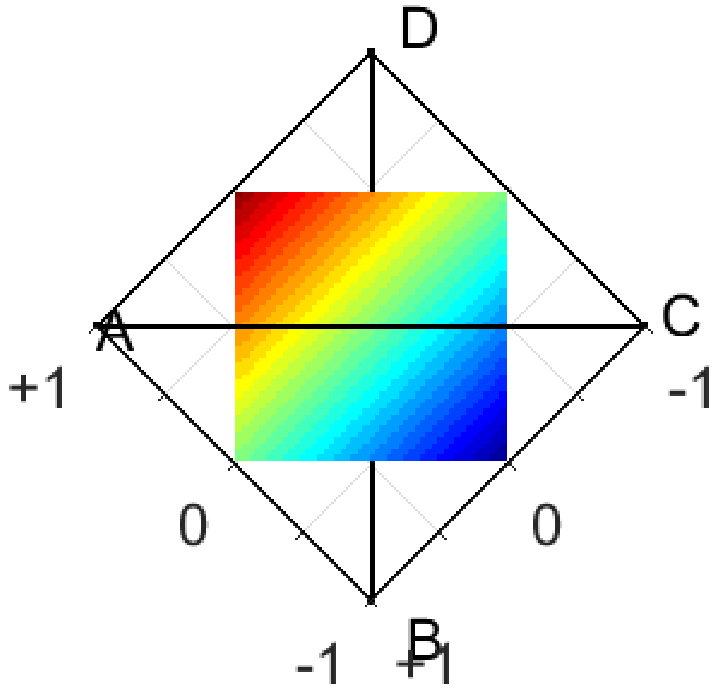


a teraz: pożądany przebieg/wygląd miar???


min
mid
max

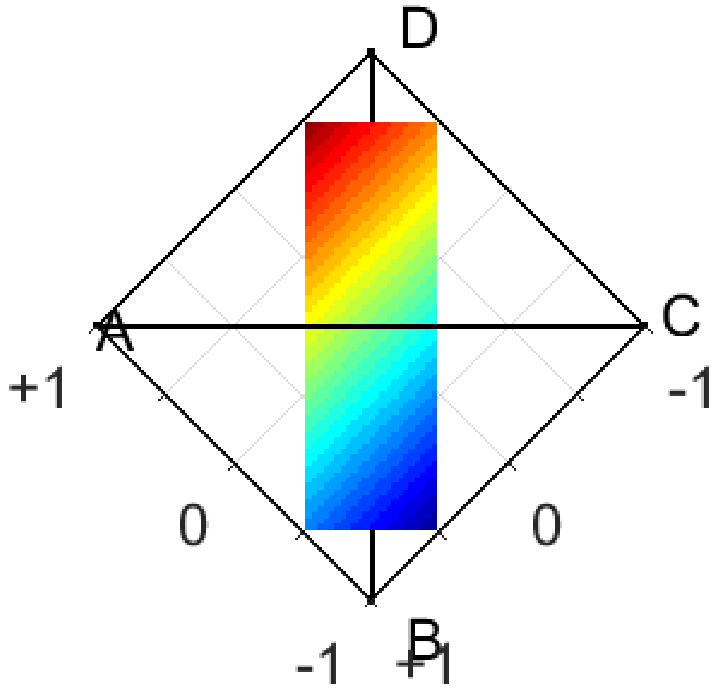
- miara na początek:
  - CA

równowaga klas



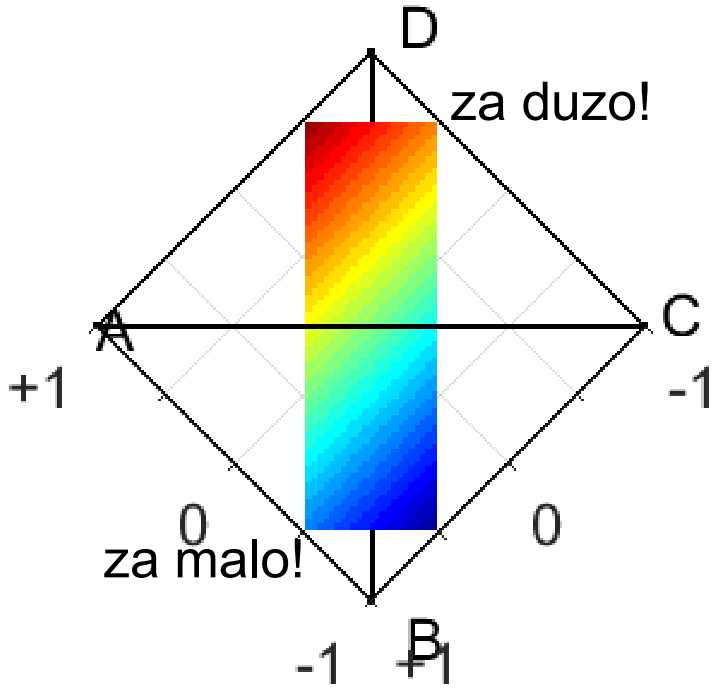
CA

przewaga klasy 0



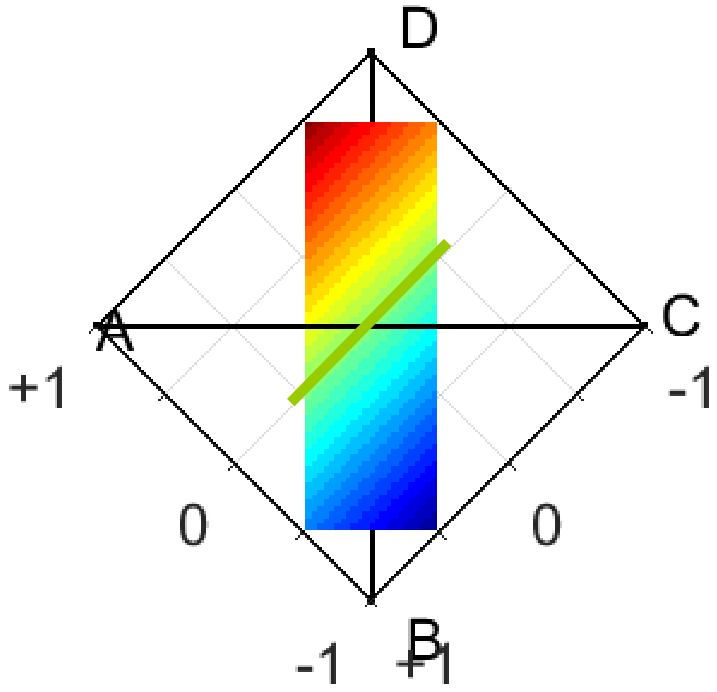
CA

przewaga klasy 0



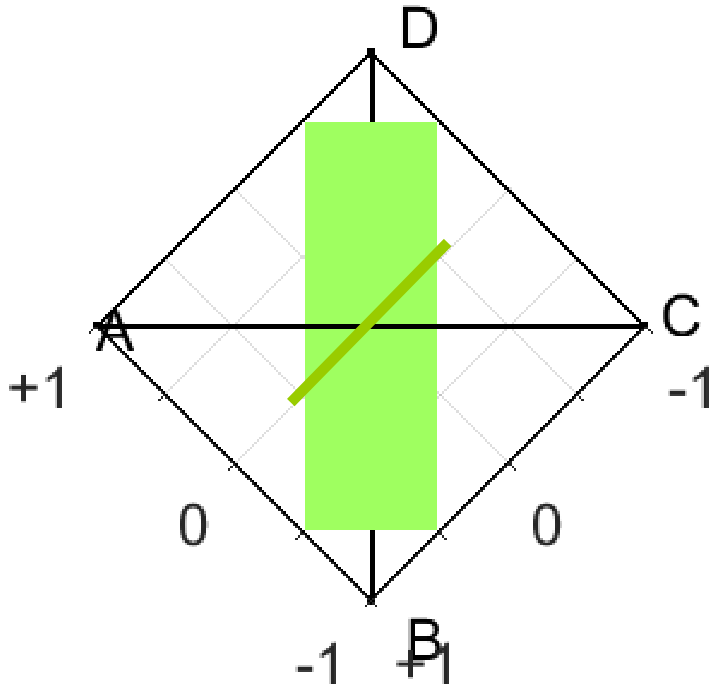
CA

przewaga klasy 0



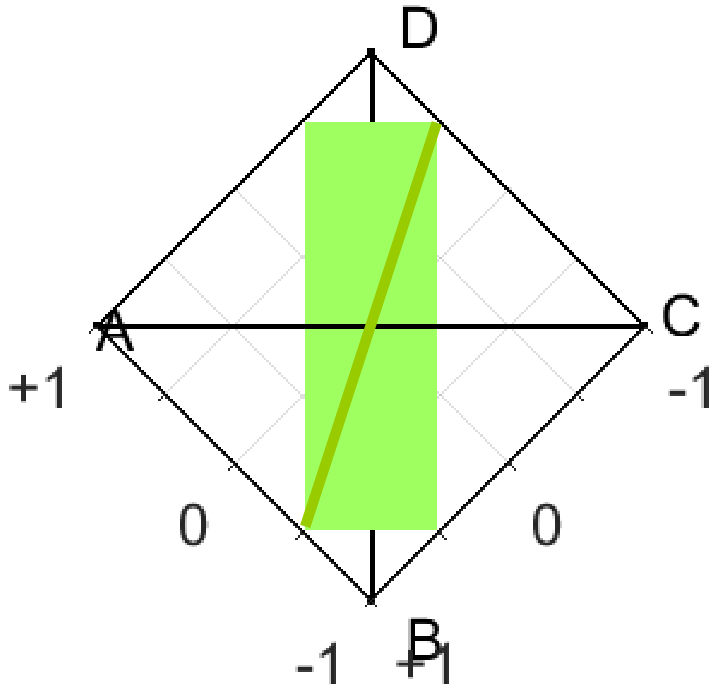
CA

przewaga klasy 0



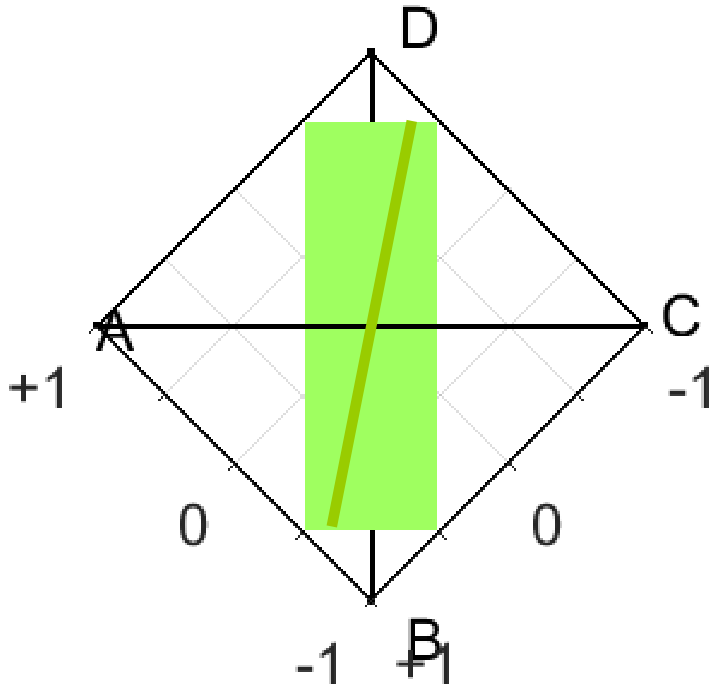
jak CA

przewaga klasy 0



lepiej? niż CA

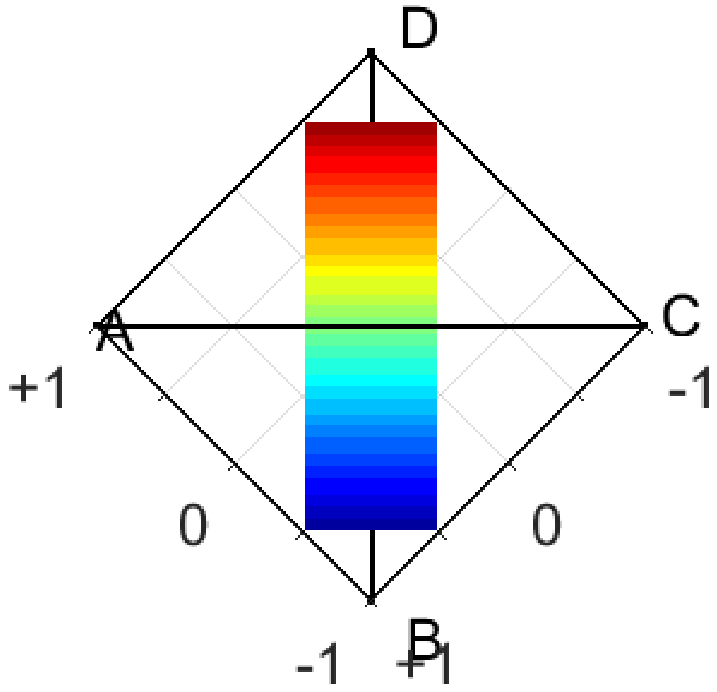
przewaga klasy 0



(przesadnie) lepiej? niż CA

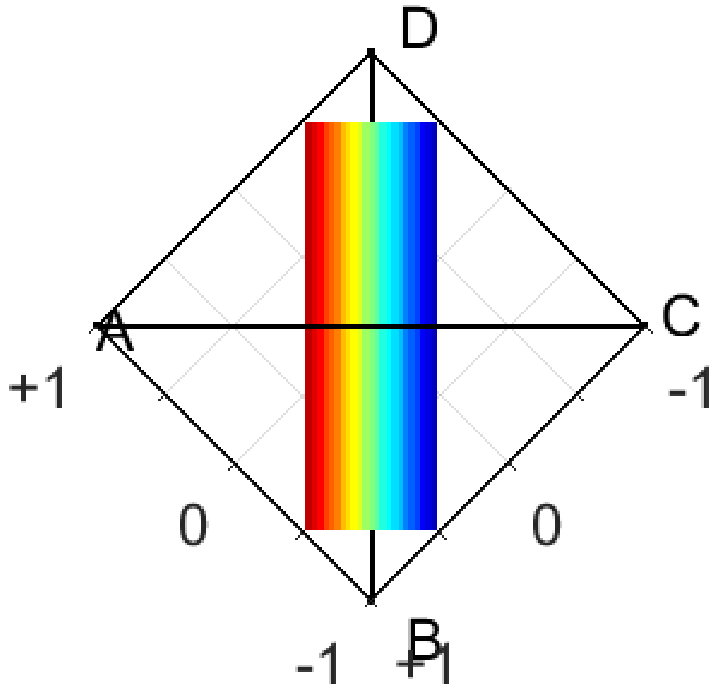
- inne miary:

przewaga klasy 0



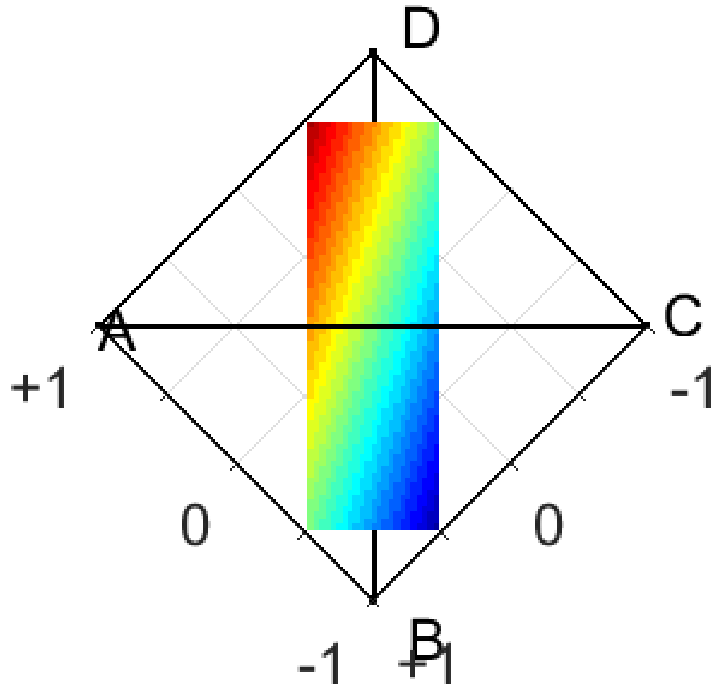
specificity

przewaga klasy 0



recall

przewaga klasy 0



amean(specificity,recall)

- kombinacje wypukłe
  - $p, q \geq 0$
  - $p+q = t$, gdzie $t > 0$
  - wtedy
    - $p/(p+q) \geq 0$, $q/(p+q) \geq 0$
    - $p/(p+q) + q/(p+q) = 1$
    - współczynniki kombinacji wypukłej
    - ułamki $p/(p+q)$ i $q/(p+q)$ sterują proporcjami p i q
      - $p/(p+q) = 0$ oznacza, że $p = 0$, $q = t$
      - $p/(p+q) = 1/2$ oznacza, że $p = t/2$, $q = 2/t$
      - $p/(p+q) = 1$ oznacza, że $p = t$, $q = 0$
      - (analogicznie dla $q/(p+q)$

- warunki dot. a,b,c,d
  - $a, b, c, d \geq 0$
  - $a+b+c+d = 1$

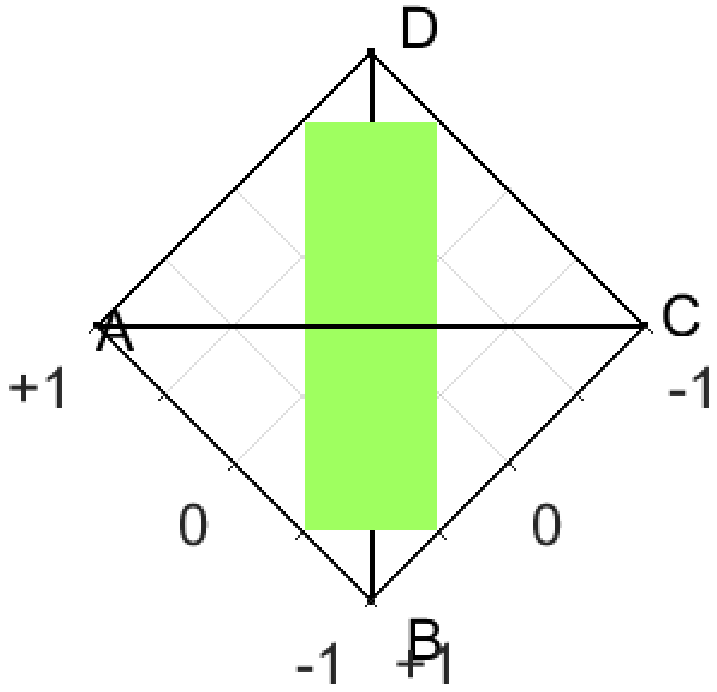
- przypadki dotyczące proporcji klas
  - a+c = p
  - b+d = q

  gdzie
    - p, q $\geq$ 0
    - p+q = 1 > 0

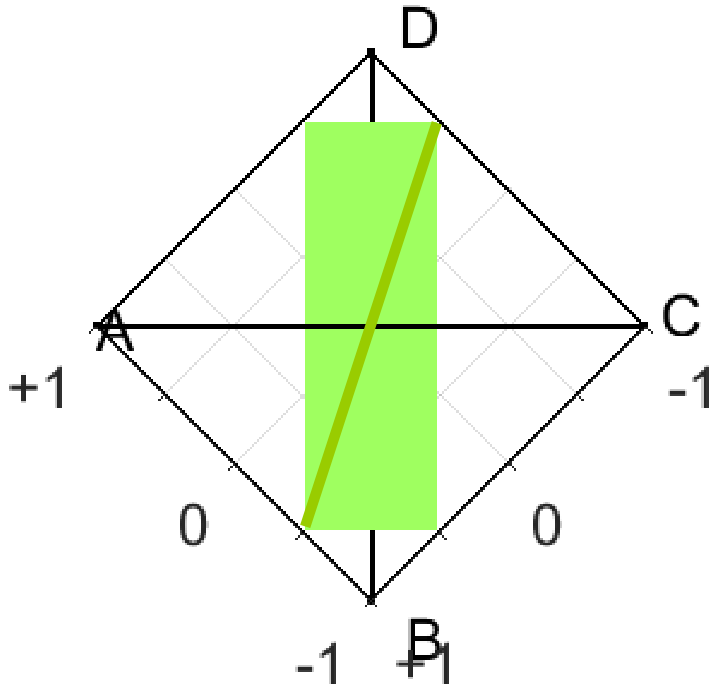
- wtedy
  - ułamki (a+c)/(a+b+c+d) i (b+d)/(a+b+c+d) sterują proporcjami klas
  - dla (a+c)/(a+b+c+d) = p mamy (b+d)/(a+b+c+d) = 1 – p = q

- dla ustalonych a+c = p i b+d = q zmianom mogą podlegać
  - a i c (w ramach sumy a+c = p)
    - współczynniki kombinacji wypukłej
    - ułamki a/(a+c) i c/(a+c) sterują proporcjami a i c
      - a/(a+c) = recall
      - c/(a+c) = 1 – recall
  - b i d (w ramach sumy b+d = q)
    - współczynniki kombinacji wypukłej
    - ułamki b/(b+d) i d/(b+d) sterują proporcjami b i d
      - b/(b+d) = 1 – specificity
      - d/(b+d) = specificity

przewaga klasy 0

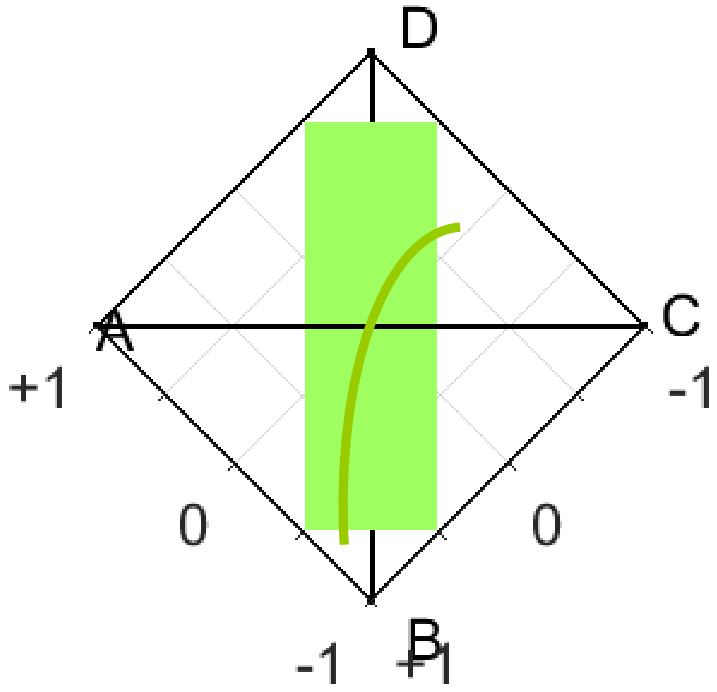


przewaga klasy 0



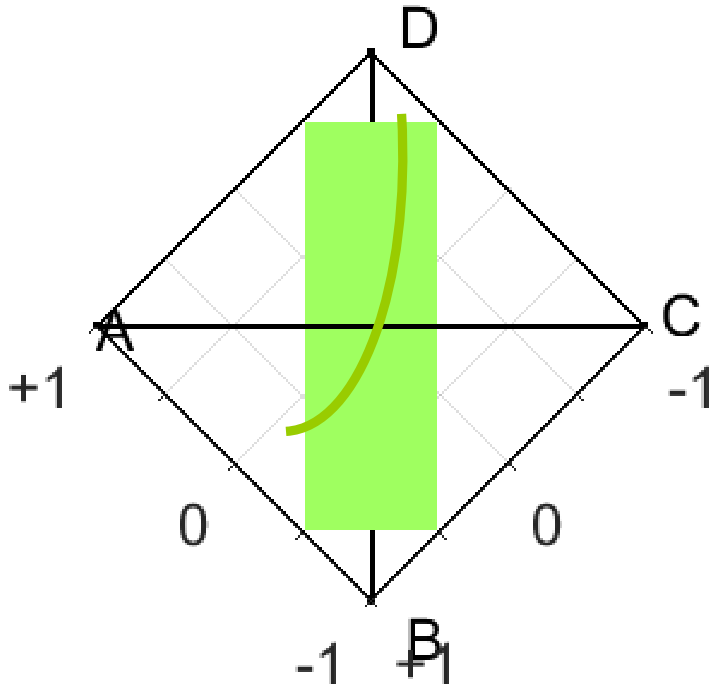
agregacja typu amean

przewaga klasy 0



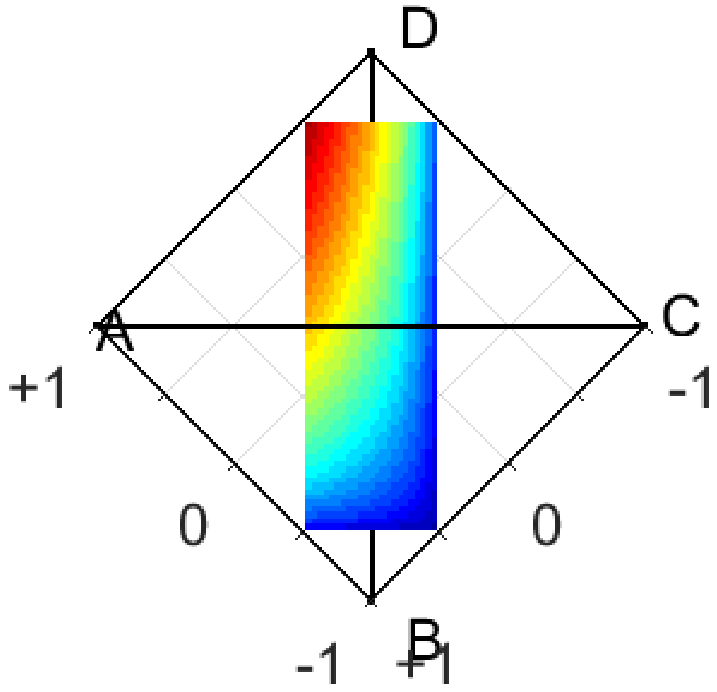
agregacja typu max

przewaga klasy 0



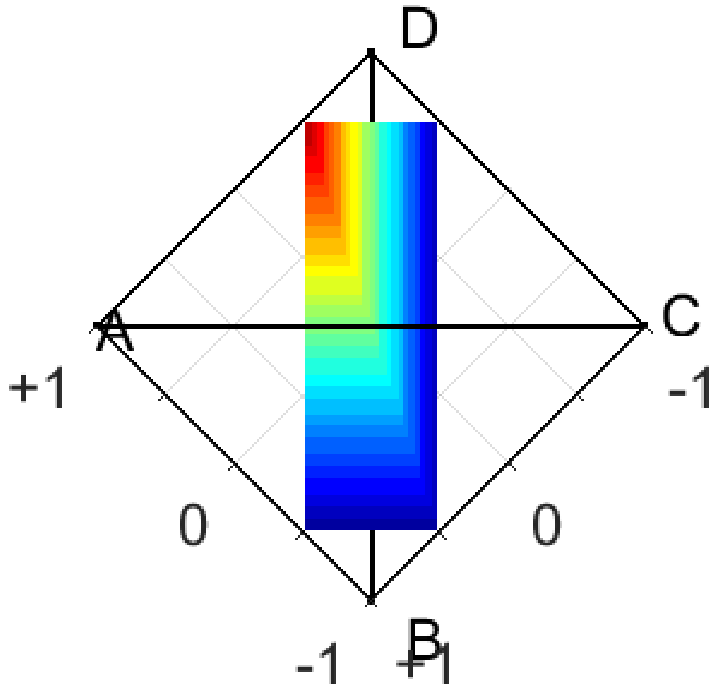
agregacja typu min
(także gmean/hmean)

przewaga klasy 0



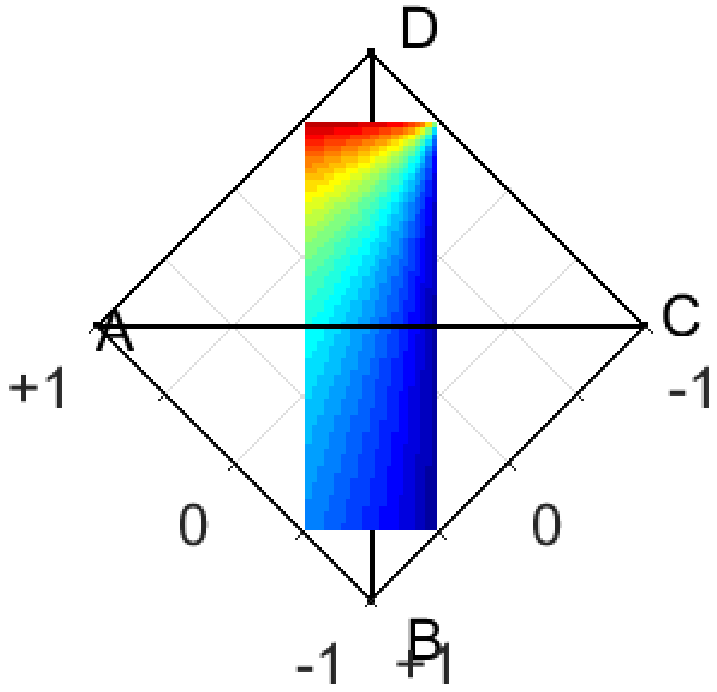
G-mean

przewaga klasy 0



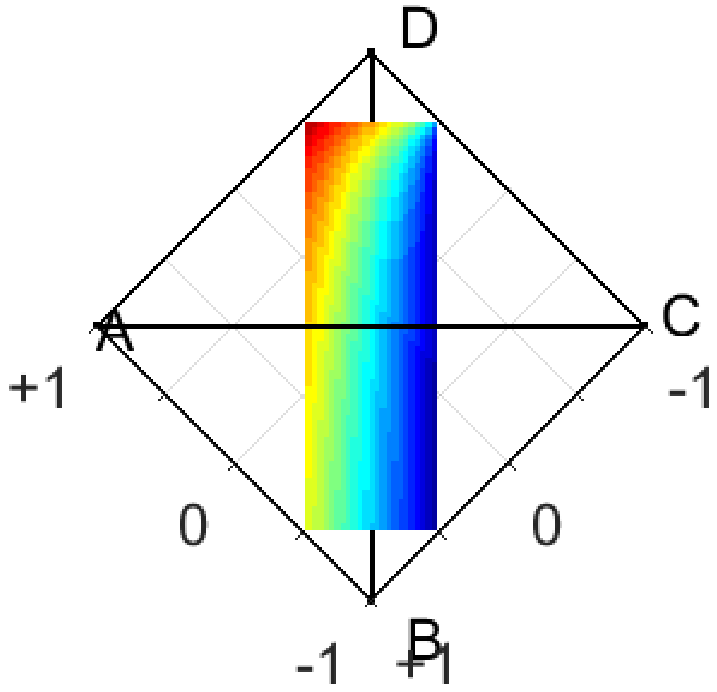
min(specificity,recall)

przewaga klasy 0



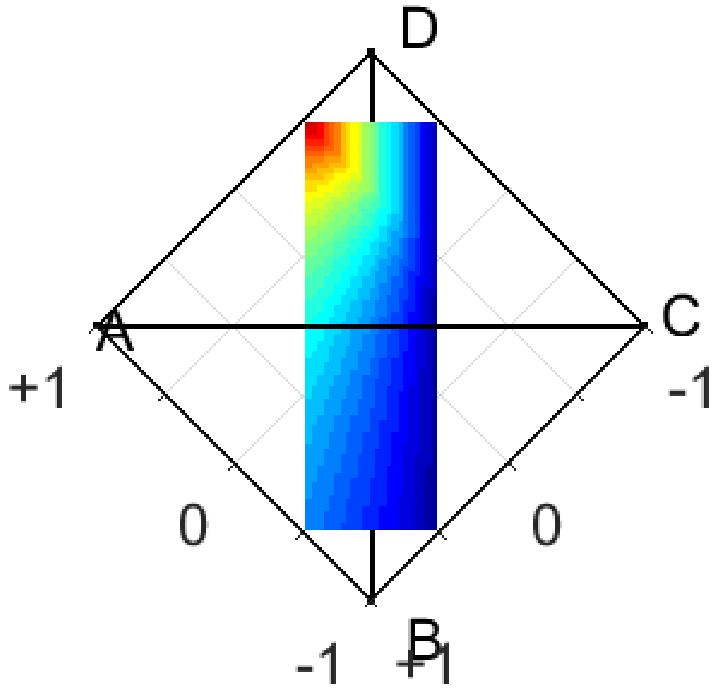
precision

przewaga klasy 0



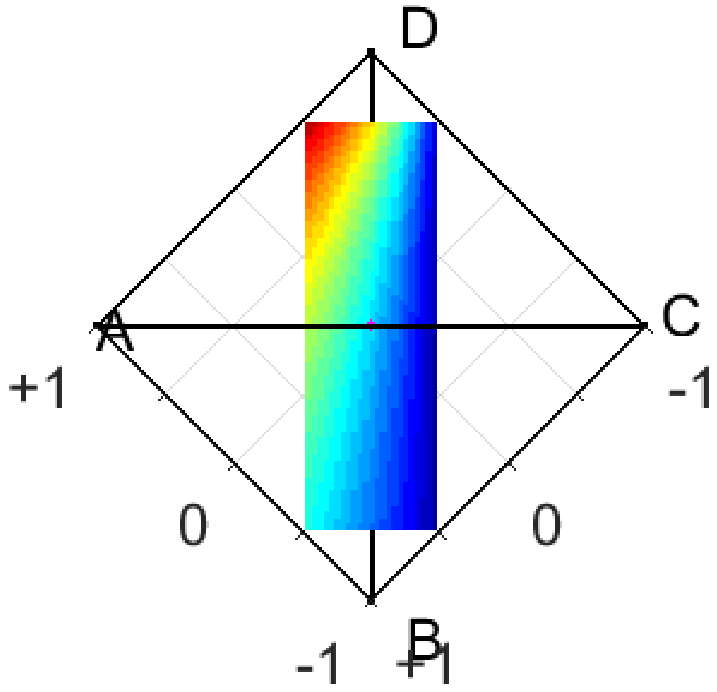
amean(precision,recall)

przewaga klasy 0



min(precision,recall)

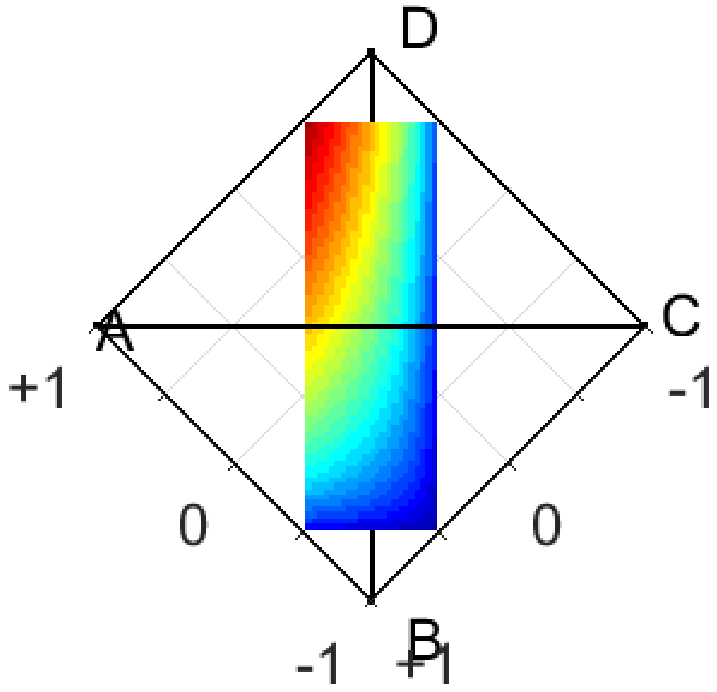
przewaga klasy 0



F_1

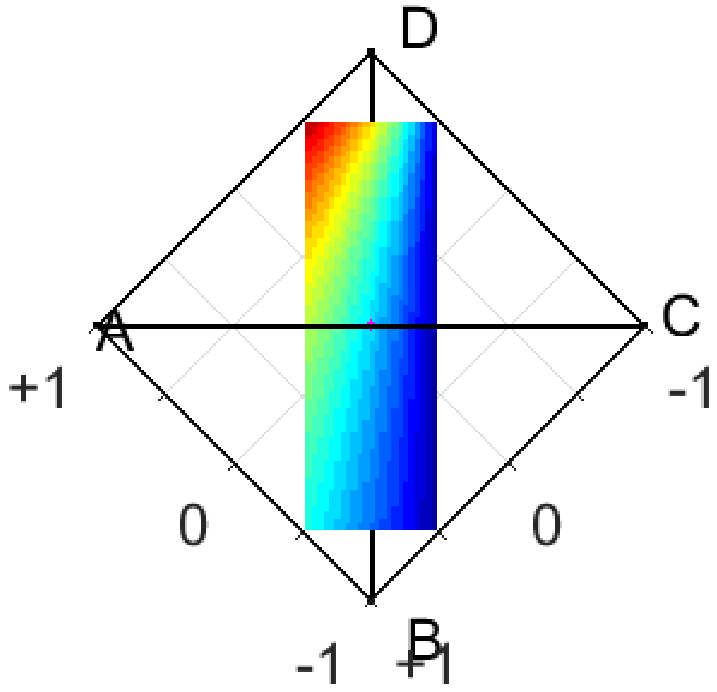
- F_1 vs G-mean dla rosnącej nierównowagi

przewaga klasy 0



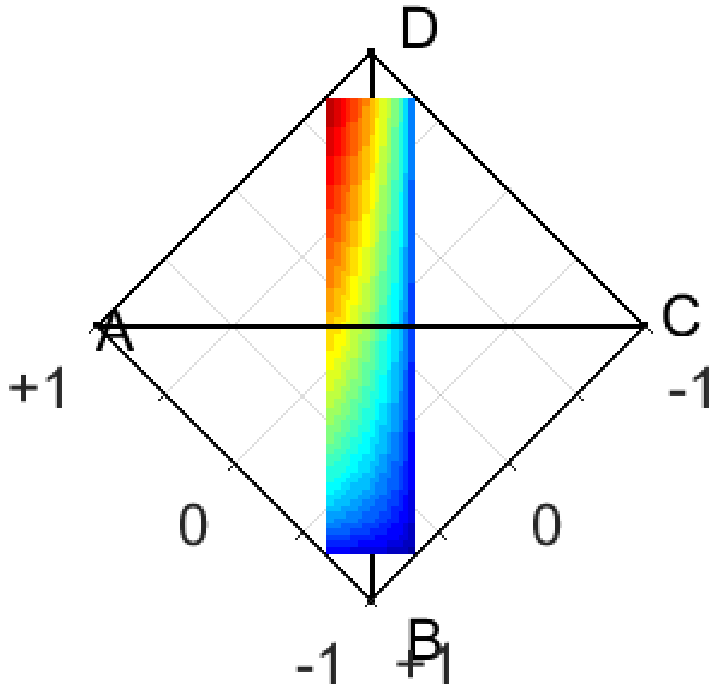
G-mean

przewaga klasy 0



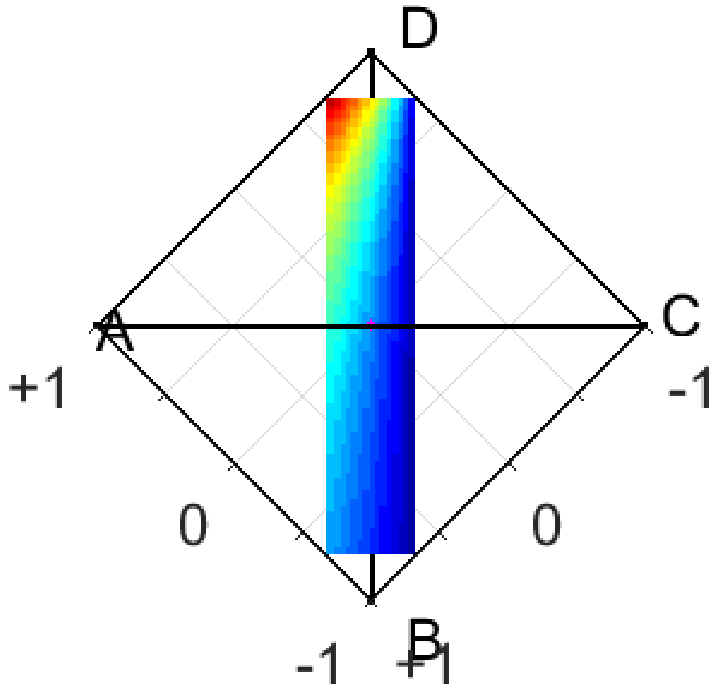
F_1

przewaga klasy 0



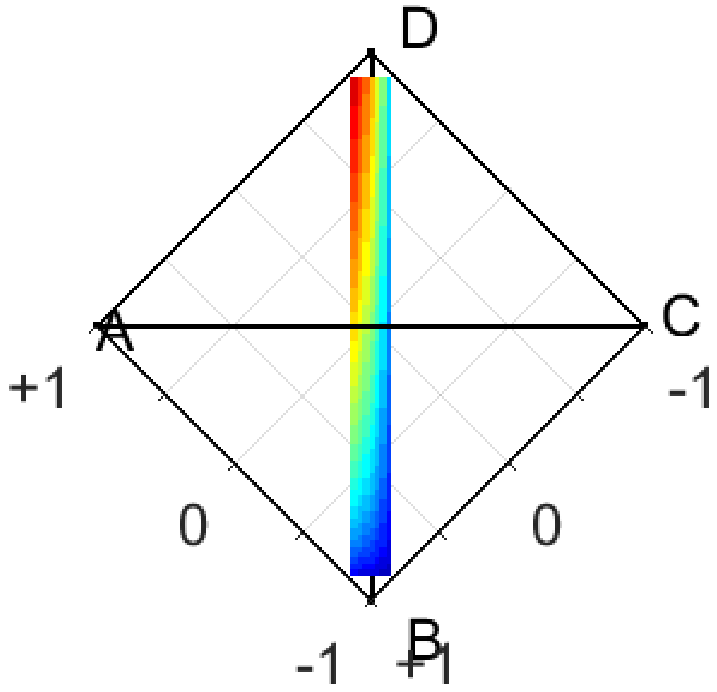
G-mean

przewaga klasy 0



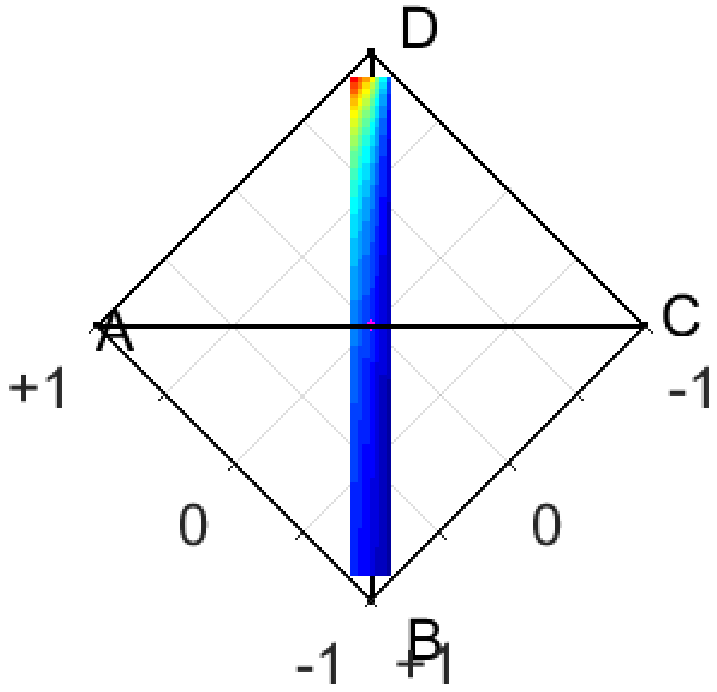
F_1

przewaga klasy 0



G-mean

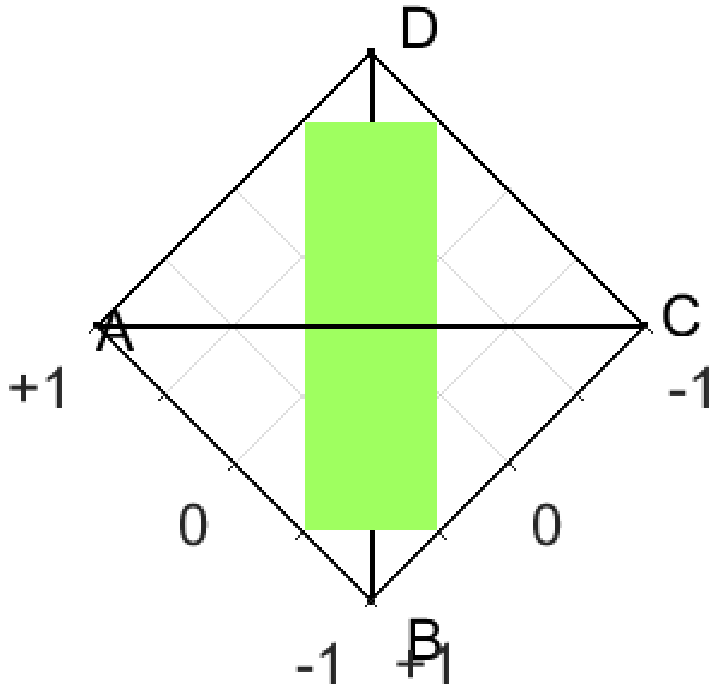
przewaga klasy 0



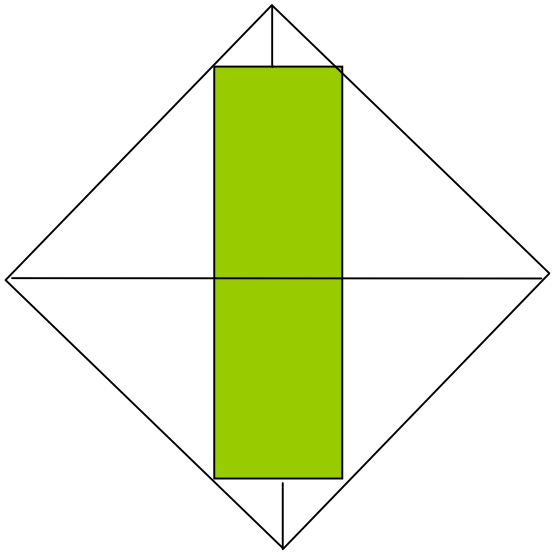
F_1

...

widok z góry



D

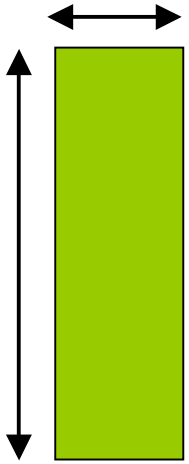


A

C

B

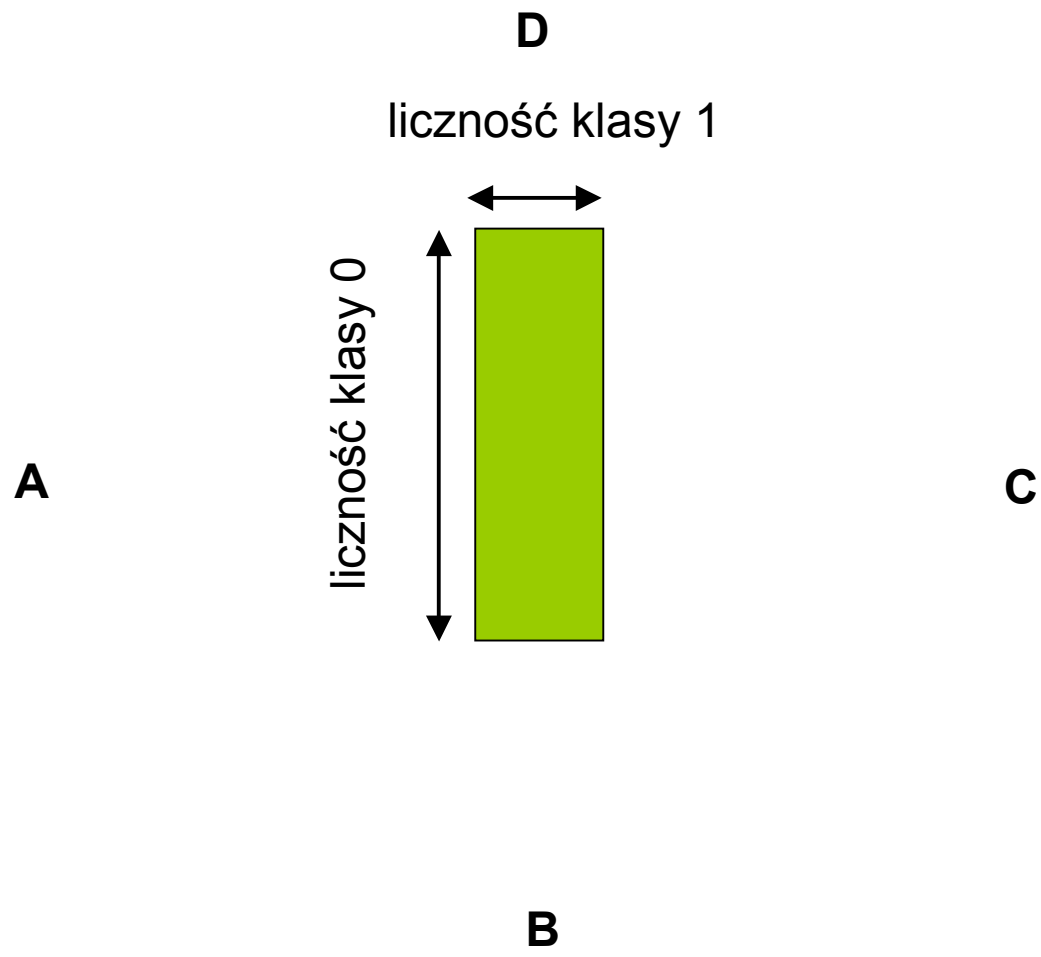
D

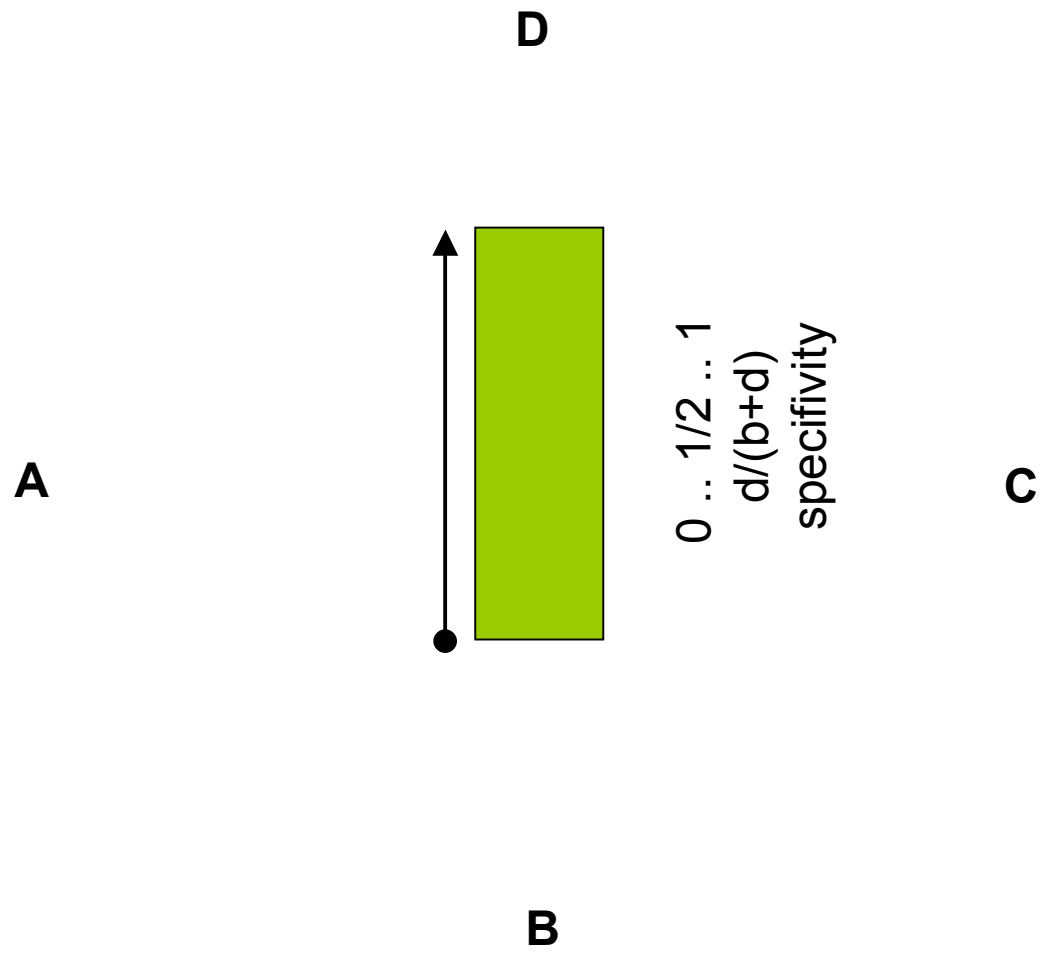


A

C

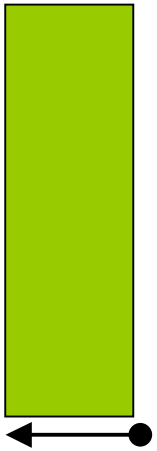
B


D
liczność klasy 1
liczność klasy 0
A
C
B


D
A
0 .. 1/2 .. 1
d/(b+d)
specifivity
C
B

D

A



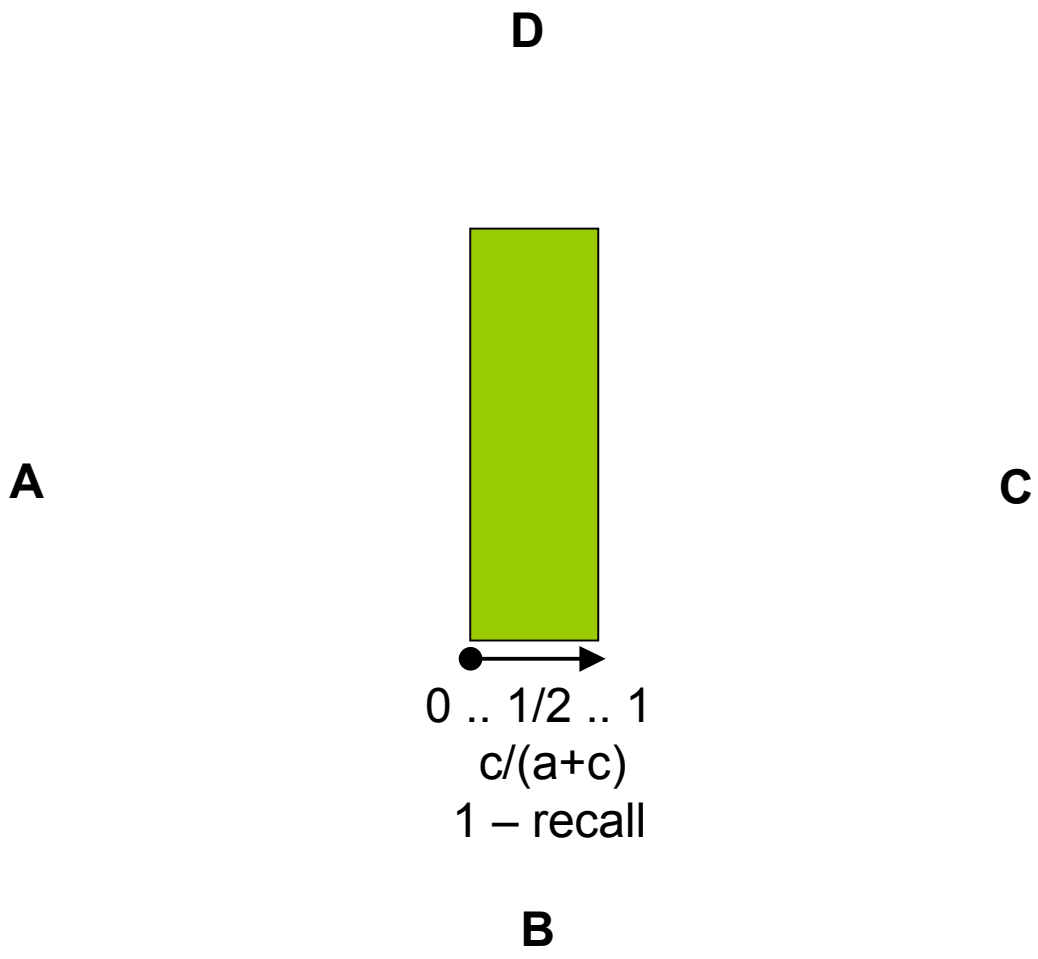
C

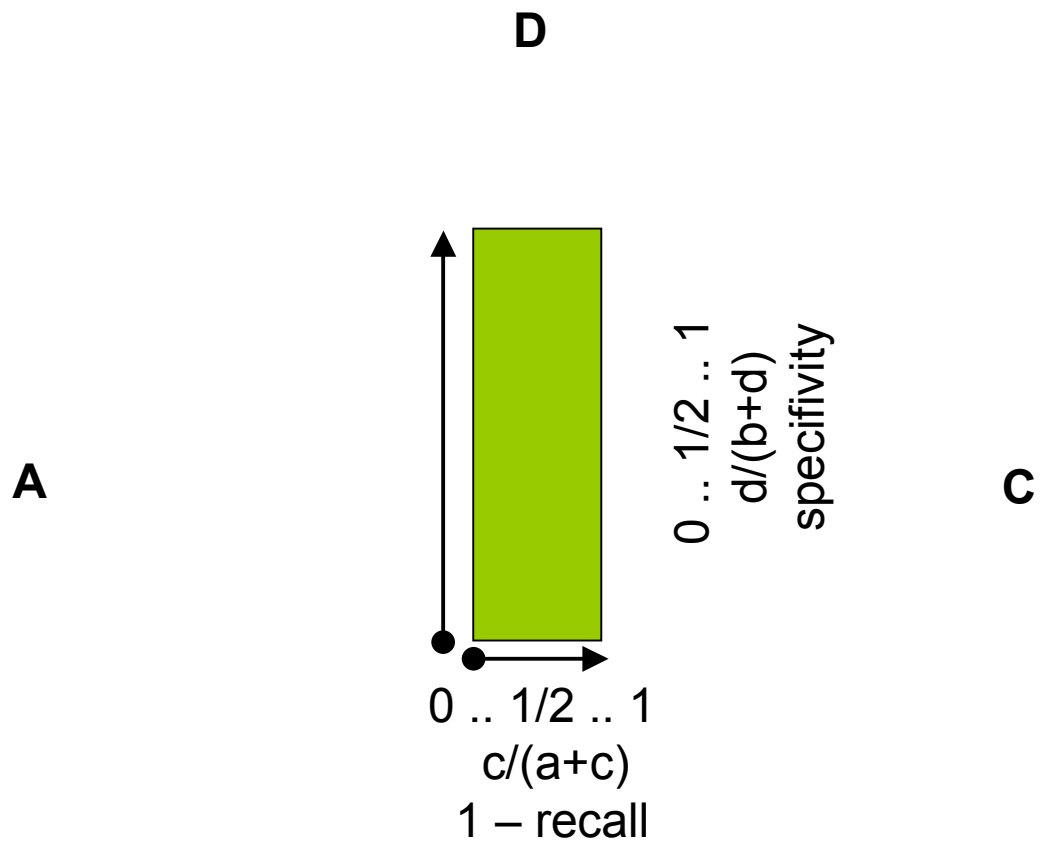
1 .. 1/2 .. 0

a/(a+c)

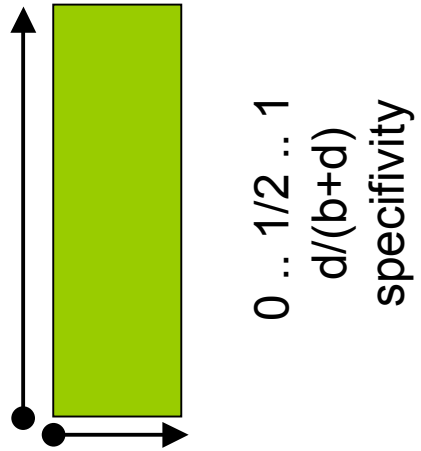
recall

B

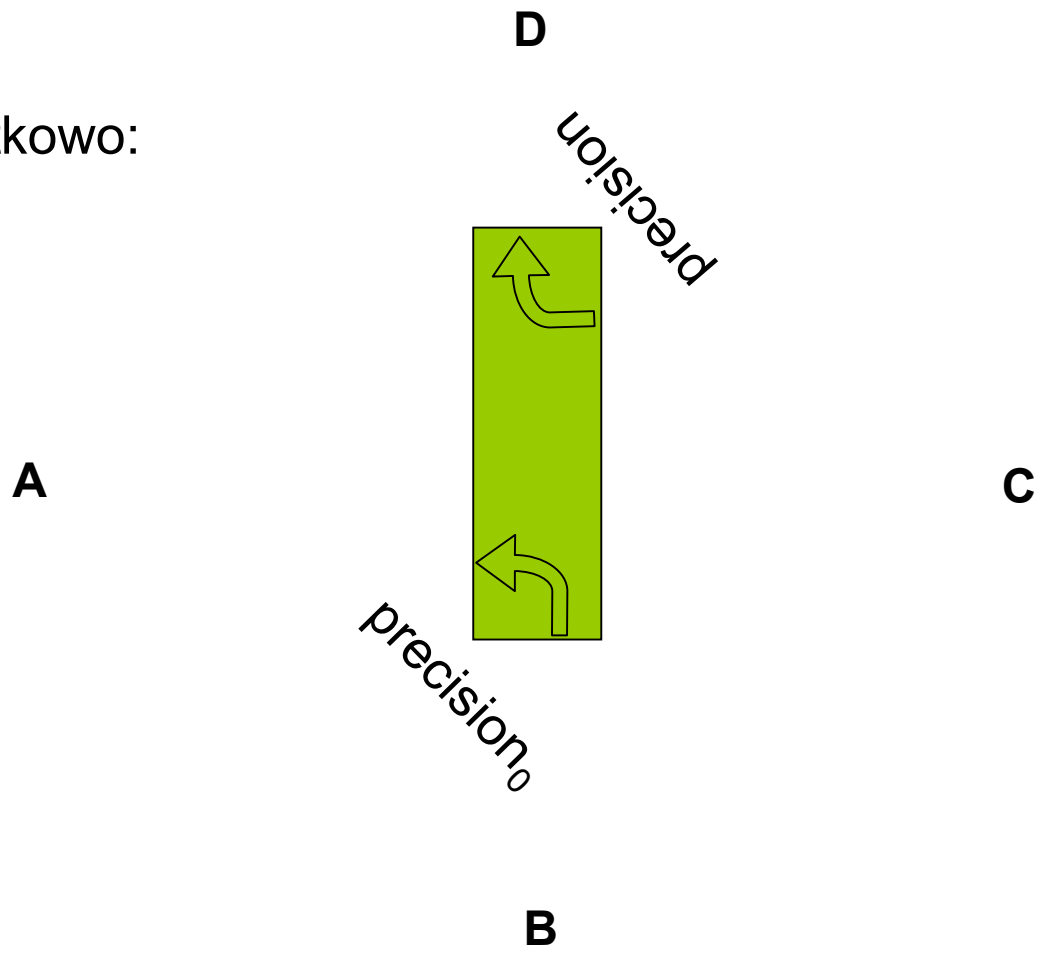

D
A
C
0 .. 1/2 .. 1
c/(a+c)
1 – recall
B


D
A
C
B
0 .. 1/2 .. 1
d/(b+d)
specifivity
0 .. 1/2 .. 1
c/(a+c)
1 – recall

(1 – recall) vs specificity
(układ współ. jak dla AUC)

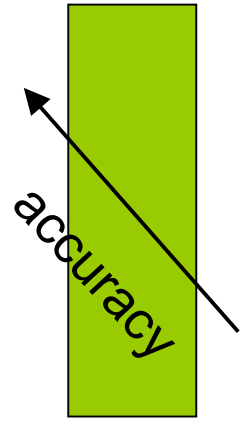


dodatkowo:



D

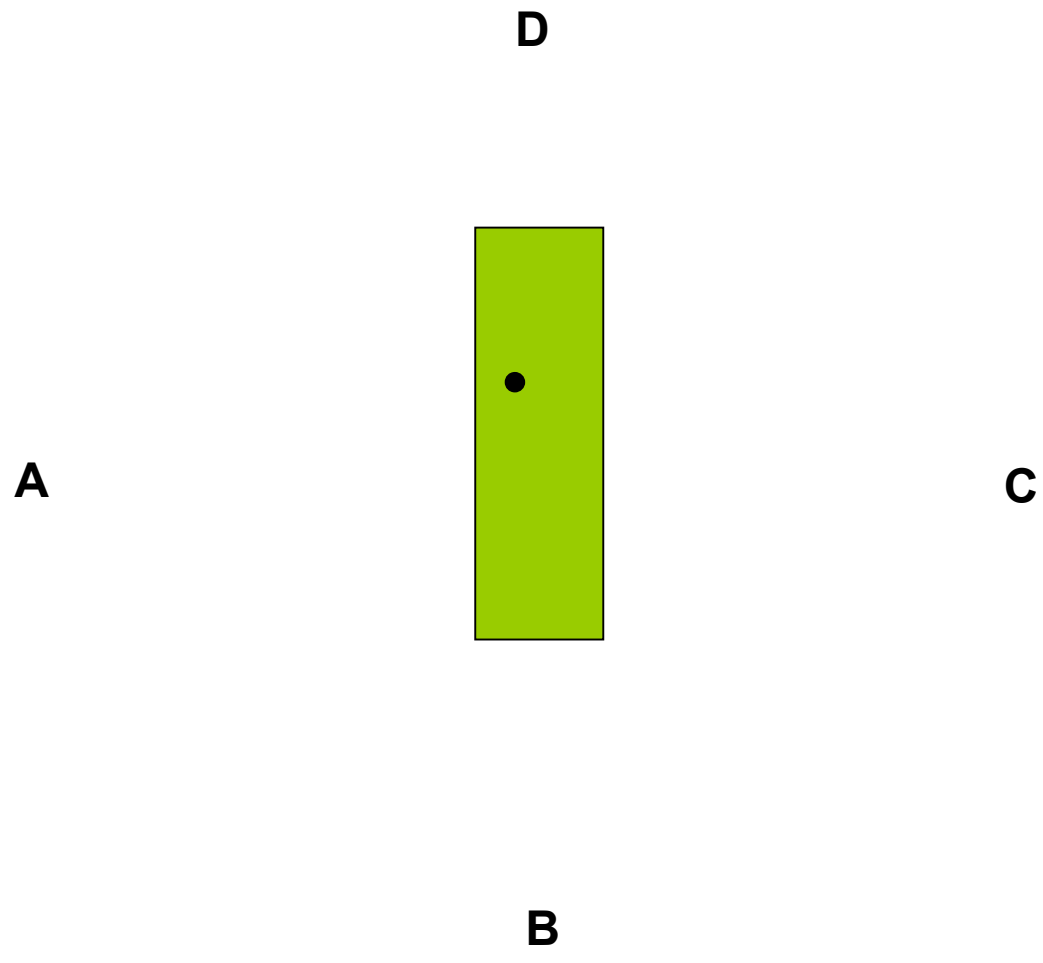
dodatkowo:

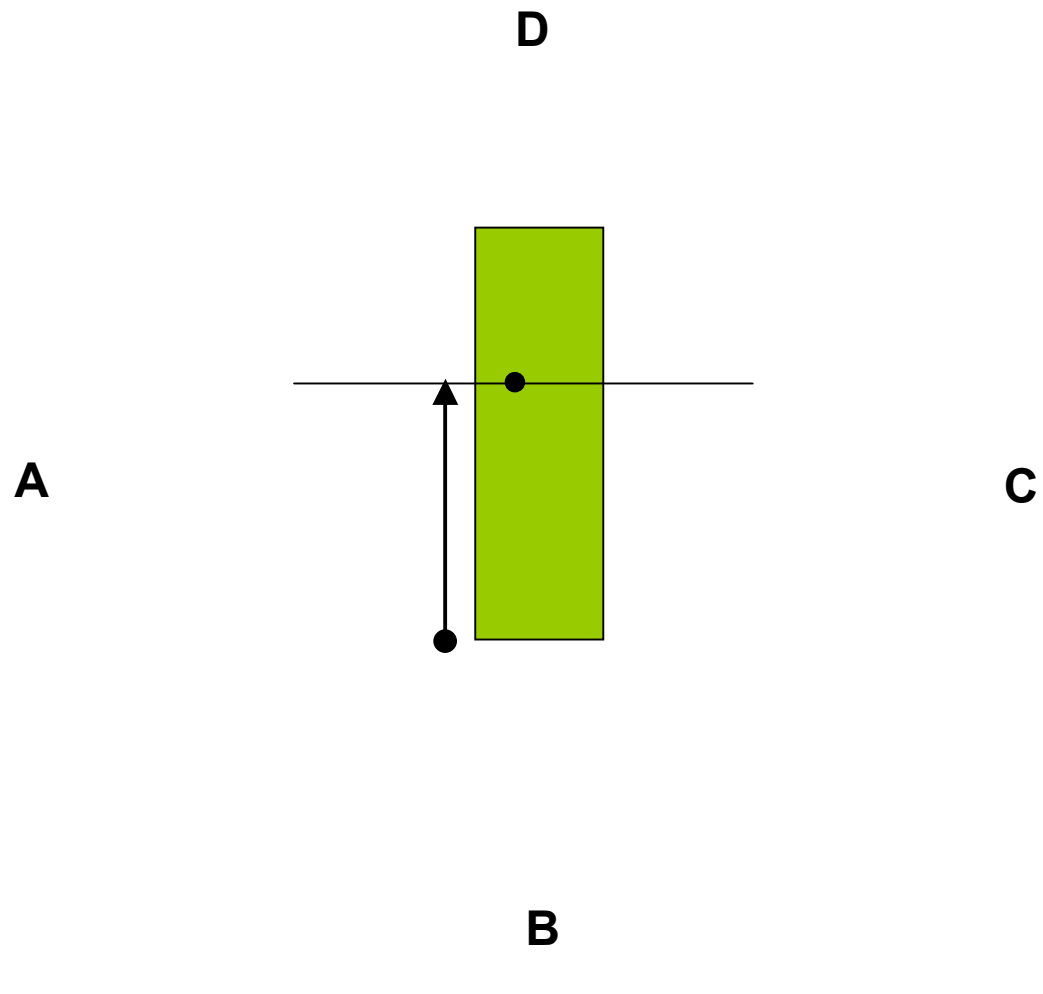


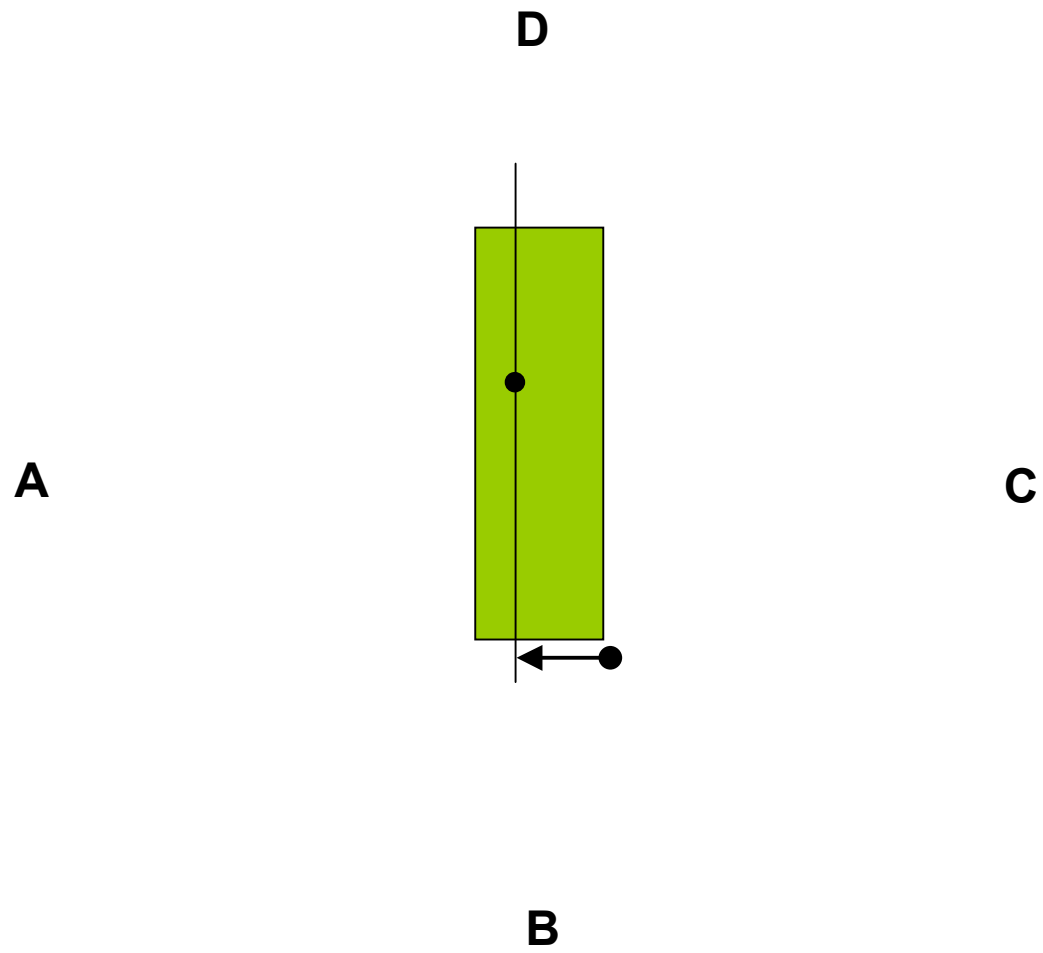
A

C

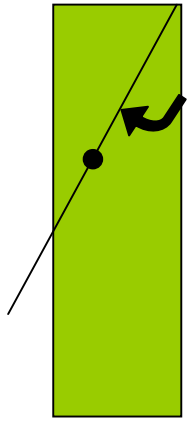
B


D
A
C
B


D
A
C
B


D
A
C
B

D

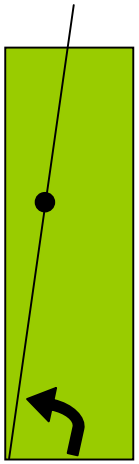


A

C

B

D

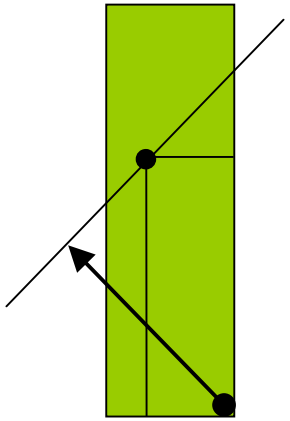


A

C

B

D



A

C

B

...