

**ETAPY REALIZACJI PROJEKTU** (dotyczy projektu nr 4)

**A. Etapy realizacji projektu w ujęciu ogólnym:**

- I. Przygotowanie danych dla programu DETREEX 4.0 (adaptacja przygotowanych wcześniej danych dla programu NEURONIX 4.0).

Prezentacja w postaci graficznej (z wykorzystaniem możliwości programu NEURONIX 4.0):

- a) danych uczących,
- b) danych testowych,
- c) danych weryfikujących,

tak aby obiekty (punkty) należące do poszczególnych klas były oznaczone różnymi symbolami graficznymi (identycznie jak w projektach nr 1 i nr 2).

Prezentacja w postaci graficznej obrazów uczących i weryfikujących przygotowanych w ramach projektu nr 3.

- II. Konstruowanie drzew decyzyjnych z wykorzystaniem programu DETREEX 4.0.

Konstrukcja maksymalnego drzewa decyzyjnego oraz wszystkich możliwych do uzyskania jego wariantów (odpowiadających różnym stopniom przycięcia drzewa maksymalnego) i odczyt parametrów umożliwiających określenie zależności procentu poprawnych decyzji systemu dla danych testowych i uczących w funkcji rozmiaru drzewa.

- III. Ilustracja graficzna uzyskanych reguł decyzyjnych (dla danych dwuwymiarowych).

- IV. (*opcjonalnie*) Wykorzystanie programu PC-SHELL 4.0 z pakietu SPHINX 4.0 do generowania odpowiedzi systemu decyzyjnego dla danych weryfikujących.

## B. Etapy realizacji projektu w ujęciu bardziej szczegółowym:

### I. Przygotowanie danych dla programu DETREEX 4.0 (adaptacja przygotowanych wcześniej danych dla programu NEURONIX 4.0).

- utworzenie nowego projektu (wskazane jest korzystanie z *Kreatora nowego projektu* ale tylko do pewnego momentu jego działania): <plik> → <Nowy projekt> → <Kreator nowego projektu> → w oknie dialogowym 'Nowy projekt', w 'Zapisz w' **wybrać dysk 'D' i wejść do katalogu 'SYSTEMY INTELIGENTNE'**, nazwać projekt (najlepiej jedną literą, np. 'a') i zapisać → 4 razy kliknąć przycisk '>' w żółtym oknie Kreatora a następnie przycisk 'Koniec' w tym oknie → dokonać edycji plików: uczącego 'a.lrn' i testowego 'a.tst' (**zgodnie z zasadami edycji plików dla programu DETREEX**) wykorzystując przygotowane przez siebie dane,
- prezentacja w postaci graficznej danych uczących, testowych i weryfikujących (każde w innym oknie; okna te należy następnie zwinąć do paska na dole ekranu) – dotyczy danych dwuwymiarowych [identycznie jak w projektach nr 1 i nr 2]:
  - dla danych uczących: otworzyć nowy arkusz → otworzyć plik 'a.lrn' → zaznaczyć w pliku 'a.lrn' **części wejściowe** (kolumny typu 'we') dla punktów należących do pierwszej klasy i przekopiować je do pierwszych dwóch kolumn w nowym arkuszu → zaznaczyć w pliku 'a.lrn' **części wejściowe** (kolumny typu 'we') dla punktów należących do drugiej klasy i przekopiować je do kolejnych dwóch kolumn w nowym arkuszu → czynności te powtórzyć dla punktów należących do trzeciej klasy → zaznaczyć całą zawartość nowego arkusza (dla trzech klas zawierających po 20 punktów będzie to sześć 20-elementowych kolumn) → wybrać <Plik> → <Nowy> → <Wykres> → zapisać arkusz pod nazwą np. 'a-dane uczace' → wybrać z 'Gallery' '2D', 'XY(Scat)' → 'Dalej' → style '5' → 'Dalej' → w 'Chart Title' wpisać 'Dane uczace' → 'Dalej' → 'Zakończ' → okno zwinąć do paska na dole ekranu (przy jego zamykaniu zapisać do pliku 'a-dane uczace'); zamknąć arkusz 'a-dane uczace'; każdy plik można potem otworzyć przy pomocy <Plik> <Otwórz>,
  - dla danych testowych – analogicznie (nazwa 'a-dane testowe'),
  - dla danych weryfikujących – analogicznie (nazwa 'a-dane weryfikujace').
- prezentacja w postaci graficznej obrazów uczących i weryfikujących przygotowanych w ramach projektu nr 3.

Uwaga: W związku z tym, że w ramach niniejszych zajęć przetwarza się niezależnie dwuwymiarowe dane liniowo separowalne, dwuwymiarowe dane liniowo nieseparowalne oraz obrazy, można, dla każdego z tych problemów, utworzyć – w katalogu 'SYSTEMY INTELIGENTNE' – odrębne podkatalogi.



## II. Konstruowanie drzew decyzyjnych z wykorzystaniem programu DETREEX 4.0.

- **konstrukcja maksymalnego drzewa decyzyjnego oraz wszystkich możliwych do uzyskania jego wariantów (odpowiadających różnym stopniom przycięcia drzewa maksymalnego) i odczyt parametrów umożliwiających określenie zależności procentu poprawnych decyzji systemu dla danych testowych i uczących w funkcji rozmiaru drzewa – wg poniższego scenariusza**

Uwaga: w programie DETREEX 4.0 drzewo decyzyjne można przycinać poprzez manipulowanie dwoma parametrami: Parametr 1 (*Min. liczba przykładów tworzących liść drzewa*) oraz Parametr 2 (*Przycinanie drzewa decyzyjnego*). W poniższych eksperymentach zmieniany będzie **wyłącznie Parametr 1** (ze względu na łatwość jego interpretacji); Parametr 2 nie będzie wykorzystywany.

Oznaczenia parametrów odczytywanych w programie DETREEX 4.0:

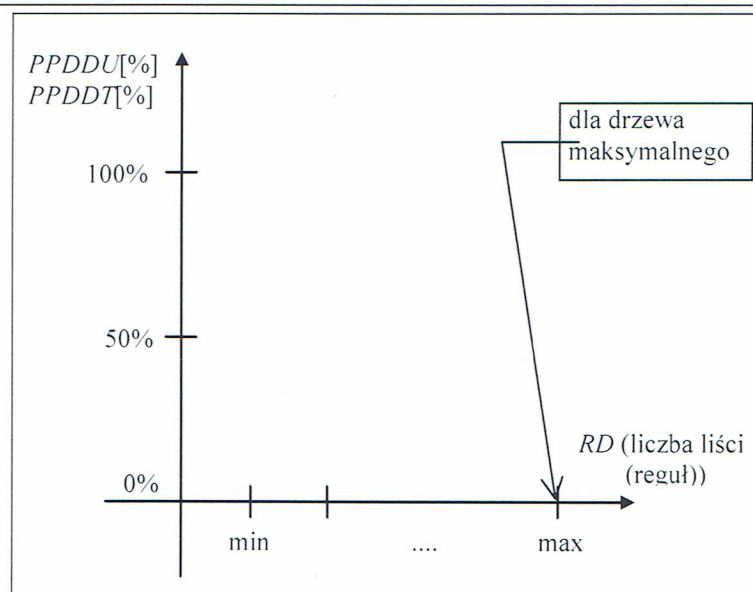
- RD* – Rozmiar Drzewa (liczba liści (reguł)),
- PPBS-DT*[%] – Procent Przykładów Błędnie Sklasyfikowanych (*Dane Testowe*),
- PPBS-DU*[%] – Procent Przykładów Błędnie Sklasyfikowanych (*Dane Uczące*), oraz parametrów wyznaczanych na ich podstawie (przez odjęcie od 100%):
- PPDDT*[%] – Procent Poprawnych Decyzji dla *Danych Testowych* (*PPDDT*[%]),
- PPDDU*[%] – Procent Poprawnych Decyzji dla *Danych Uczących*,

| Parametr 1:<br><i>Min. liczba przykładów tworzących liść drzewa</i> | Parametr 2:<br><i>Przycinanie drzewa decyzyjnego</i> | <i>PPBS-DU</i><br>[%] | <i>PPBS-DT</i><br>[%] | <i>RD</i><br>(liczba liści (reguł)) | <i>PPDDU</i><br>[%] | <i>PPDDT</i><br>[%] |
|---|--|-----------------------|-----------------------|-------------------------------------|---------------------|---------------------|
| <b>1</b>  | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |
| ↑ <i>n1</i>   | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |
| ↑ <i>n2</i>   | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |
| ↑ <i>n3</i>   | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |
| ↑ ...   | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |
| ↑ ...   | ● <i>Nie przycinaj drzewa</i>                        |                       |                       |                                     |                     |                     |

drzewo maksymalne

*n1, n2 n3,...* różne, większe od 1 liczby naturalne; odnośnie doboru tych parametrów – patrz Uwaga 1 na następnej stronie

tę część tabeli należy uporządkować wg rosnącego *RD*, usunąć identyczne wiersze i przedstawić na rys. jak niżej



**‘UWAGI’**

- 1) Nie ma potrzeby i sensu rozważania – jako  $n_1, n_2, n_3, \dots$  – kolejnych liczb naturalnych 2, 3, 4, ....  
Należy wychwycić wartości, którym odpowiadają drzewa o różnych rozmiarach  $RD$  (od drzewa maksymalnego do minimalnego) ale dla których system wciąż operuje na pełnej liczbie klas (3 klasy dla danych dwuwymiarowych i 5 klas dla obrazów).

Generalnie, im większa wartość Parametru 1 tym silniej przycięte drzewo uzyskuje się. Dla zbyt dużych wartości Parametru 1, system zmuszony do znacznego przycięcia drzewa, łączy ze sobą obiekty różnych klas (zmniejsza się liczba klas) co jest niedopuszczalne. Łatwo stąd wydedukować, że wartość Parametru 1 nie powinna być większa niż liczba rekordów dla poszczególnych klas w zbiorze uczącym.

- 2) W związku z tym, że  $RD$  jest jedynie przybliżoną miarą złożoności drzewa (nie uwzględniając liczby węzłów) mogą się zdarzyć wiersze w tabeli z poprzedniej strony o identycznych  $RD$  ale różnych  $PPDU$  i  $PPDT$  – na wykresie (patrz rysunek pod tabelą na poprzedniej stronie) należy je nanieść jako odrębne punkty.

- 3) Zawartość tabeli z poprzedniej strony najrozsądniej wpisywać na bieżąco do nowego arkusza.

**4) Zasady przeprowadzania eksperymentów z tabeli z poprzedniej strony:**

- a) ze względu na specyfikę programu DETREEX 4.0, najpierw najrozsądniej jest dokonać odczytów tylko parametru  $PPBS-DT$  (dla danych testowych) dla wszystkich rozważanych wartości Parametru 1 z pierwszej kolumny tabeli –

mając wygenerowane aktualne drzewo uzyskuje się to poprzez kliknięcie ikonki ✓,

- b) następnie należy dokonać następującego zabiegu:

plik ‘a.tst’ przekopiować do roboczego pliku ‘a.rob’,

plik ‘a.lrn’ przekopiować do pliku ‘a.tst’

(nie należy dopuścić do zmiany zawartości pliku ‘a.lrn’),

po czym należy przeprowadzić eksperymenty jak w punkcie 4a) – tym razem uzyskuje się wartości parametru  $PPBS-DU$  (dla danych uczących ale ‘podstawionych’ systemowi jako dane testowe),

na końcu plik ‘a.rob’ przekopiować do pliku ‘a.tst’.

**III. Ilustracja graficzna uzyskanych reguł decyzyjnych (dla danych dwuwymiarowych)**

- na podstawie postaci maksymalnego drzewa decyzyjnego (lub, co równoważne, zbioru uzyskanych z tego drzewa reguł), należy nanieść – na tle danych uczących i testowych – linie graniczne oddzielające obszary w których znajdują się obiekty z tych samych klas (segmentacja przestrzeni atrybutów na obszary decyzyjne);

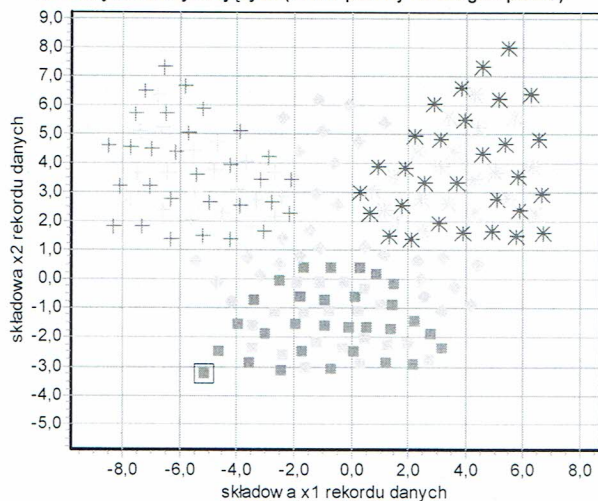
najwygodniej jest wykorzystać w tym celu możliwości, które oferuje EDYTOR DANYCH DWUWYMIAROWYCH

- przykład (na następnej stronie)

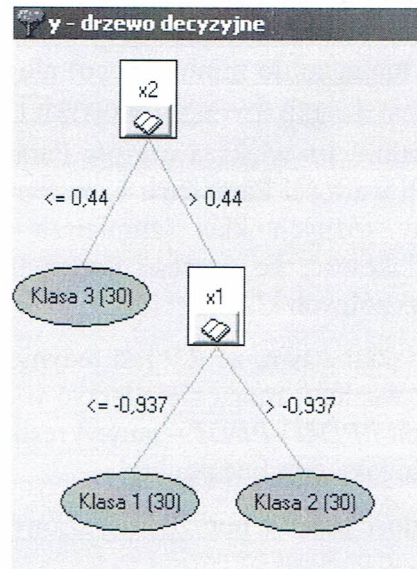


a) zbiór danych (bez 'czarnych' linii szkicujących granice pomiędzy skupiskami)

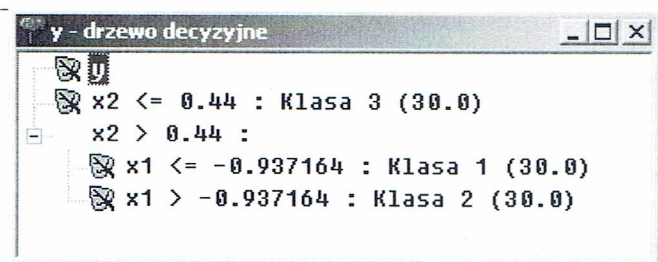
Zbiory danych: uczących (kolorowe punkty na pierwszym planie), testowych i weryfikujących (szare punkty na drugim planie)



b) maksymalne drzewo decyzyjne

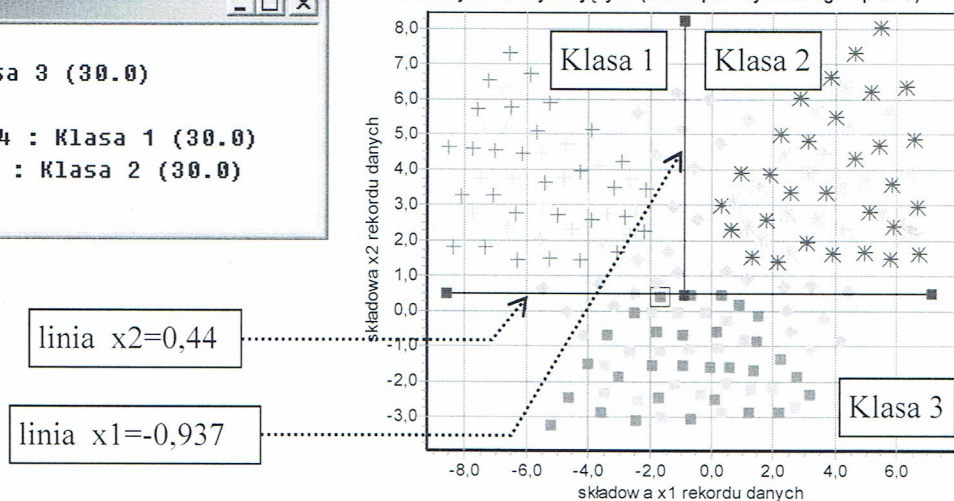


c) zbiór reguł dla drzewa decyzyjnego z rys. b)



d) ilustracja graficzna uzyskanych reguł decyzyjnych

Zbiory danych: uczących (kolorowe punkty na pierwszym planie), testowych i weryfikujących (szare punkty na drugim planie)



- w miarę możliwości przedstawić analogiczną ilustrację dla drzew decyzyjnych przyciętych w różnym stopniu

#### IV. (opcjonalnie) Wykorzystanie programu PC-SHELL 4.0 z pakietu SPHINX 4.0 do generowania odpowiedzi systemu decyzyjnego dla danych weryfikujących

- dla drzewa maksymalnego kliknąć ikonkę 'Zapisz reguły do bazy wiedzy' lub ikonkę 'Zapisz reguły do źródła wiedzy',
- kliknąć ikonkę 'Uruchom bazę/źródło wiedzy',
- zgłasza się program PC-SHELL 4.0 – podając mu wartości atrybutów opisujących dany przypadek uzyskuje się odpowiedź systemu; dodatkowo, klikając przycisk 'Jak?' uzyskuje się uzasadnienie odpowiedzi,
  - przeprowadzić eksperymenty dla szeregu danych weryfikujących.