GDAŃSK UNIVERSITY OF TECHNOLOGY

Eksploracja danych

OCENA KLASYFIKATORÓW

Wojciech Waloszek

wowal@eti.pg.gda.pl

Teresa Zawadzka

tegra@eti.pg.gda.pl

Katedra Inżynierii Oprogramowania
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

GDAŃSK UNIVERSITY OF TECHNOLOGY

Ocena wyjścia

- Omówione metody budują klasyfikatory, które osiągają bardzo dobre wyniki na *zbiorze uczącym*,
- Często okazuje się, że klasyfikator osiągający dobry wynik na zbiorze uczącym ma bardzo słabe właściwości predykcyjne,
- Zjawisko to nazywamy *błędem nadmiernego dopasowania.*

GDAŃSK UNIVERSITY OF TECHNOLOGY

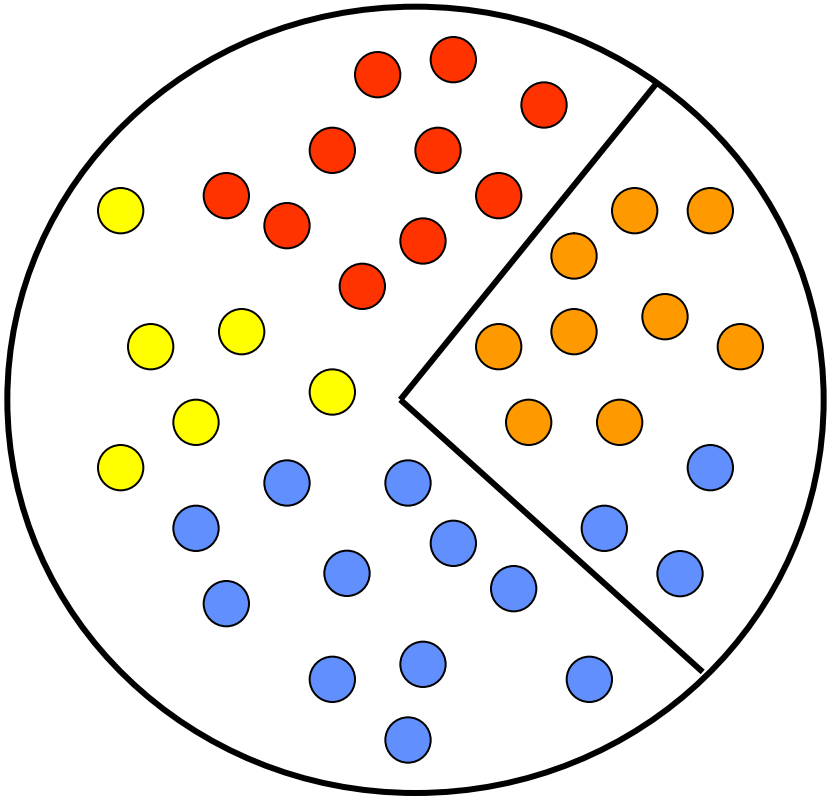
Błąd nadmiernego dopasowania

Nazwisko	Wiek	Wykształcenie	Sam.	Z.K.
Abacki	32	wyższe	tak	tak
Babacka	35	średnie	tak	tak
Cabacki	26	podstawowe	nie	nie
Dabacka	45	wyższe	nie	tak
Ebacki	38	średnie	tak	tak
Fabacki	28	wyższe	nie	nie
Gabacka	65	średnie	tak	nie
Habacka	22	średnie	nie	nie
Ibacki	43	podstawowe	tak	nie

```
if Nazwisko=Abacki then Z.K.=tak
if Nazwisko=Babacka then Z.K.=tak
if Nazwisko=Cabacki then Z.K.=nie
  ...
```

GDAŃSK UNIVERSITY OF TECHNOLOGY

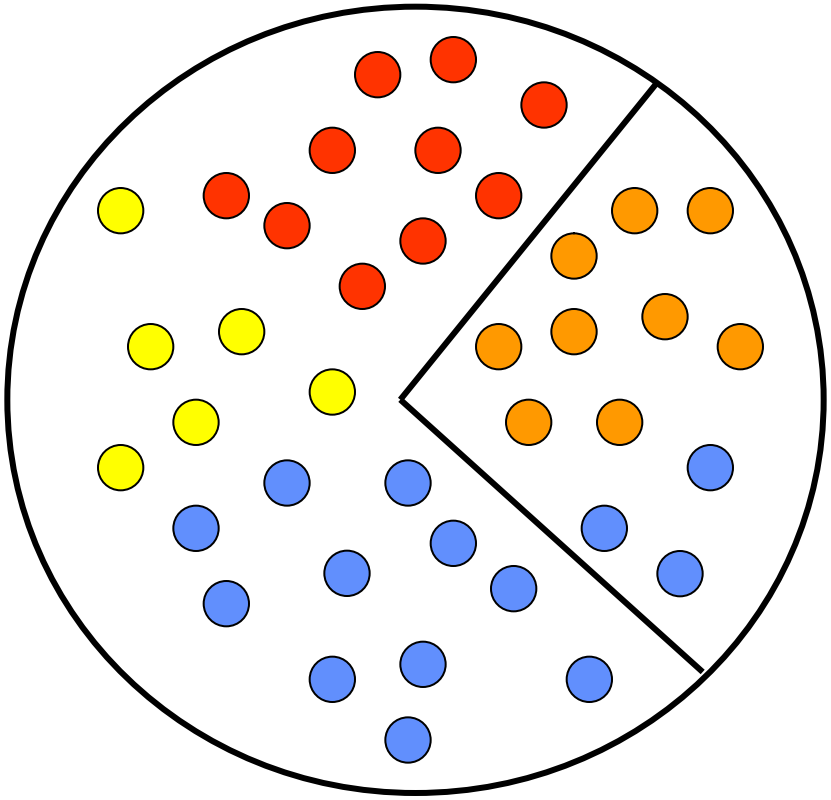
Zbiór uczący i weryfikujący



- Standardowe podejście do oceny klasyfikatora to podział zbioru przykładów (zbioru uczącego) na *trenujący* i *testujący*

GDAŃSK UNIVERSITY OF TECHNOLOGY

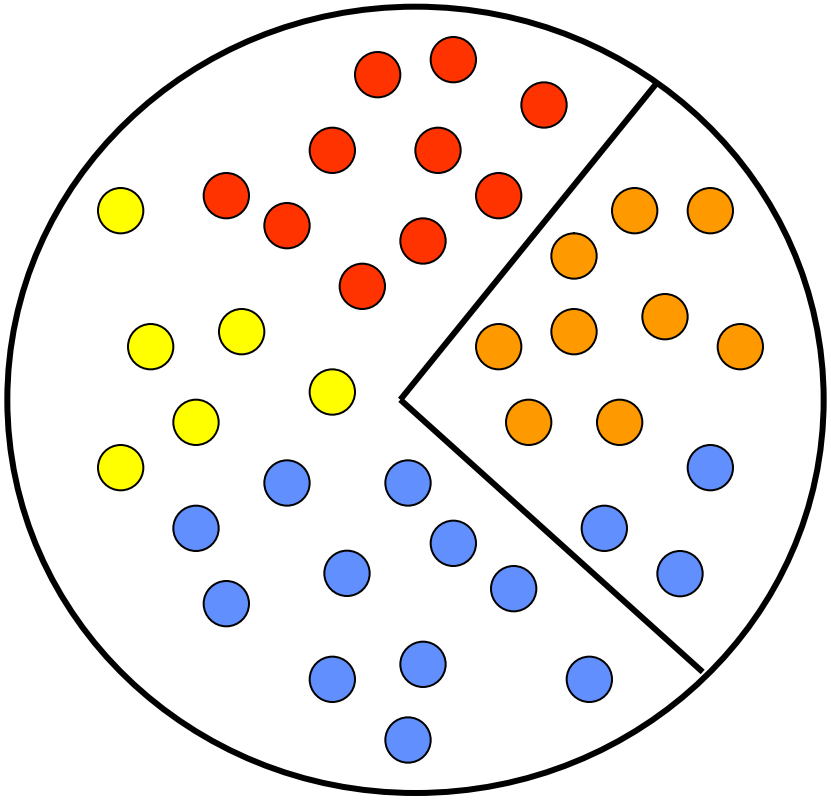
Zbiór uczący i weryfikujący (2)



- Klasyfikator budujemy na podstawie zbioru trenującego, a następnie badamy jego skuteczność na zbiorze testującym

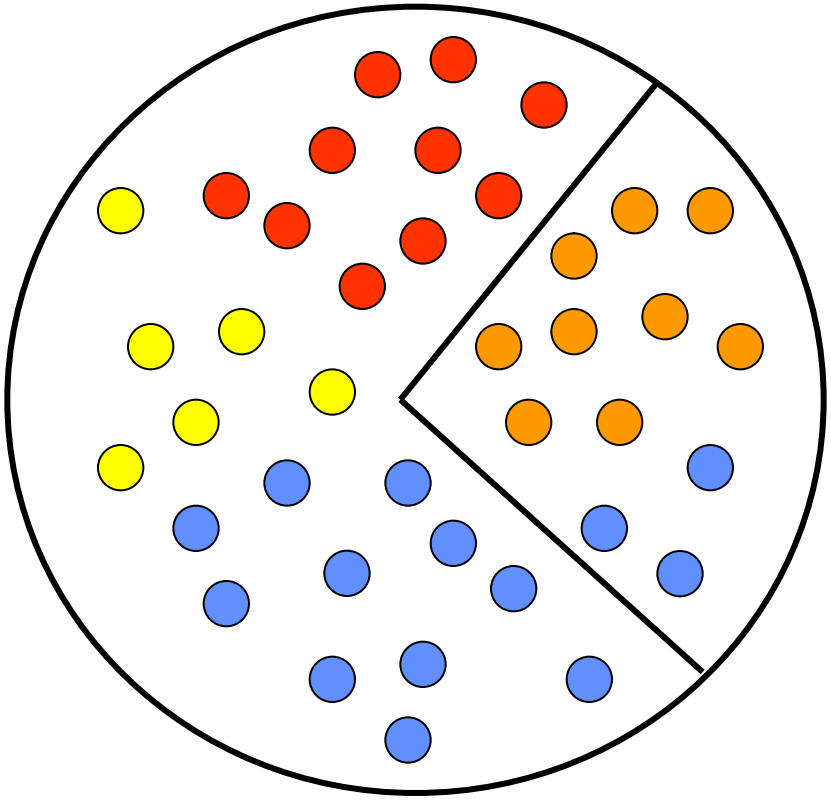
GDAŃSK UNIVERSITY OF TECHNOLOGY

Zbiór uczący i weryfikujący (3)



- Podejście to najczęściej stosujemy tylko do oceny siły predykcyjnej, a właściwy klasyfikator budujemy używając całego zbioru przykładów.

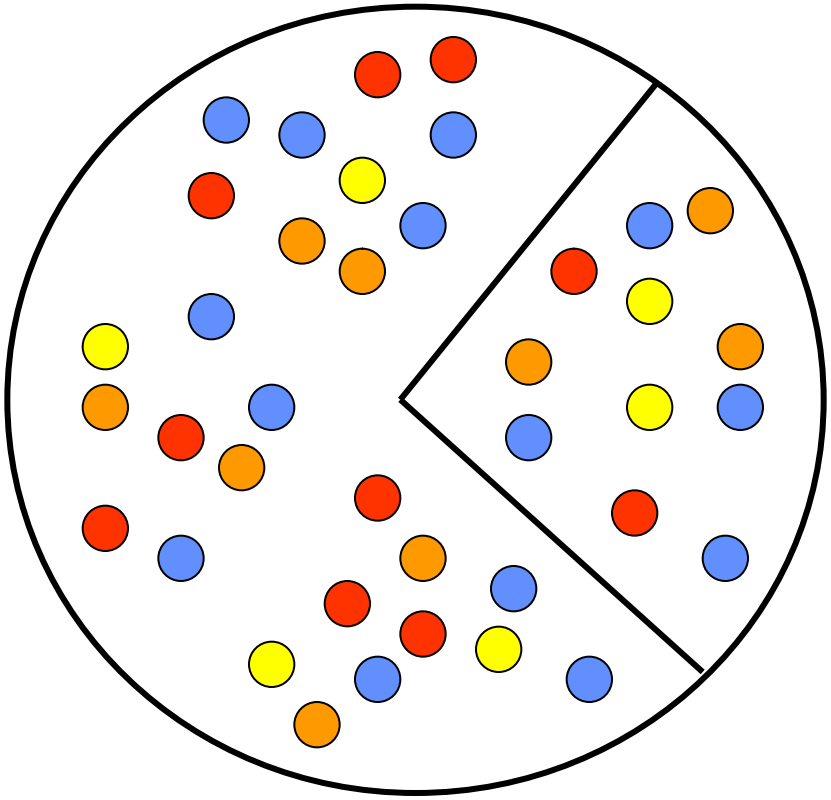
Stratyfikacja



- Przy wyborze zbioru trenującego istnieje niebezpieczeństwo nierównomiernej reprezentacji niektórych klas,
- Temu efektowi zapobiega ***stratyfikacja***.

GDAŃSK UNIVERSITY OF TECHNOLOGY

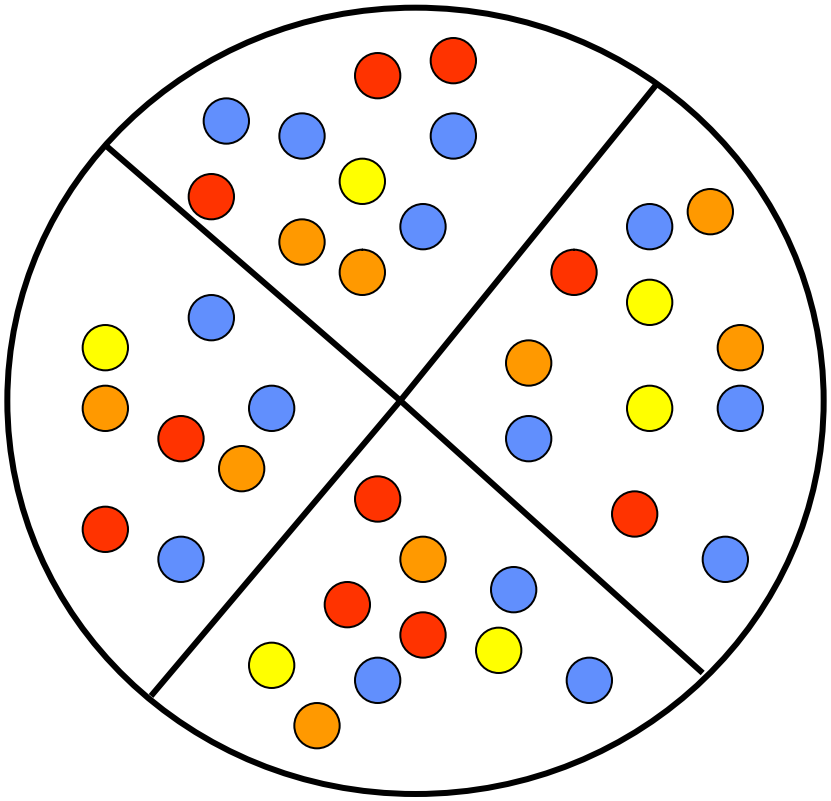
Efekt stratyfikacji



- Stratyfikacja polega na dokonywaniu takich podziałów zbioru przykładów, aby klasy były równomiernie reprezentowane w każdej jego części

GDAŃSK UNIVERSITY OF TECHNOLOGY

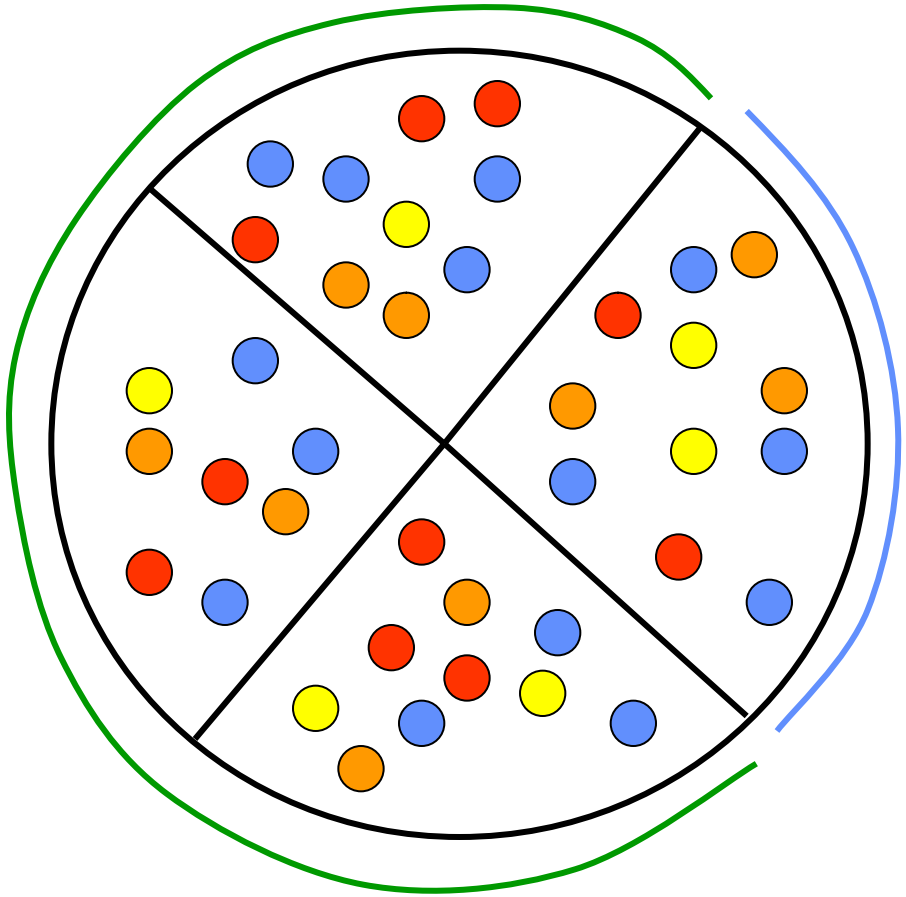
Podział zbioru na *n* części



- Pomimo dokonania stratyfikacji, wciąż możemy mieć wątpliwości co do wyboru zbioru testującego,
- Odpowiedzią może być podział zbioru na *n* części

GDAŃSK UNIVERSITY OF TECHNOLOGY

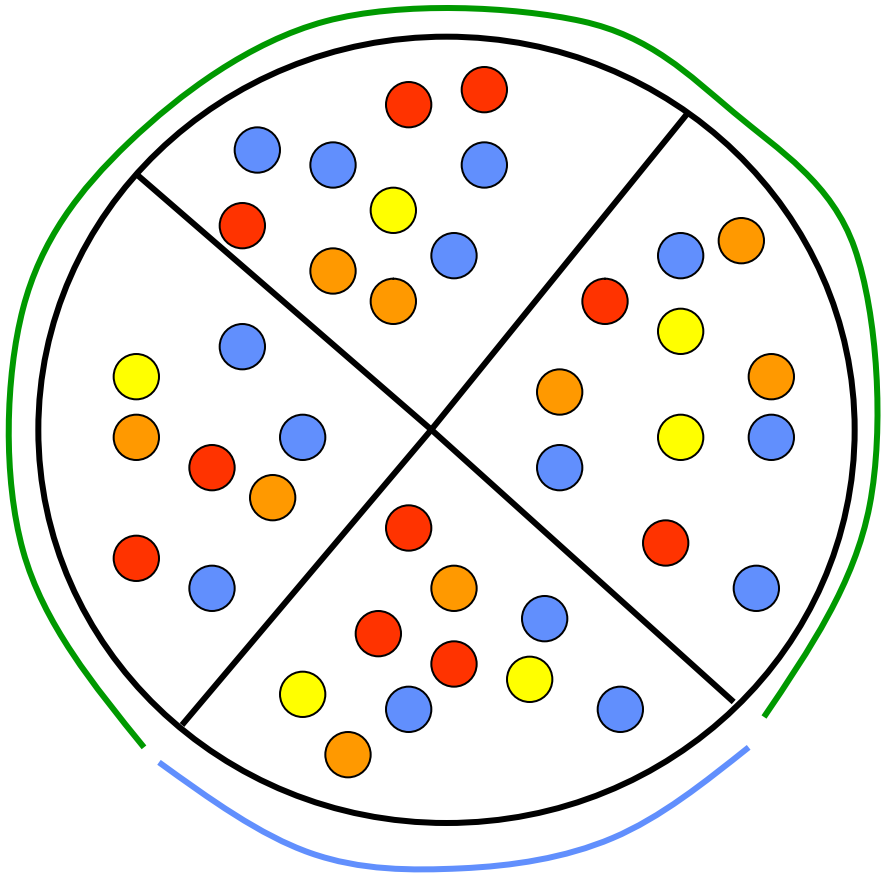
Podział zbioru na *n* części (2)



- Przy zastosowaniu tego podejścia proces uczenia powtarzamy *n* razy,
- Przy każdym powtórzeniu jedną część zbioru (tu: ¼) przyjmujemy za zbiór testujący, resztę jako trenujący (tu: ¾)

GDAŃSK UNIVERSITY OF TECHNOLOGY

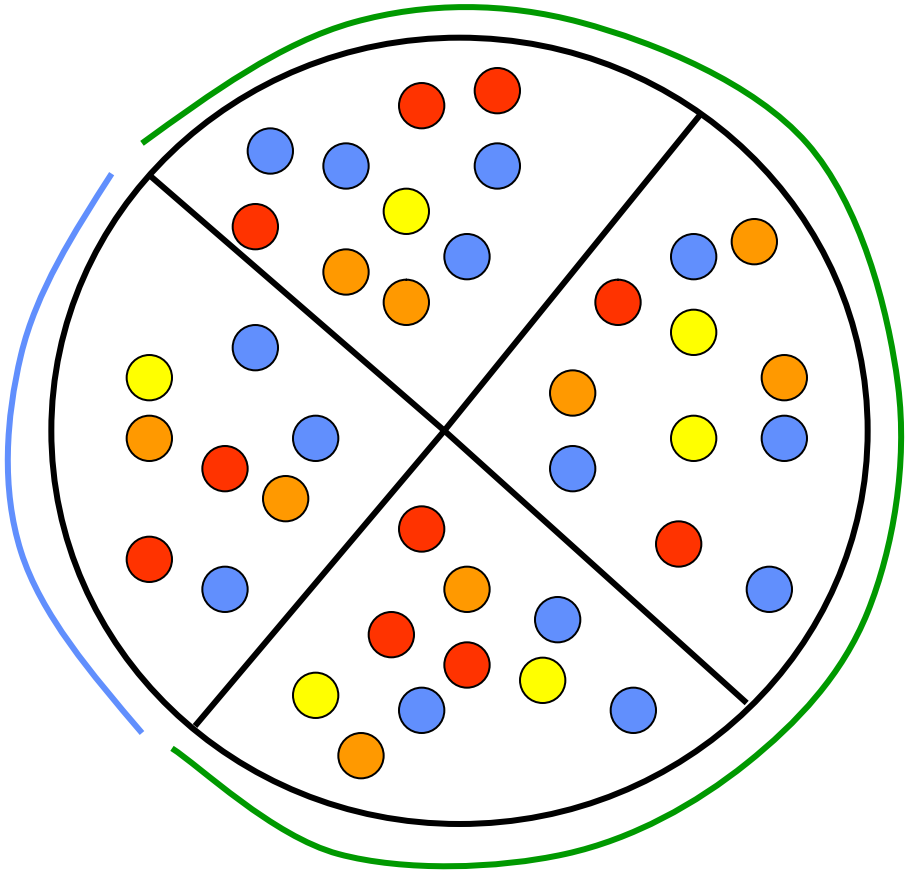
Podział zbioru na *n* części (2)



- Przy zastosowaniu tego podejścia proces uczenia powtarzamy *n* razy,
- Przy każdym powtórzeniu jedną część zbioru (tu: ¼) przyjmujemy za zbiór testujący, resztę jako trenujący (tu: ¾)

GDAŃSK UNIVERSITY OF TECHNOLOGY

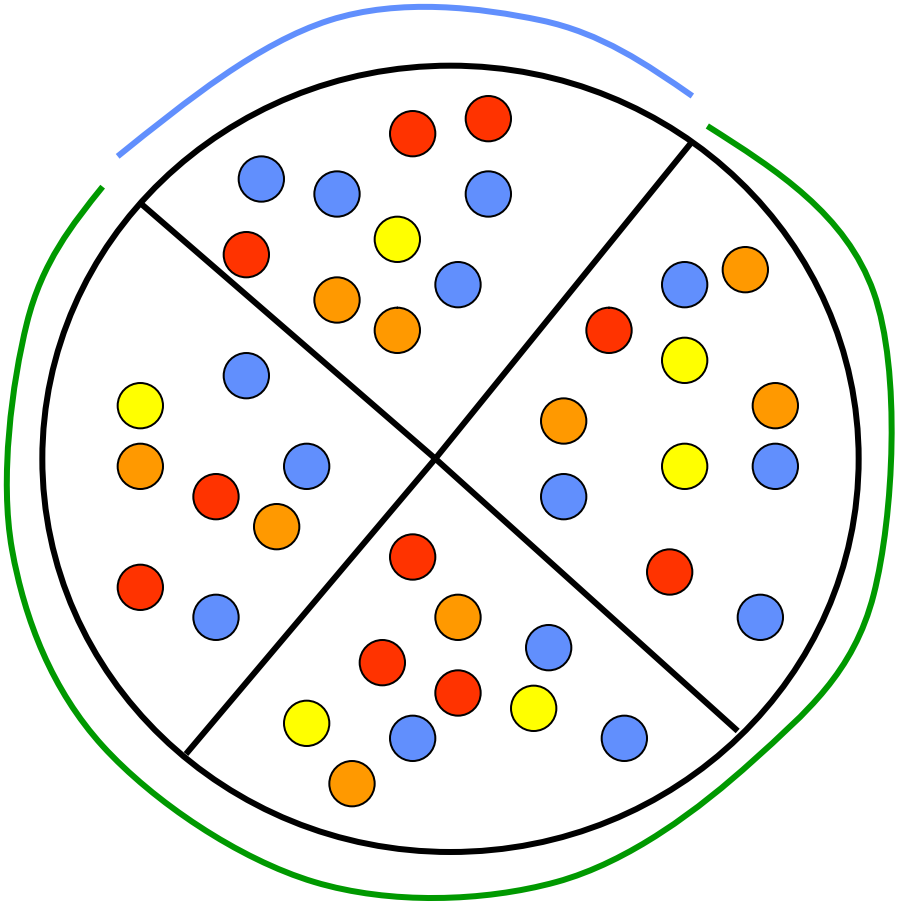
Podział zbioru na *n* części (2)



- Przy zastosowaniu tego podejścia proces uczenia powtarzamy *n* razy,
- Przy każdym powtórzeniu jedną część zbioru (tu: ¼) przyjmujemy za zbiór testujący, resztę jako trenujący (tu: ¾)

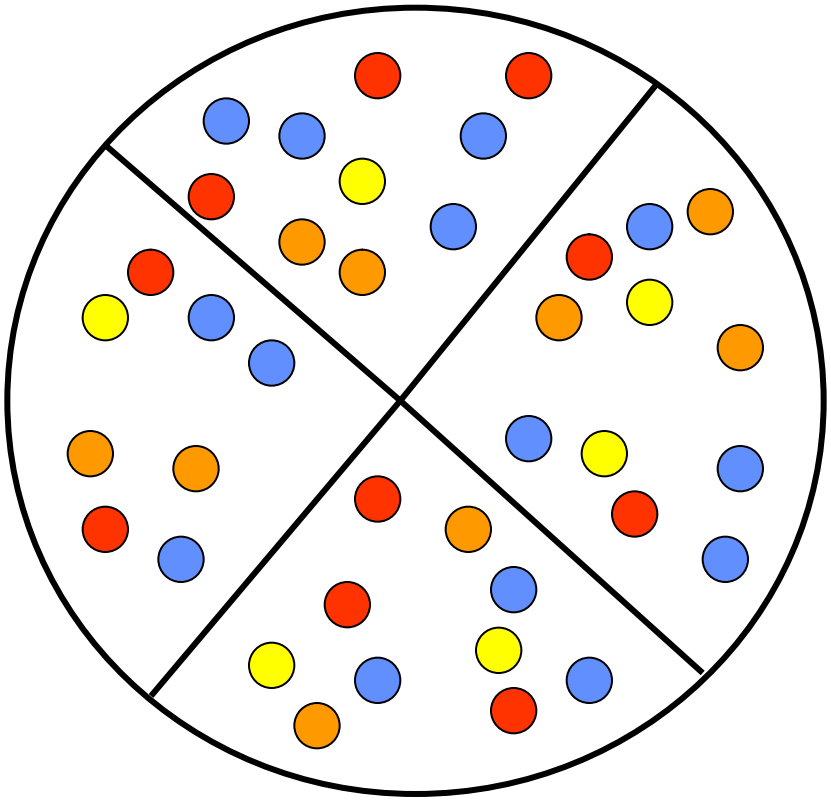
GDAŃSK UNIVERSITY OF TECHNOLOGY

Podział zbioru na *n* części (2)



- Przy zastosowaniu tego podejścia proces uczenia powtarzamy *n* razy,
- Przy każdym powtórzeniu jedną część zbioru (tu: ¼) przyjmujemy za zbiór testujący, resztę jako trenujący (tu: ¾)

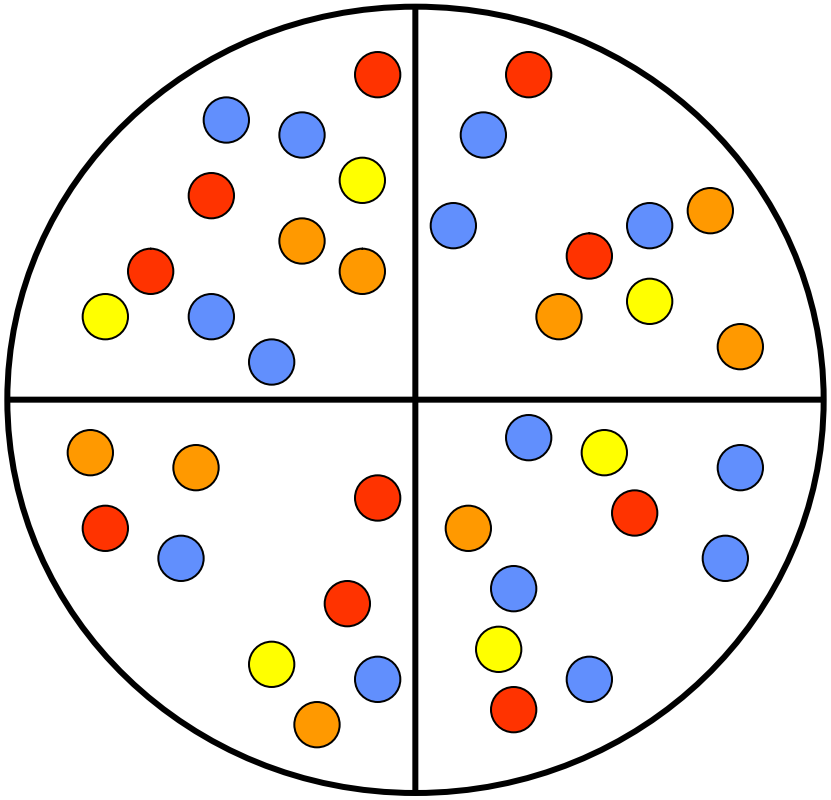
m-krotne dokonanie podziału



- Można jednak wskazać same granice podziału jako potencjalne źródło błędów,
- Rozwiązanie może być *m*-krotne powtórzenie podziału na *n* części,
- Proces uczenia powtarzamy wówczas $m \cdot n$ razy

GDAŃSK UNIVERSITY OF TECHNOLOGY

m-krotne dokonanie podziału



- Można jednak wskazać same granice podziału jako potencjalne źródło błędów,
- Rozwiązanie może być *m*-krotne powtórzenie podziału na *n* części,
- Proces uczenia powtarzamy wówczas $m \cdot n$ razy

GDAŃSK UNIVERSITY OF TECHNOLOGY

Ocena jakości klasyfikatora

Walidacja krzyżowa

Walidacja skrośna

Kroswalidacja

- Opisana metoda jest jedną z najpopularniejszych metod oceny klasyfikatorów:

  ***m* *n*-fold cross-validation with stratification**

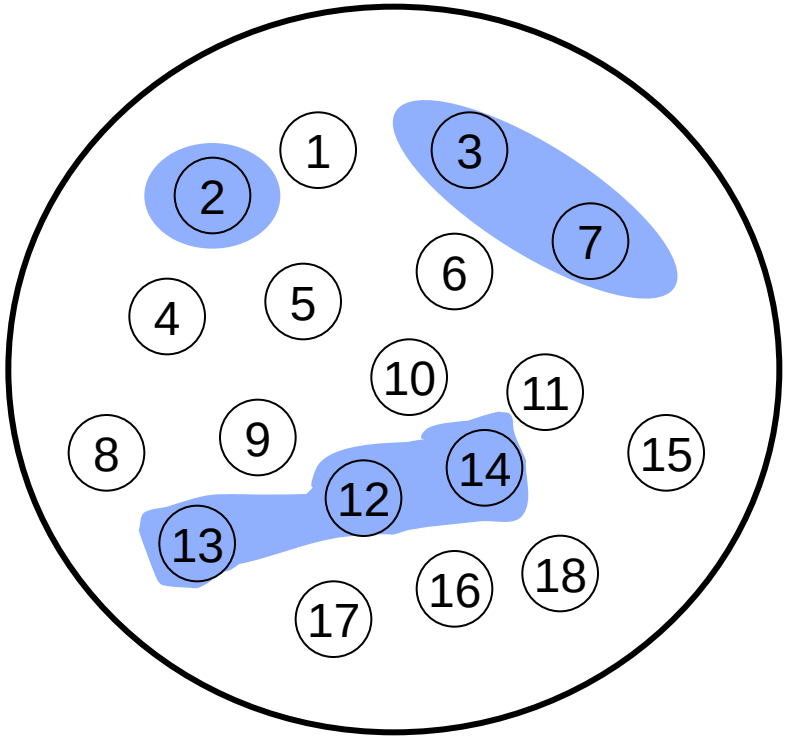
- Najczęściej przyjmujemy m=n=10 otrzymując:

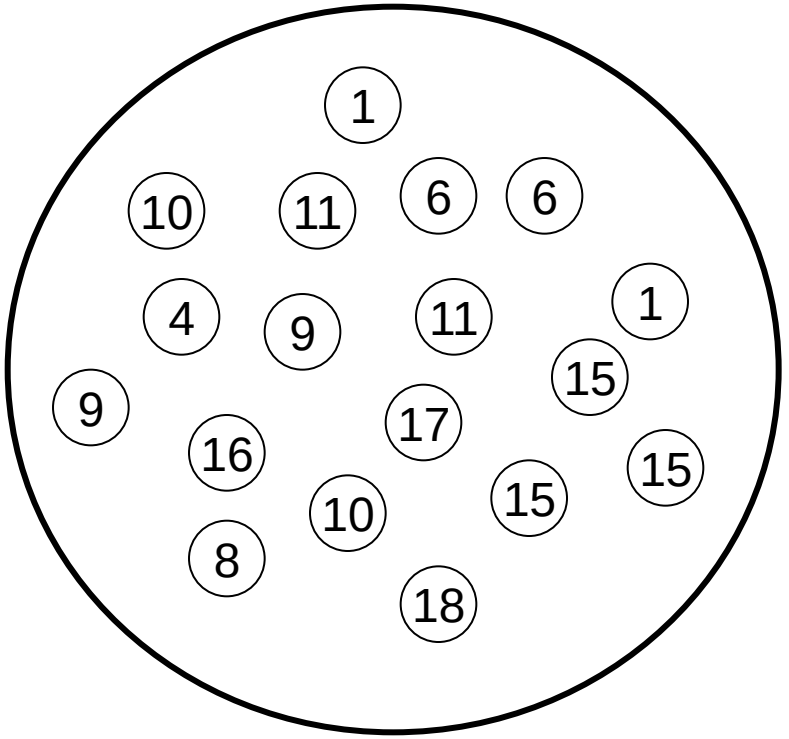
  10 10-fold cross-validation with stratification

- Metoda ta wymaga jednak wielu powtórzeń procesu uczenia i dlatego czasem stosowane są metody mniej zasobochłonne

GDAŃSK UNIVERSITY OF TECHNOLOGY

0.632 bootstrap



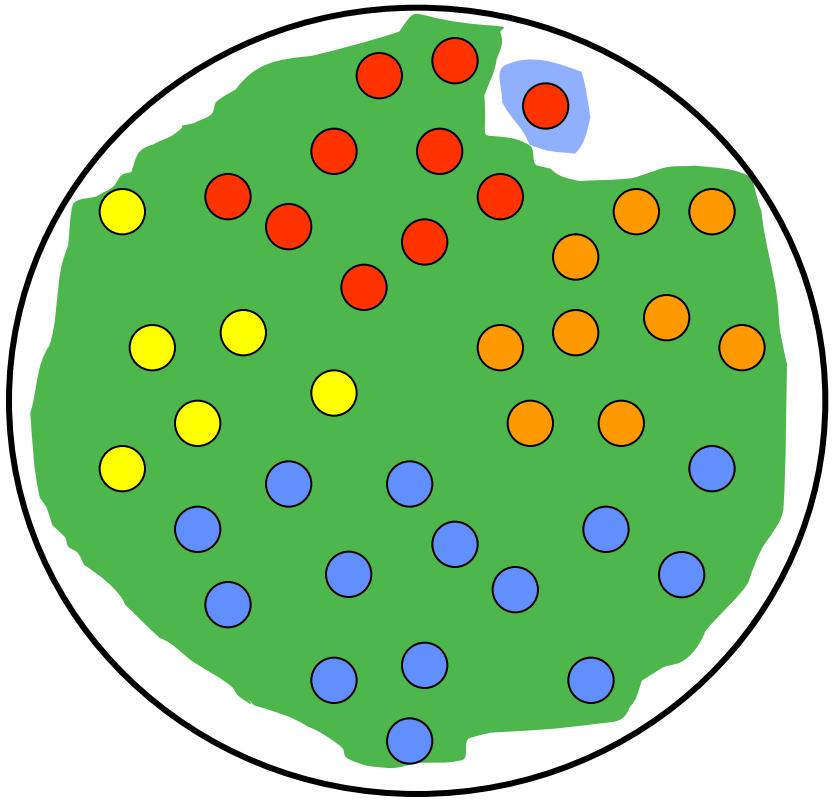


***k* razy wybieramy losowo przykład, który umieszczamy w zbiorze trenującym (*k* – moc zbioru przykładów, tu: 18)**

Przykłady niewybrane ani razu stają się naszym zbiorem testującym

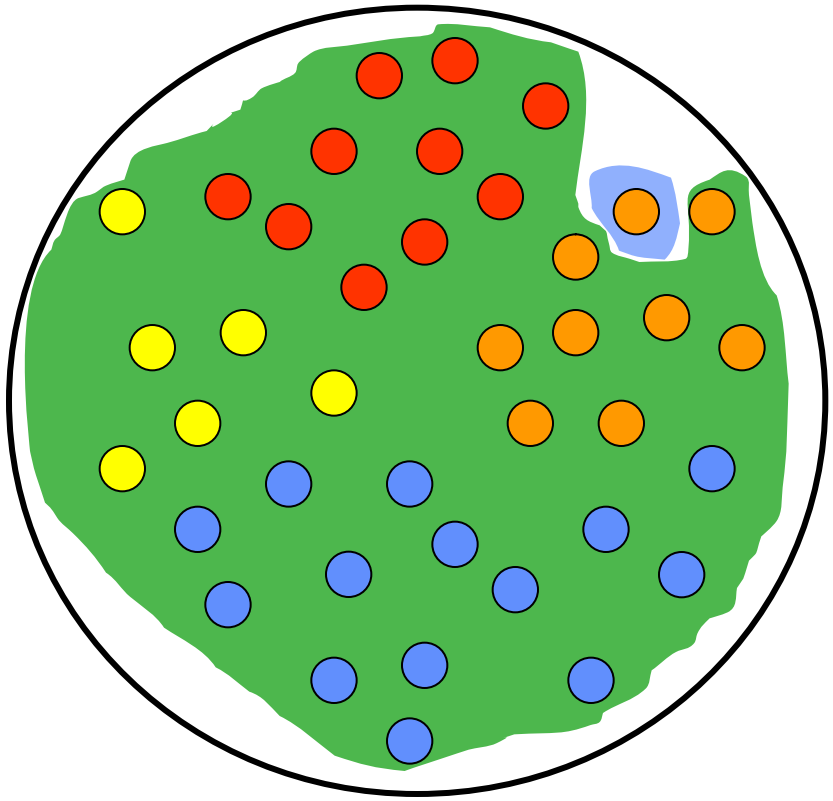
Liczba przykładów niewybranych stanowi statystycznie 0.368 *k* (wybranych 0.632)

Leave-one-out



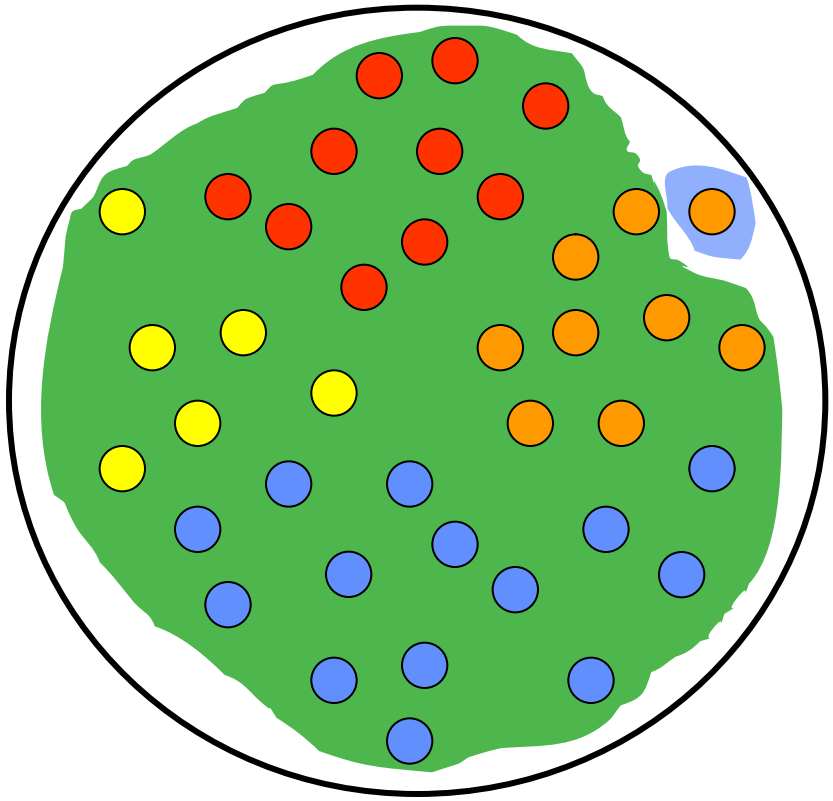
- Tutaj za zbiór testujący przyjmujemy jeden przykład,
- Proces uczenia powtarzamy *k* razy, ale nie musimy *m* razy „dokonywać podziału” ani wykonywać kosztownego procesu stratyfikacji

Leave-one-out



- Tutaj za zbiór testujący przyjmujemy jeden przykład,
- Proces uczenia powtarzamy *k* razy, ale nie musimy *m* razy „dokonywać podziału” ani wykonywać kosztownego procesu stratyfikacji

Leave-one-out



itd.

- Tutaj za zbiór testujący przyjmujemy jeden przykład,
- Proces uczenia powtarzamy *k* razy, ale nie musimy *m* razy „dokonywać podziału” ani wykonywać kosztownego procesu stratyfikacji

GDAŃSK UNIVERSITY OF TECHNOLOGY

Prezentacja efektów oceny

Analiza wyników:

- Macierze błędu,
- Wykres wzrostu (przyrostu),
- Krzywe ROC

GDAŃSK UNIVERSITY OF TECHNOLOGY

Macierz błędów

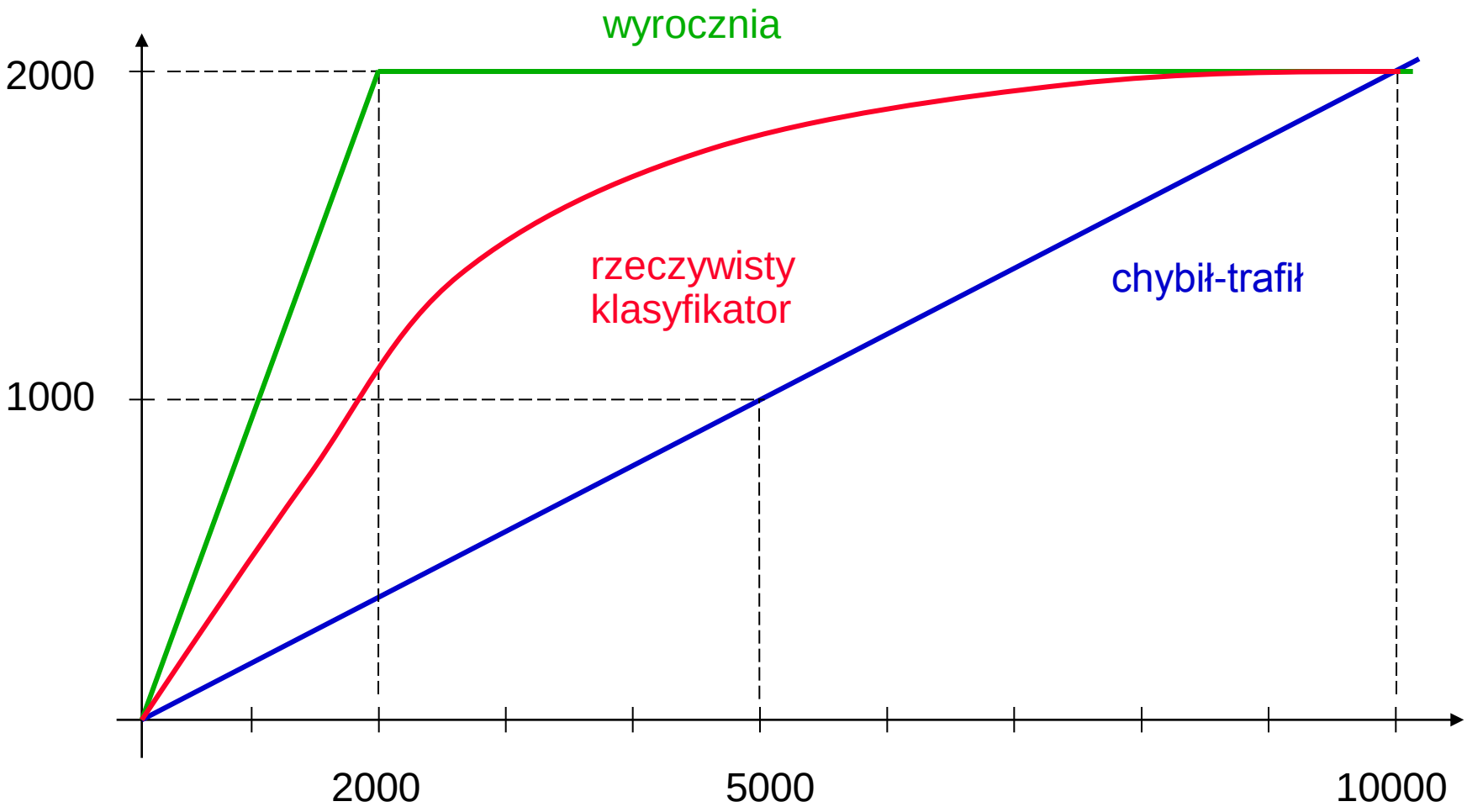
	tak (przewidywana)	nie (przewidywana)
tak (rzeczywista)	1053	229
nie (rzeczywista)	283	682

Uwaga 1: Niektóre klasyfikatory można „ukierunkować” na popełnianie konkretnych (mniej kosztownych) rodzajów błędów (np. NKB).

Uwaga 2: Część klasyfikatorów udziela odpowiedzi „niepewnej” (p_j dla j-tej klasy). Jakość takich odpowiedzi możemy mierzyć bardziej skomplikowaną funkcją, np. Σ_j ($p_j - a_j$), przy czym a_j = 1 wtwg. przykład faktycznie należy do j-tej klasy.

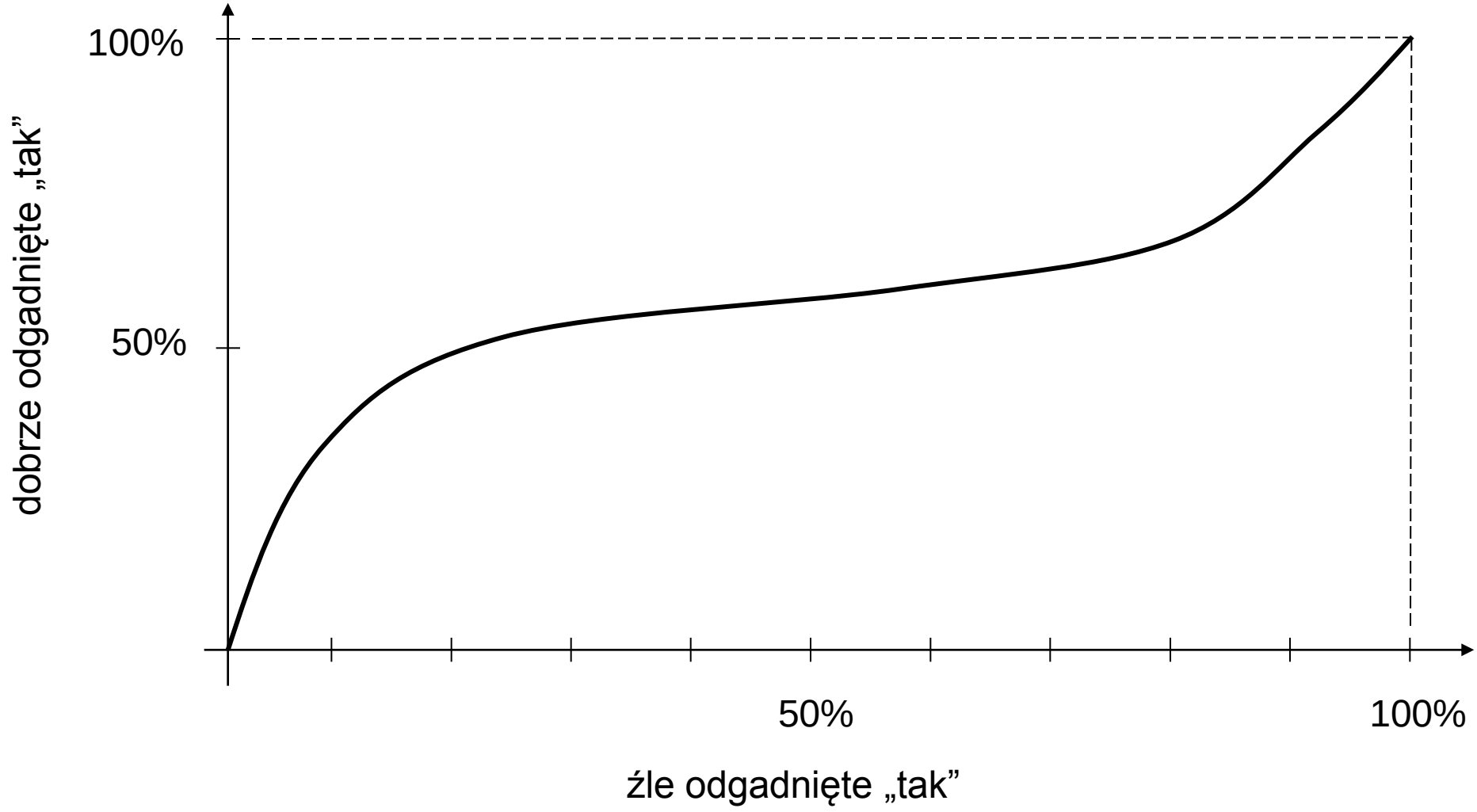
GDAŃSK UNIVERSITY OF TECHNOLOGY

Wykres wzrostu



GDAŃSK UNIVERSITY OF TECHNOLOGY

Krzywa ROC



Ocena jakości predykcji numerycznej

błąd średni $\frac{(p_1 - a_1)^2 + \ldots + (p_n - a_n)^2}{n}$

błąd średniokwadratowy $\sqrt{\frac{(p_1 - a_1)^2 + \ldots + (p_n - a_n)^2}{n}}$

błąd średni bezwzględny $\frac{|p_1 - a_1| + \ldots + |p_n - a_n|}{n}$

względny błąd średni $\frac{(p_1 - a_1)^2 + \ldots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \ldots + (a_n - \bar{a})^2}$

względny błąd średniokwadratowy $\sqrt{\frac{(p_1 - a_1)^2 + \ldots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \ldots + (a_n - \bar{a})^2}}$

względny błąd bezwzględny $\frac{|p_1 - a_1| + \ldots + |p_n - a_n|}{|a_1 - \bar{a}| + \ldots + |a_n - \bar{a}|}$

UWAGA:
p_i to odpowiedź dla *i*-tego przykładu
a_i to wartość rzeczywista dla *i*-tego przykładu
$\bar{a}$ to wartość średnia dla danego atrybutu
n to liczba przykładów w zbiorze

GDAŃSK UNIVERSITY OF TECHNOLOGY

Dziękujemy za uwagę

Zapraszamy na wykład:

ELEMENTY INŻYNIERII WEJŚCIA I WYJŚCIA