KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Prezentacja multimedialna
współfinansowana przez Unię Europejską
w ramach Europejskiego Funduszu Społecznego
w projekcie

*„Innowacyjna dydaktyka bez ograniczeń
– zintegrowany rozwój Politechniki Łódzkiej –
zarządzanie Uczelnią,
nowoczesna oferta edukacyjna
i wzmacniania zdolności do zatrudniania
osób niepełnosprawnych”*

dr inż. Artur Klepaczko

Eksploracja danych
Drzewa i reguły decyzyjne

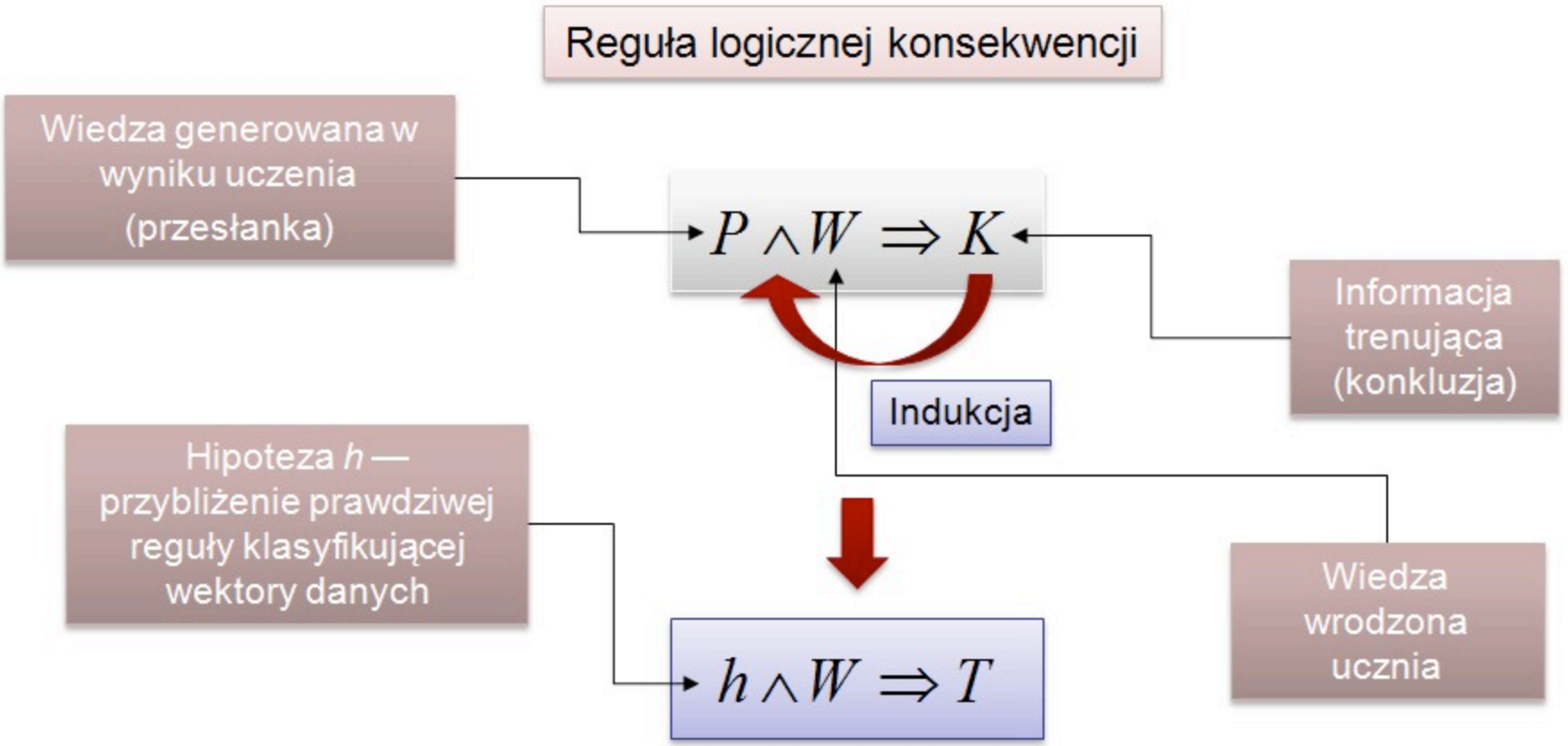
Zadanie nr 13 – Studia podyplomowe „Przetwarzanie i analiza obrazów biomedycznych”

Politechnika Łódzka
Instytut Elektroniki

90-924 Łódź, ul. Żeromskiego 116,
tel. 042 631 28 83
www.kapitalludzki.p.lodz.pl

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

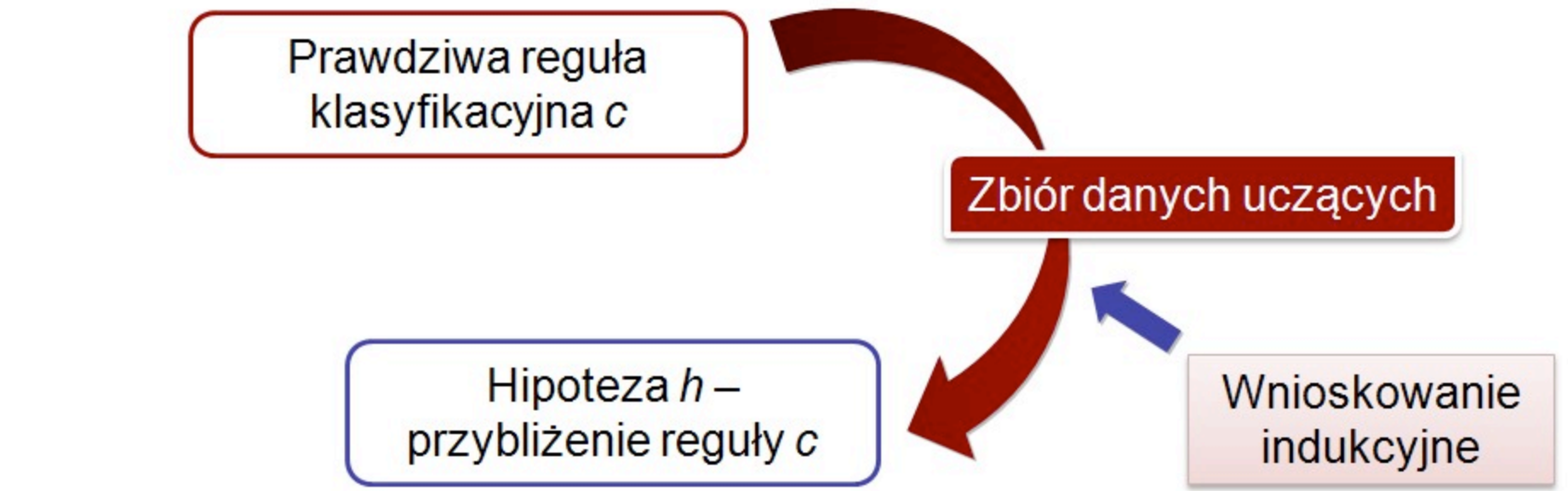
Uczenie przez indukcję



We wnioskowaniu indukcyjnym zakładamy, że informacja trenująca dostępna w procesie uczenia jest logiczną konsekwencją hipotezy *h* oraz być może niepustej wiedzy wrodzonej ucznia *W*.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Indukcja reguł decyzyjnych



Postać reguły decyzyjnej

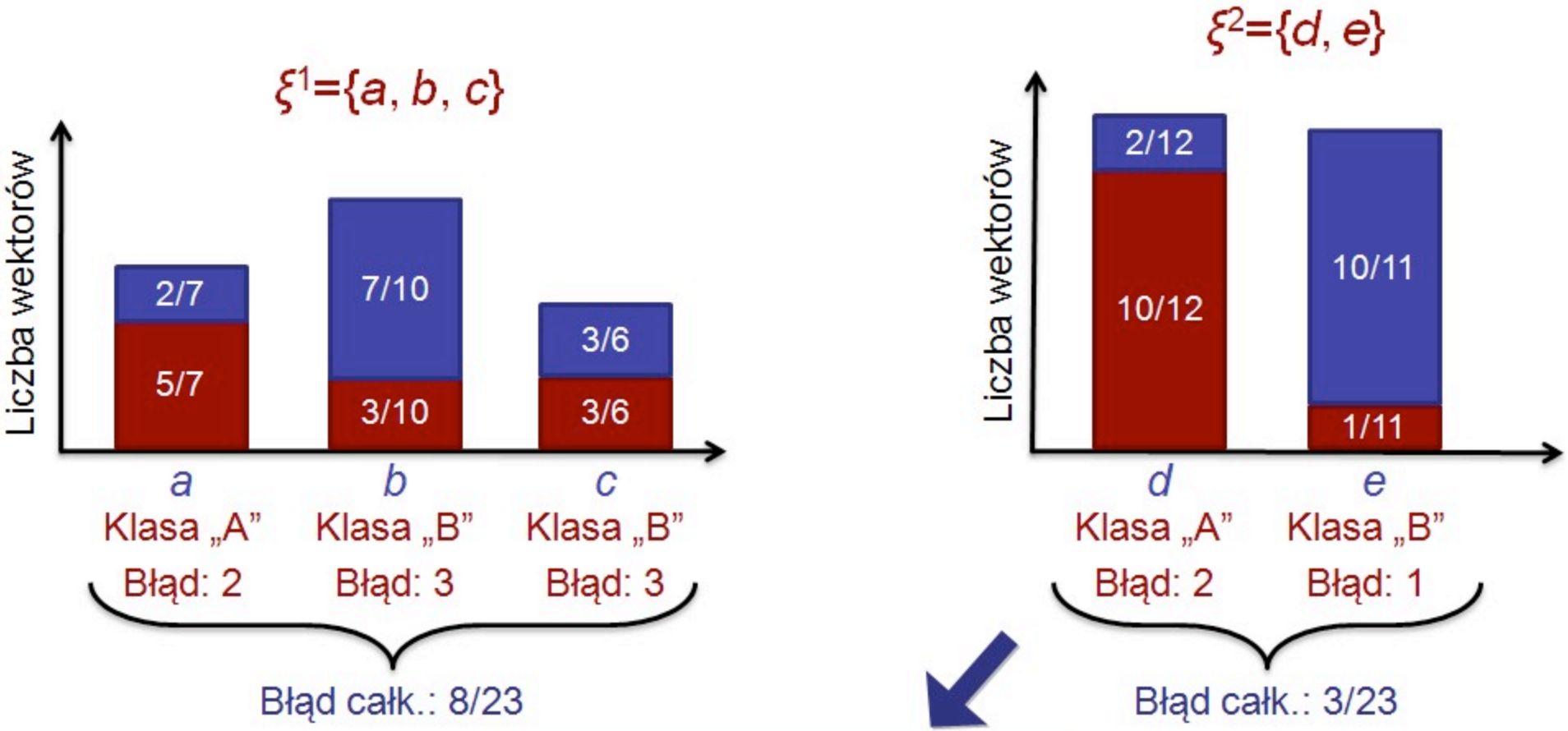
JEŚLI *Warunek#1* & *Warunek#2* & … *Warunek#n* **TO** *Decyzja*

Część warunkowa

Część decyzyjna

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Reguły proste – algorytm 1-R



JEŻELI $\xi^2 = d$ **TO** Klasa „A”
$= e$ **TO** Klasa „B”

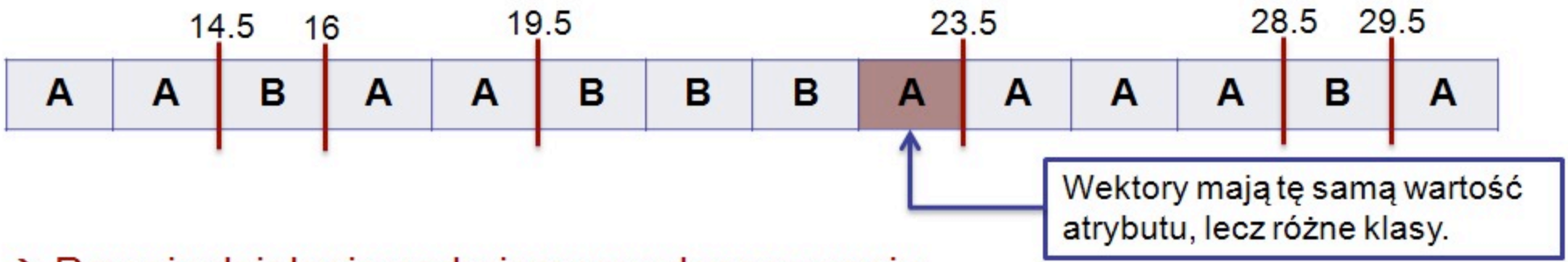
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Dyskretyzacja atrybutów ilościowych

→ Porządkowanie według atrybutu numerycznego

12	14	15	17	19	20	21	22	22	25	27	28	29	30
A	A	B	A	A	B	B	B	A	A	A	A	B	A

→ Wydzielenie przedziałów wartości atrybutu, dla których klasa nie zmienia się



→ Przeciwdziałanie nadmiernemu dopasowaniu

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zbiór danych — *Iris*

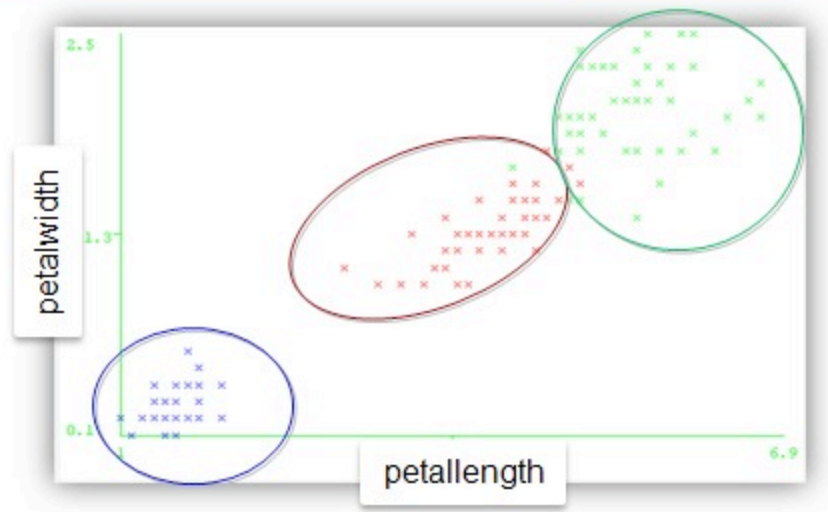
Iris – zbiór pochodzący z repozytorium wzorcowych zbiorów danych UCI*

Źródło: pl.wikipedia.org.
Grafika dostępna na zasadach GFDL

Liczba wektorów danych: 150

Liczba klas: 3

Liczba cech 4 (długość i szerokość działki kielicha (ang. *sepal*) oraz płatka kwiatowego (ang. *petal*))



*UCI – University of California, Irvine. Machine Learning Repository. http://archive.ics.uci.edu/ml/index.html

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Algorytm 1-R w zastosowaniu do zbioru danych *Iris*

```
=== Classifier model (full training set) ===

petallength:
        < 2.45      -> Iris-setosa
        < 4.85      -> Iris-versicolor
        >= 4.85     -> Iris-virginica

(143/150 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         143               95.3333 %
Incorrectly Classified Instances         7                4.6667 %
```

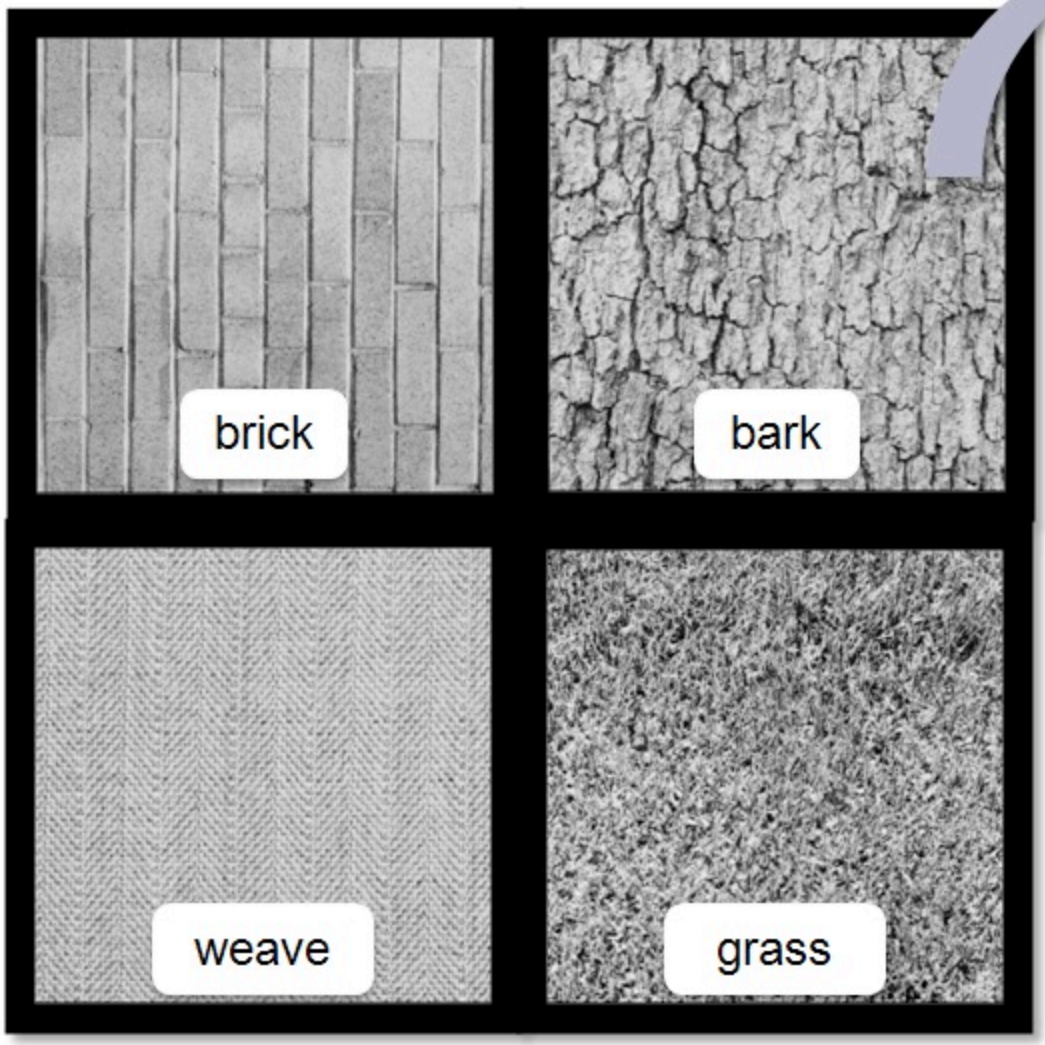
Wynik uzyskany przy użyciu programu Weka Explorer

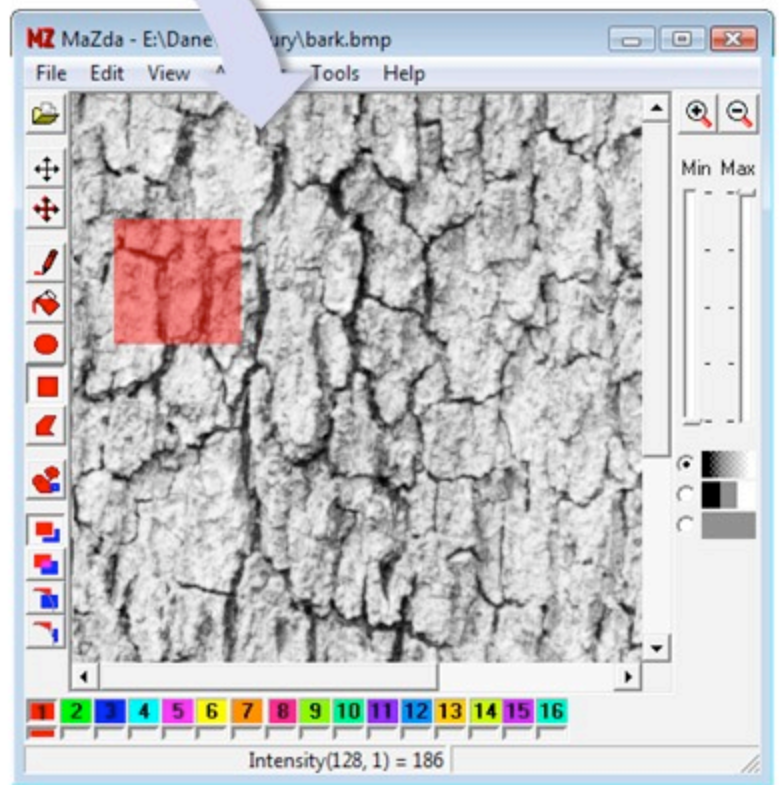
Algorytm 1-R

- skuteczny w przypadku wielu rzeczywistych zbiorów danych
- wydajny (ze względu na krótki czas obliczeń)
- prowadzi do prostej, łatwej w interpretacji hipotezy prawdziwej reguły decyzyjnej

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zbiór danych — *Tekstury Brodatza**





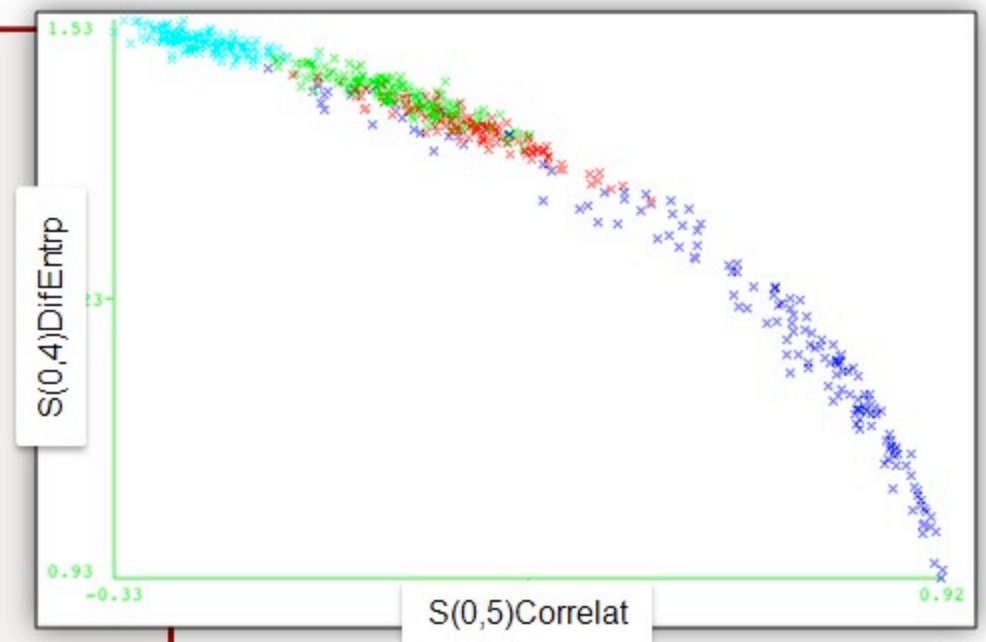
MaZda
wyznaczenie i selekcja cech tekstury

*Brodatz P., *A photographic album for artists and designers*, Dover, New York 1966.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Algorytm 1-R – klasyfikacja tekstur

```
=== Classifier model (full training set) ===

S(0,4)DifEntrp:
                < 1.3324805  -> brick
                < 1.4217385  -> calf_leather
                < 1.4260485  -> grass
                < 1.4309895  -> calf_leather
                < 1.441526   -> grass
                < 1.4443815  -> calf_leather
                < 1.4808575  -> grass
                >= 1.4808575  -> herringbone_weave
(533/640 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         533               83.2813 %
Incorrectly Classified Instances       107               16.7188 %
```



- Wektory należące do różnych klas tekstur nie tworzą wyraźnie odrębnych skupień.
- Klasyfikator 1-R mimo to osiąga dokładność na poziomie 80%.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Walidacja krzyżowa – ocena skuteczności uczenia

Zbiór wektorów danych o liczności Q

Podział na trzy podzbiory

S_1 (1/3*Q)	S_2 (1/3*Q)	S_3 (1/3*Q)

Trzy cykle uczenia:

Nr	Podzbiór uczący	Podzbiór testowy	Błąd
1	S_1, S_2	S_3	e_1
2	S_1, S_3	S_2	e_2
3	S_2, S_3	S_1	e_3

Błąd na zbiorze uczącym:

$$e = (e_1+e_2+e_3)/3$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Uwiarygodniona ocena klasyfikatora – przykład

Ocena ze zbiorym uczącym

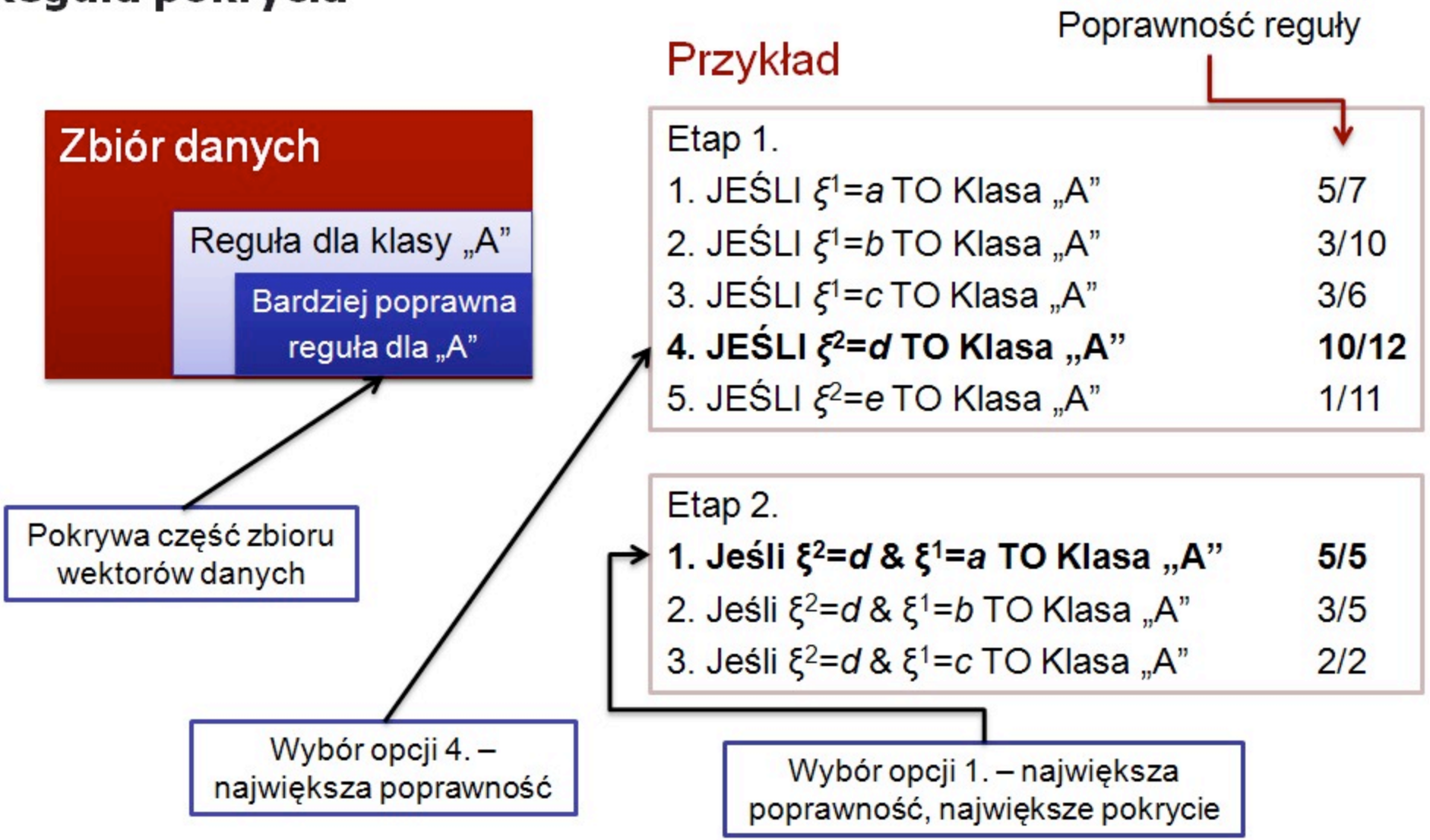
```
=== Classifier model (full training set) ===

S(0,4)DifEntrp:
                < 1.3324805  -> brick
                < 1.4217385  -> calf_leather
                < 1.4260485  -> grass
                < 1.4309895  -> calf_leather
                < 1.441526   -> grass
                < 1.4443815  -> calf_leather
                < 1.4808575  -> grass
                >= 1.4808575 -> herringbone_weave
(533/640 instances correct)

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         533               83.2813 %
Incorrectly Classified Instances       107               16.7188 %
```

Walidacja krzyżowa

```
=== Classifier model (full training set) ===

S(0,4)DifEntrp:
                < 1.3324805  -> brick
                < 1.4217385  -> calf_leather
                < 1.4260485  -> grass
                < 1.4309895  -> calf_leather
                < 1.441526   -> grass
                < 1.4443815  -> calf_leather
                < 1.4808575  -> grass
                >= 1.4808575  -> herringbone_weave
(533/640 instances correct)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         510               79.6875 %
Incorrectly Classified Instances       130               20.3125 %
```

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Reguła pokrycia



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Reguła pokrycia w zastosowaniu do zbioru *Iris*

```
=== Classifier model (full training set) ===

Prism rules
----------
If petal
If petal
If petal
If petal
If petal

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         147               98 %
Incorrectly Classified Instances         3                2 %
```

Algorytm pokrycia

- prowadzi do wielu, skomplikowanych reguł
- niekoniecznie zapewnia mniejszy błąd klasyfikacji niż reguła 1R
- poszczególne reguły mogą pokrywać ten sam podzbiór wektorów danych

Wynik uzyskany przy użyciu programu Weka Explorer

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przycinanie reguł

Wiele rozbudowanych reguł – czy warto?

TAK → Zwiększona dokładność klasyfikacji

NIE → Niewielka poprawa, *Overfitting*

IREP* – inkrementacyjne przycinanie ze zmniejszeniem błędu

Zbiór rozrostu → konstruowanie reguły – np. algorytm pokrycia

Zbiór przycinający → przycinanie – usuwanie poszczególnych warunków i test reguły

$$\frac{[p+(N-n)]}{T}$$

T – liczność zbioru przycinającego
p – liczba prawidłowo pokrytych wektorów
n – liczba nieprawidłowo pokrytych wektorów
N – liczba wszystkich negatywnych wektorów

*IREP – *incremental reduced-error prunning*

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Reguły z wyjątkami

→ Klasa 1 (domyślna)

Jak wybrać?

Np. klasa, która posiada najwięcej wektorów danych w zbiorze uczącym

Z wyjątkiem

JEŚLI *warunek* α **TO** Klasa 2

Alternatywnie

JEŚLI *warunek* β **TO** Klasa 3

Z wyjątkiem

JEŚLI *warunek* γ **TO** Klasa 2

Jak to rozumieć?

Wszystkie wektory danych należą do klasy 1 z wyjątkiem tych, które spełniają warunek α — one należą do klasy 2, lub tych, które spełniają warunek β — one należą do klasy 3 z wyjątkiem tych, które spełniają warunek γ — one także należą do klasy 2.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przycięte reguły z wyjątkami w zastosowaniu do zbioru *Iris*

```
=== Classifier model (full training set) ===

RIpple DOwn Rule Learner(Ridor) rules
--------------------------------------

class = Iris-setosa  (150.0/100.0)
        Except (petallength > 2.45) => class = Iris-versicolor  (67.0/0.0) [33.0/0.0]
                Except (petalwidth > 1.75) and (petallength > 4.85) => class = Iris-virginica  (29.0/0.0) [14.0/0.0]

Total number of rules (incl. the default rule): 3

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         141               94 %
Incorrectly Classified Instances         9                6 %
```

Algorytm Ridor

- reguły na najniższym poziomie są ignorowane
- reguły na wyższych poziomach są przycinane (za pomocą metody IREP)
- można ustalić minimalny stosunek liczności podzbiorów danych pokrytego i wykluczonego przez regułę

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Decyzje na podstawie prawdopodobieństwa

ξ^1			ξ^2			Klasa	
	A	B		A	B	A	B
a	5/11	2/12	*d*	10/11	2/12	11/23	12/23
b	3/11	7/12	*e*	1/11	10/12		
c	3/11	3/12					

Prawdopodobieństwo tego, że dla wektora danych należącego do klasy „A" atrybut ξ^1=a.

Prawdopodobieństwo tego, że wektor danych należy do klasy „B"

Jaka będzie klasa nowego wektora $\mathbf{x}_{test} = \{b, d\}$?

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Naiwny klasyfikator Bayes'a

Wektor testowy: $\mathbf{x}_{test}=\{b, d\}$

Prawdopodobieństwo *a priori* klasy „A”
Pr(A) = 11/23

Prawdopodobieństwo *a priori* klasy „B”
Pr(B) = 12/23

Prawdopodobieństwo *a posteriori* klasy „A”
Pr(Klasa „A” | $\mathbf{x}_{test}$)

Prawdopodobieństwo *a posteriori* klasy „B”
Pr(Klasa „B” | $\mathbf{x}_{test}$)

Iloczyny prawdopodobieństw atrybutów

3/11 × 10/11 × 11/23 ≈ 0,12

7/12 × 2/12 × 12/23 ≈ 0,05

Obliczenia są poprawne przy (naiwnym) założeniu niezależności atrybutów.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Klasyfikator Bayesa dla atrybutów ciągłych

- Zamiast prawdopodobieństw atrybutów oblicza się gęstość prawdopodobieństwa
- Należy założyć określony typ rozkładu i funkcji gęstości prawdopodobieństwa, np. dla rozkładu normalnego →

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Dla każdej klasy C
  - Dla każdego atrybutu ξ^j
    - Wartość średnia μ^j →

$$\frac{\text{suma wartości atrybutu } \xi^j \text{ wektorów z klasy } C}{\text{liczność klasy } C}$$

    - Odchylenie standardowe σ^j →

$$\sqrt{\frac{\text{suma } (\xi^j - \sigma^j)^2 \text{ dla wektorów z klasy } C}{\text{liczność klasy } C - 1}}$$

- Klasę dla wektora testowego wyznacza się obliczając funkcje gęstości prawdopodobieństw poszczególnych atrybutów

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

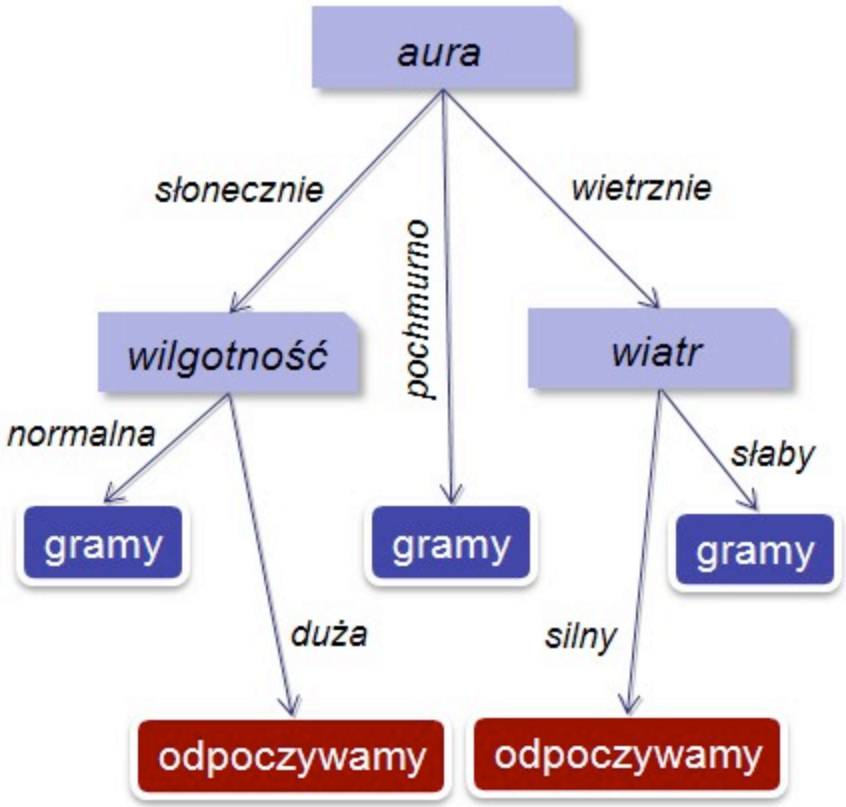
Naiwny klasyfikator Bayesa w zastosowaniu do zbioru tekstur

```
=== Classifier model (full training set) ===

Naive Bayes Classifier

                    Class
Attribute             brick     calf_leather        gras
                     (0.25)       (0.25)         (0.25)
=========================================
S(0,5)DifEntrp
  mean                1.2085       1.4297          1.4
  std. dev.           0.1393       0.0253          0.0

...

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         512               80               %
Incorrectly Classified Instances       128               20               %
```

```
=== Classifier model (full training set) ===

S(0,4)DifEntrp:
                < 1.3324805  -> brick
                < 1.4217385  -> calf_leather
                < 1.4260485  -> grass
                < 1.4309895  -> calf_leather
                < 1.441526   -> grass
                < 1.4443815  -> calf_leather
                < 1.4808575  -> grass
                >= 1.4808575  -> herringbone_weave
(533/640 instances correct)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         510               79.6875 %
Incorrectly Classified Instances       130               20.3125 %
```

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Drzewa decyzyjne – metoda „dziel-i-rządź”



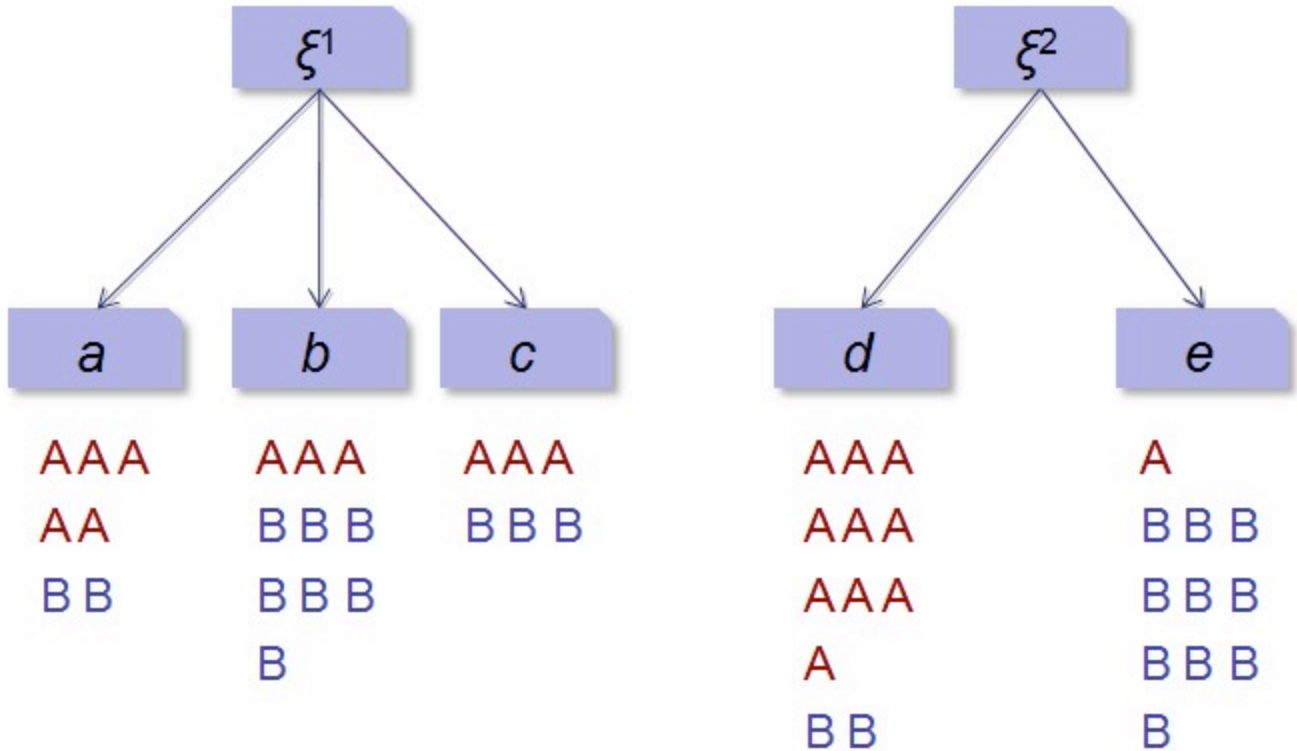
Jak to zostało skonstruowane?

- Wybierz atrybut, który zostanie umieszczony w węźle–**korzeniu**.
- **Podziel** zbiór wektorów danych na podzbiory odpowiadające poszczególnym wartościom wybranego atrybutu.
- Jeśli w utworzonych podzbiorach wszystkie wektory należą do tej samej klasy, to reprezentujący je węzeł staje się **liściem**.
- W przeciwnym przypadku, węzeł staje się **rodzicem**, w któym należy wybrać kolejny atrybut rozdzielający wektory danych na podzbiory.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Konstruowanie drzew decyzyjnych – wybór atrybutu

Dwie możliwości – który atrybut lepiej nadaje się na *korzeń*?



Potrzebna jest obiektywna miara do oceny, jak poszczególne atrybuty nadają się do podziału wektorów w danym węźle drzewa.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Entropia – chaos i porządek

Definicja termodynamiczna

- *Miara stopnia nieuporządkowania układu makroskopowego*

Definicja teorioinformacyjna

- *Ilość informacji zawartej w wiadomości wysłanej przez źródło*

Chaos

AAAAA BB | AAA BBB | AAA BBBBBBB

Informacja niejednoznaczna — obie klasy są równie prawdopodobne.

Porządek

AAAAAAAAAA BB | A BBBBBBBBBB

Informacja jednoznaczna — klasa „A” jest bardziej prawdopodobna.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zysk informacyjny

Entropia informacyjna	• $\text{Entrp}(p_1, p_2, \ldots, p_n) = -p_1 \log p_1 - p_2 \log p_2 \ldots - p_n \log p_n$
Entropia zbioru w węźle-korzeniu (11 wektorów klasy „A”, 12 wektorów klasy „B”)	• $\text{Entrp}(11/23, 12/23) = -11/23 \times \log(11/23) - 12/23 \times \log(12/23) = \mathbf{0,99\ bit}$
Średnia entropia w węźle ξ^1	• $7/23 \times \text{Entrp}(2/7,5/7) + 6/23 \times \text{Entrp}(3/6,3/6) + 10/23 \times \text{Entrp}(3/10,7/10) = 0,26 + 0,26 + 0,38 = \mathbf{0,90\ bit}$
Średnia entropia w węźle ξ^2	• $11/23 \times \text{Entrp}(1/11,10/11) + 12/23 \times \text{Entrp}(10/12,2/12) = 0,21 + 0,34 = \mathbf{0,55\ bit}$
Zysk informacyjny w przypadku wyboru atrybutu ξ^1	• $0,99 - 0,90 = \mathbf{0,09\ bit}$
Zysk informacyjny w przypadku wyboru atrybutu ξ^2	• $0,99 - 0,55 = \mathbf{0,44\ bit}$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Drzewa decyzyjne dla atrybutów rzeczywistych

12	14	15	17	19	20	21	22	22	25	27	28	29	30
B	**A**	**B**	**A**	**B**	**B**	**B**	**A**	**A**	**A**	**A**	**A**	**B**	**A**

- Porządkowanie wektorów danych według atrybutu numerycznego
- Podział wektorów na dwa podzbiory
  - Drzewa binarne
  - Kryterium podziału – zysk informacyjny
- Możliwe są kolejne testy tego samego atrybutu
  - Wada: konstruowane drzewa są zbyt rozbudowane i trudne w interpretacji
- Alternatywnie: drzewa niebinarne (więcej niż dwa rozgałęzienia w węźle)
  - Wada: trudniejsze w implementacji
- Alternatywnie: dyskretyzacja atrybutu
  - Wada: utrata części informacji

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Indukcja drzewa decyzyjnego – przykład

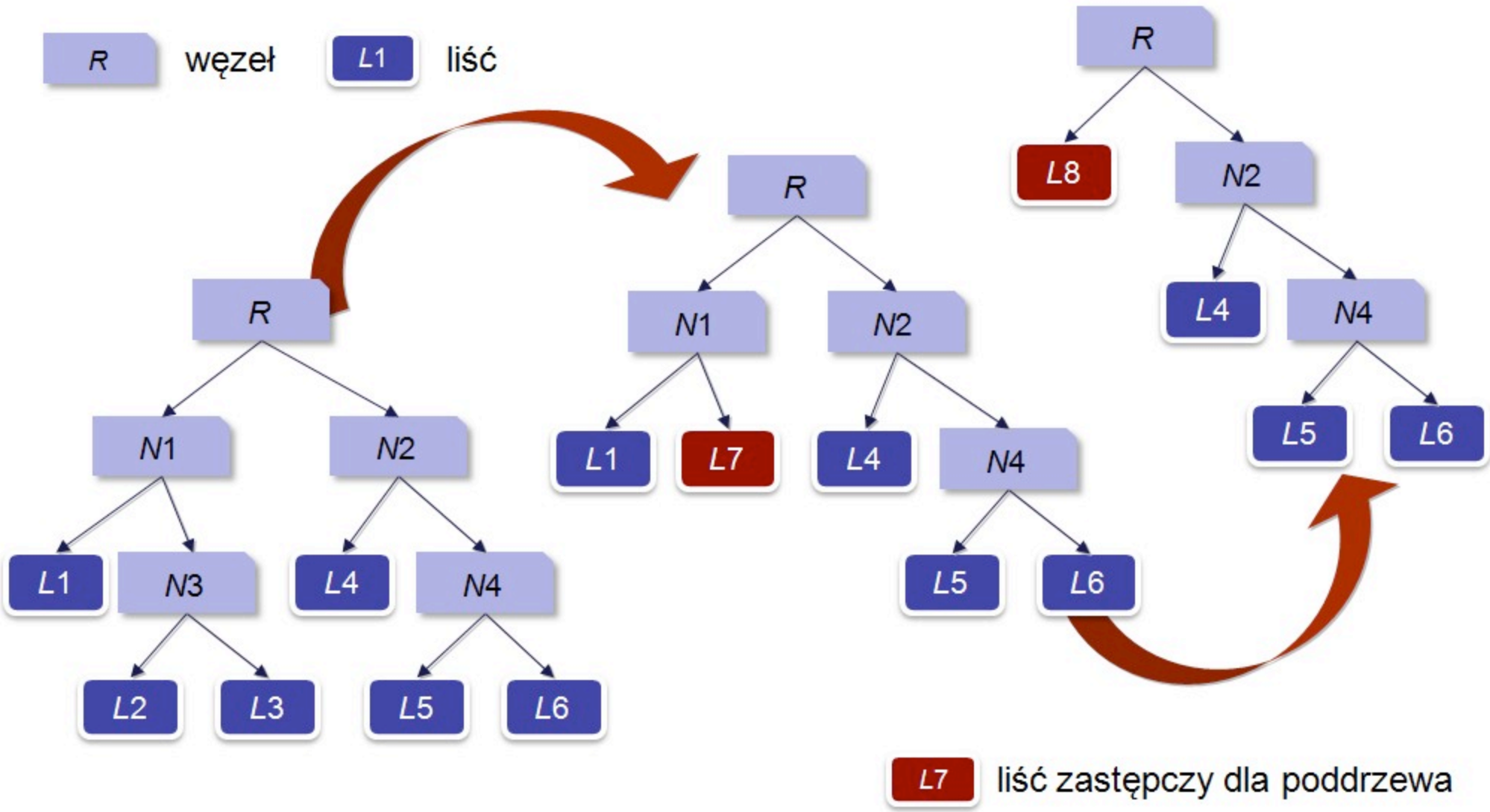
```
=== Classifier model (full training set) ===

J48 unpruned tree
------------------

S(0,5)Correlat <= -0.063539
|   S(0,5)Correlat <= -0.0943: herringbone_weave (153.0/1.0)
|   S(0,5)Correlat > -0.0943
|   |   S(0,4)DifEntrp <= 1.485315
|   |   |   S(0,5)DifEntrp <= 1.492711: herringbone_weave (3.0/1.0)
|   |   |   S(0,5)DifEntrp > 1.492711: grass (5.0)
|   |   S(0,4)DifEntrp > 1.485315: herringbone_weave (5.0)

Number of Leaves  :     15
Size of the tree :      29

=== Summary ===

Correctly Classified Instances         522               81.5625 %
Incorrectly Classified Instances       118               18.4375 %
```

Proste drzewa decyzyjne

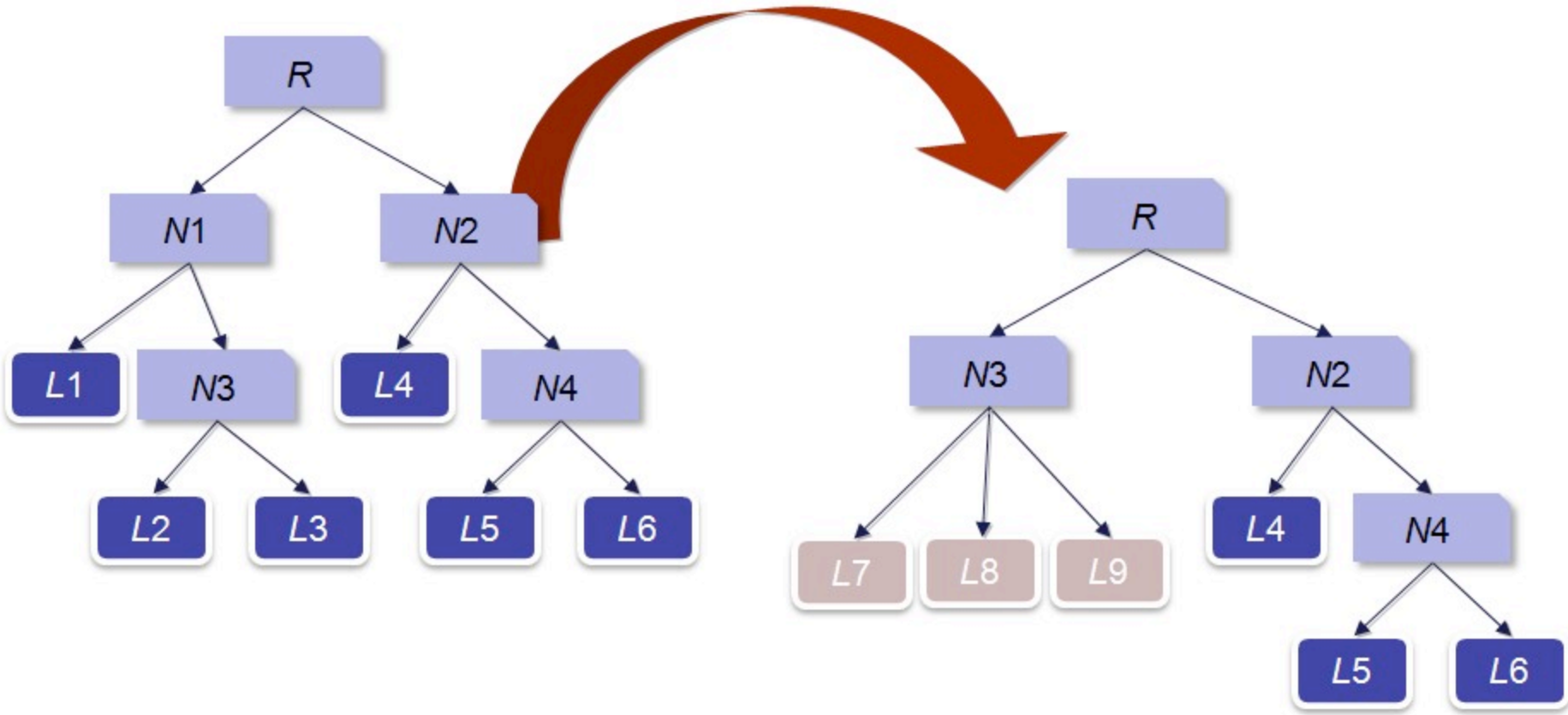
- w przypadku rzeczywistych zbiorów danych powstaje wiele rozgałęzień
- istnieje duże ryzyko nadmiernego dopasowania drzewa do zbioru treningowego
- potrzebny jest mechanizm uproszczenia drzewa — przycinanie

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przycinanie drzewa — *subtree replacement*



Przycinanie drzewa — *subtree raising*



Kiedy przycina się drzewa?

L7 nowy liść poddrzewa

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

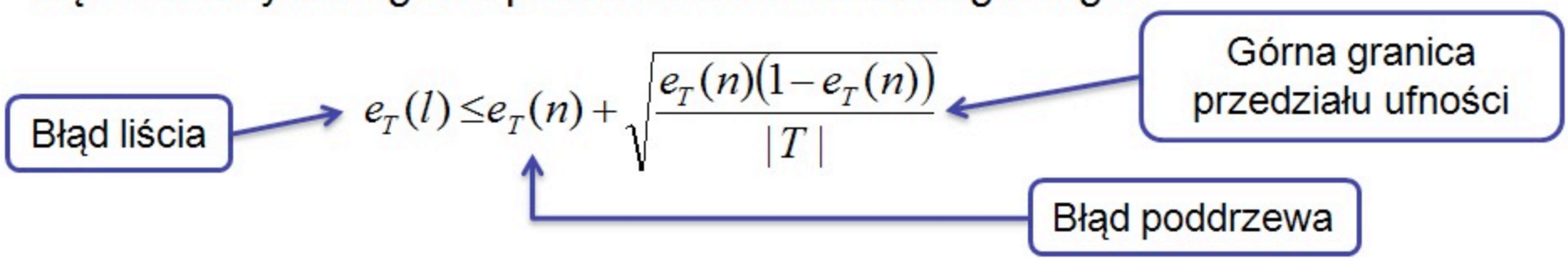
Kryterium przycinania

Przycinanie redukujące błąd

- Oddzielny zbiór testowy
- Duży zbiór trenujący, z którego można wydzielić podzbiór testowy
- Przycinanie następuje, jeśli błąd na zbiorze testowym nie zwiększy się.

Przycinanie pesymistyczne

- Brak oddzielnego zbioru testującego
- Zbiór treningowy jest zbyt mały, aby rezygnować z części wektorów danych w trakcie uczenia.
- Przycinanie następuje, jeśli nie zwiększy się pesymistyczne oszacowanie błędu rzeczywistego na podstawie zbioru treningowego.



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przycięte drzewa w zastosowaniu do zbioru tekstur

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

S(0,5)Correlat <= -0.063539
|   S(0,5)Correlat <= -0.0943: herringbone_weave (153.0/1.0)
|   S(0,5)Correlat > -0.0943
|   |   S(0,4)DifEntrp <= 1.485315
|   |   |   S(0,5)DifEntrp <= 1.492711: herringbone_weave (3.0/1.0)
|   |   |   S(0,5)DifEntrp > 1.492711: grass (5.0)
|   |   S(0,4)DifEntrp > 1.485315: herringbone_weave (5.0)

Number of Leaves  :     12
Size of the tree :      23

=== Summary ===

Correctly Classified Instances         527               82.3438 %
Incorrectly Classified Instances       113               17.6563 %
```

Algorytm C4.5

- Redukcja liczby rozgałęzień
- Poprawa (niekiedy znaczna) jakości klasyfikacji
- Implementacja zarówno metody *subtree replacement* jak i *subtree raising*
- Skalę przycięcia można kontrolować poprzez zmianę jednego z parametrów