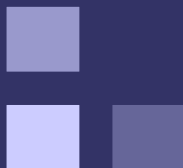


## Rozmyte drzewa decyzyjne

Łukasz Ryniewicz

Metody inteligencji obliczeniowej

21.05.2007



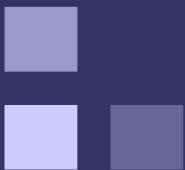
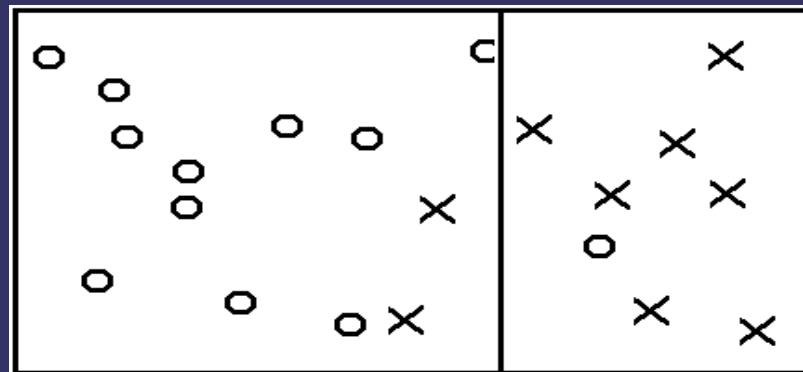
# AGENDA

- 1 Drzewa decyzyjne kontra rozmyte drzewa decyzyjne, problemy i cele
- 2 Zbiory rozmyte - teoria
- 3 Rozmyte drzewa decyzyjne, opis struktury, funkcji podziału, miara różnorodności, klasyfikacja
- 4 Wyniki badań
- 5 Modyfikacje rozmytych drzew decyzyjnych
- 6 Moje prace
- 7 Pierwsze wyniki



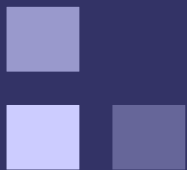
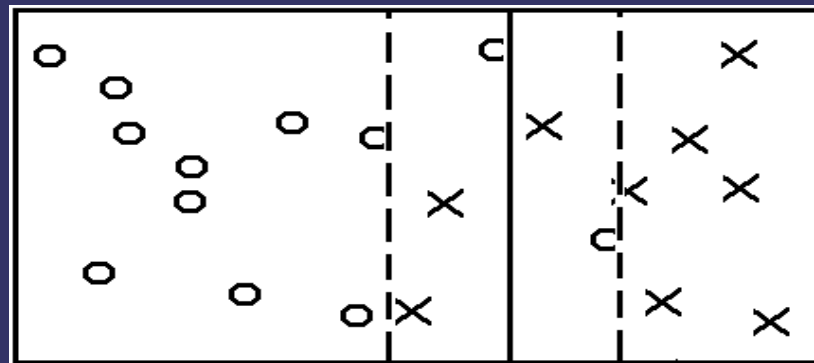
# Drzewa decyzyjne - problem

- Niewielkie zmiany w zbiorze uczącym mogą znacząco wpływać na postać drzewa
- Nieprecyzyjne dane mogą wpływać na postać drzewa
- Te cechy powodują że uzasadnione jest poszukiwanie lepszego klasyfikatora, choć sama idea drzew decyzyjnych wydaje się być słuszna

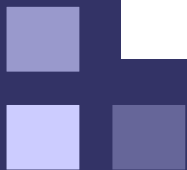
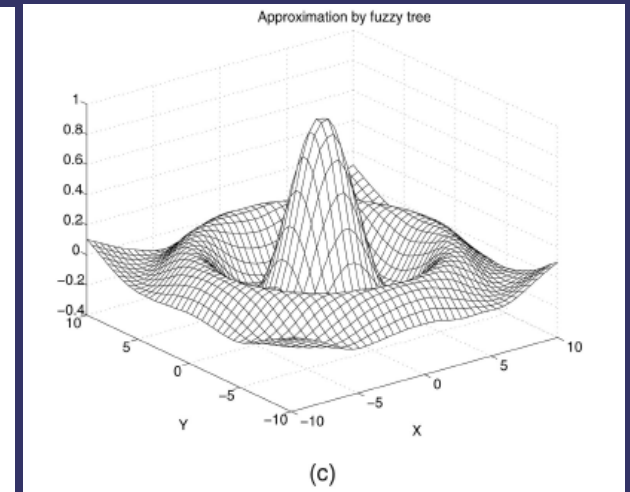
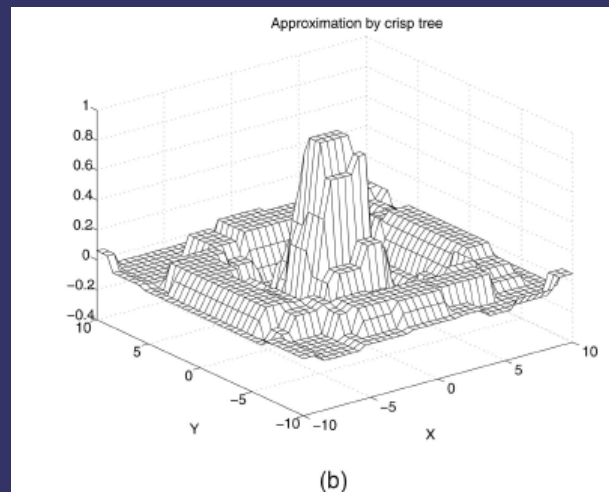
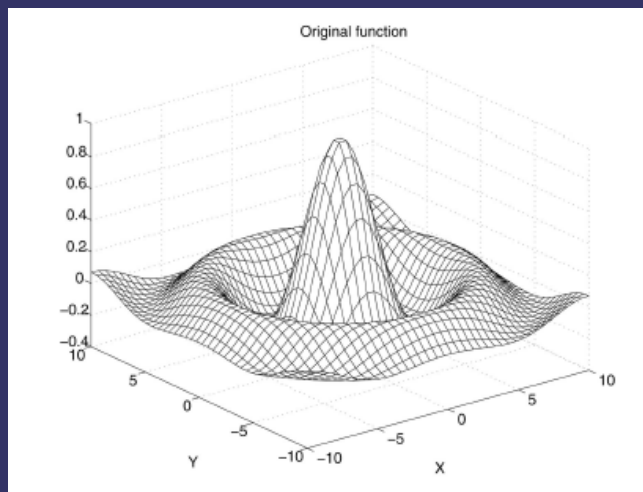
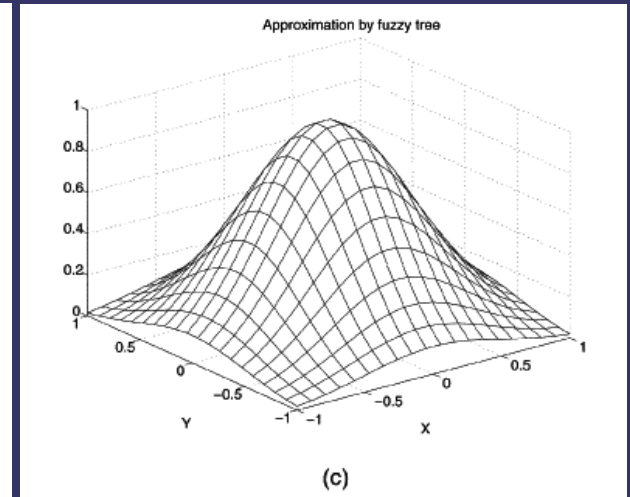
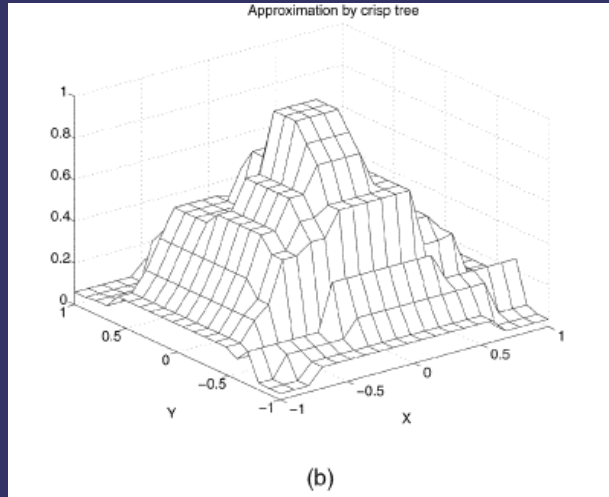
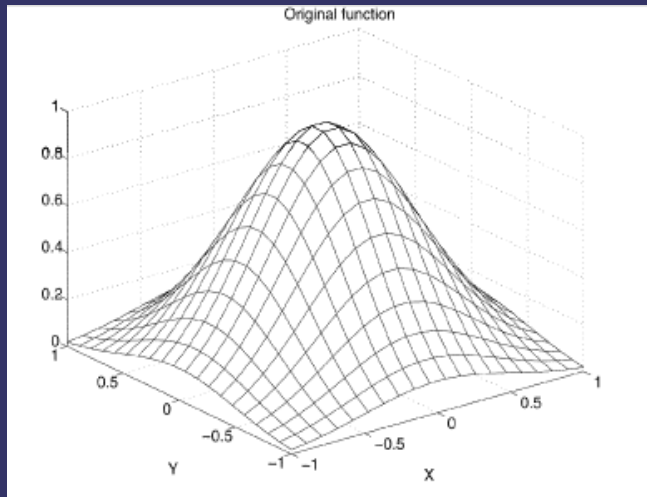


# Rozmyte drzewa decyzyjne - cel

- Uzyskanie klasyfikacji ciągłego paramertu, oraz umiejętność wyjaśnienia podjętej decyzji
- Lepsza estymacja badanego zachowania
- Bardziej „naturalna” struktura danych jaki i wyników



# Rozmyte drzewa decyzyjne

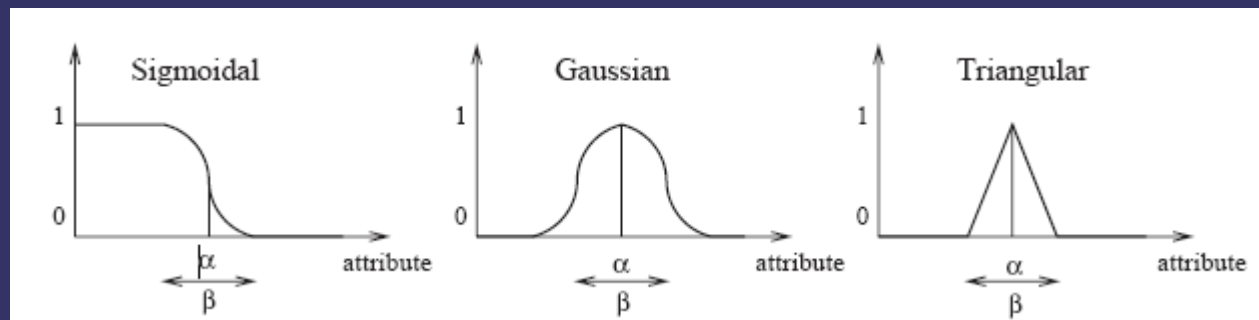


# Zbiory rozmyte - teoria

- Zbiór rozmyty  $A$  jest to pewnien „zwykły” zbiór  $X$  oraz stowarzyszona z nim funkcja przynależności:

$$u_A(x) : X \rightarrow [0,1]$$

- Zwykle jest charakteryzowany przez dwa parametry:  $\alpha$ ,  $\beta$ , będące odpowiednio parametrem podziału i rozmycia



# Zbiory rozmyte - teoria

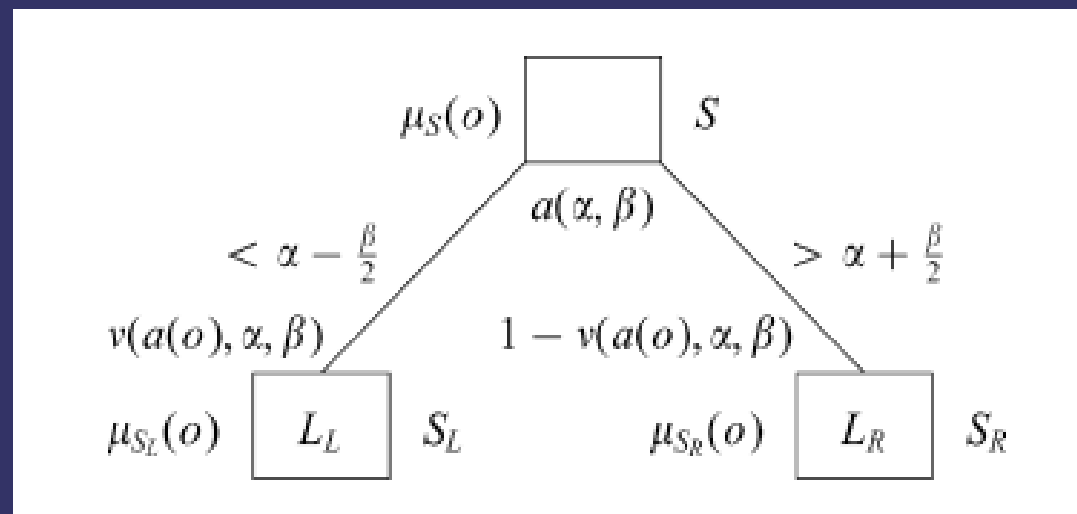
## Podstawowe operacje

- Suma: 
$$u_{A \cup B} = \underset{x \in X}{MAX} \{u_A(x), u_B(x)\}$$
- Przecięcie: 
$$u_{A \cap B} = \underset{x}{MIN} \{u_A(x), u_B(x)\}$$
- Dopełnienie: 
$$u_{A'}(x) = 1 - u_A(x)$$
- Równość: 
$$A = B \Leftrightarrow u_A(x) = u_B(x), x \in X$$
- Zawieranie: 
$$A \subset B \Leftrightarrow u_A(x) \leq u_B(x), x \in X$$
- Nośnik: 
$$supp(A) = \{x \in X : u_A(x) > 0\}$$



# Rozmyte drzewa decyzyjne – opis struktury

- Proces klasyfikacji analogicznie jak w przypadku zwykłych drzew (budowa, przycinanie, itd)
- W każdym wierzchołku znajduje się pewien zbiór obserwacji, lecz jest przedstawiony jako zbiór rozmyty
- Zwykle dla korzenia przyjmuje się zbiór pewny, tj taki dla którego wartość funkcji przynależności jest równa 1

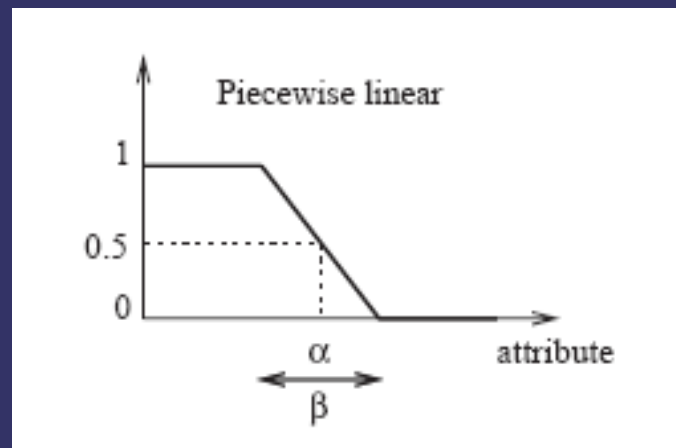




# Rozmyte drzewa decyzyjne – funkcja podziału

Podział jest określany na podstawie funkcji przynależności  $v$  dla której estymujemy parametry podziału i rozmycia (przypadek liniowej funkcji podziału, badane są też funkcje logistyczne i sigmoidalne)

- Obserwacje  $< \alpha - \beta/2$  trafiają do lewego dziecka
- Obserwacje  $> \alpha + \beta/2$  trafiają do prawego dziecka
- Pozostałe do obydwu wraz z określeniem funkcji przynależności odpowiednio:  $v(a(x), \alpha, \beta)$  oraz  $1 - v(a(x), \alpha, \beta)$



# Rozmyte drzewa decyzyjne - podziały

- Wartość funkcji przynależności do węzłów dzieci:

$$u_{S_L}(x) = u_S(x) v(a(x), a, \beta)$$

$$u_{S_R}(x) = u_S(x) (1 - v(a(x), a, \beta))$$

- Minimalizacja błędu średniokwadratowego przy każdym podziale:

$$E_S = \sum_{x \in S} u_S(x) (u_C(x) - \hat{u}_C(x))^2$$

$$\hat{u}_C(x) = v(a(x), a, \beta) L_L + (1 - v(a(x), a, \beta)) L_R$$



# Rozmyte drzewa decyzyjne – miary rozróżnialności klas

- Rozmyta licznosc zbioru  $S$ :

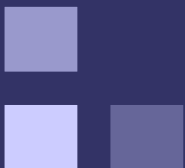
$$M(S) = \sum_{x \in S} u_S(x)$$

- Wsparcie zbioru dla  $i$ -tej klasy:

$$f_i(S) = \frac{M(S \cap C_i)}{M(S)}$$

- Entropia:

$$H(S) = \sum_{i=1,2,..k} f_i(S) \ln(f_i(S))$$



# Rozmyte drzewa decyzyjne - klasyfikacja

Klasyfikacja odbywa się analogicznie do zwykłych drzew, jednak każda obserwacja może leżeć w kilku liściach

- Gdy jest pewne że obserwacja należy do danego liścia  $L$  tzn gdy:

$$u_L(x) = 1$$

wtedy:

$$\hat{u}_c(x) = C(\underset{i}{\operatorname{argmax}} f_i(L))$$

- Przy tak konstruowanym drzewie powyższy warunek dyskryminuje przynależność danej obserwacji tylko do jednego liścia



# Rozmyte drzewa decyzyjne - klasyfikacja

W przypadku gdy dla każdego liścia:

$$u_L(x) < 1$$

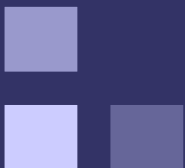
Z reguły przyjmuje się jeno z poniższych

- Klasyfikacje do liścia dla którego funkcja przynależności jest największa

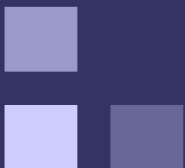
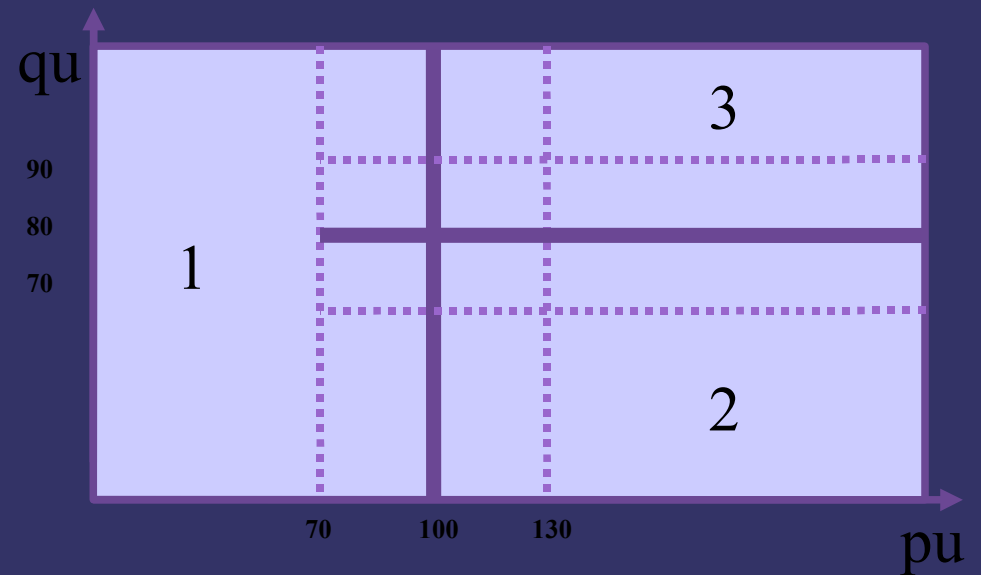
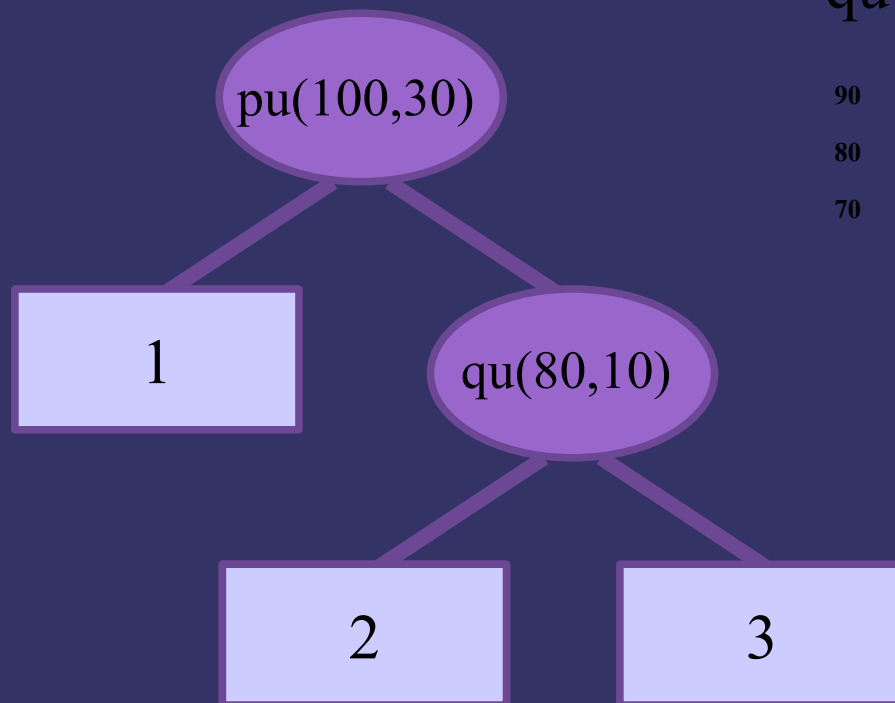
$$\hat{u}_C(x) = \sum_{Li \in Liscie} I(\max_{Lj \in Liscie} u_{Lj}(x) = u_{Li}(x)) * C(\underset{k}{argmax} f_k(Li))$$

- Obliczenie średniej ważonej z wartościami funkcji przynależności jako wagami

$$\hat{u}_C = \sum_{Li \in Liscie} C(\underset{k}{argmax} f_k(Li)) * u_{Li}(x)$$

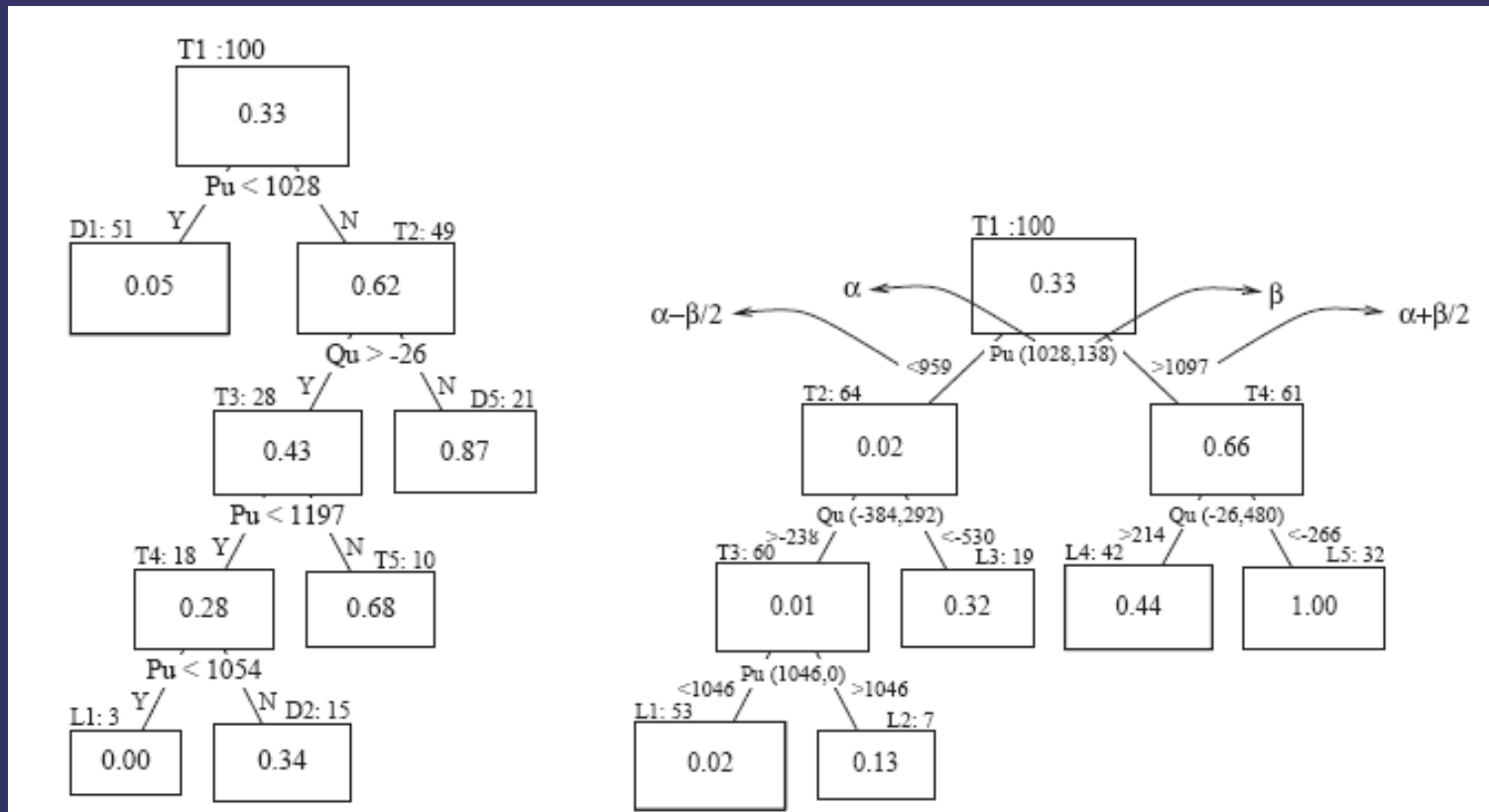


# Rozmyte drzewa decyzyjne - przykład



# Rozmyte drzewa decyzyjne - przykład

Często zbudowane rozmyte drzewo decyzyjne ma inną postać niż zwykłe



## Badane zbiory:

- Omib - baza danych o sieciach elektrycznych, 6 zmiennych ciągłych opisowych, jedna ciągła decyzyjna decyzyjna
- Twonorm - baza wygenerowana na podstawie wielowymiarowego rozkładu normalnego, 10 zmiennych opisowych, dwie klasy decyzyjne
- Waveform - baza z opisami kształtów fal, 21 zmiennych ciągłych, trzy klasy decyzyjne
- Do wyników brana średnia z 20 drzew
- Funkcja podziału - liniowa
- Warunki stopu: znajdowanie się w liściu mniej niż 1% obserwacji z oryginalnego zbioru, dla danego liścia błąd średniokwadratowy mniejszy niż  $10^{-4}$





# Rozmyte drzewa decyzyjne - wyniki

Comparing complexity and classification error rate

DB	GS size	Complexity				Error rate (%)				
		C4.5	ULG	CART	SDT	C4.5	ULG	CART	SDT (R)	SDT (B)
Omib	50	3.8	1.4	3.4	20.2	19.5	21.6	19.2	12.2	11.6
	250	12.4	7.7	22.2	52.1	11.2	11.9	11.0	4.9	4.7
	500	20.1	12.2	45.9	52.5	10.1	10.5	9.2	4.6	4.2
	1000	33.8	24.1	93.0	54.9	8.3	8.1	7.6	4.2	4.0
	1500	46.0	37.5	164.0	60.2	7.4	7.9	7.1	4.3	3.9
	2000	56.6	40.4	223.8	59.2	7.0	7.0	6.5	4.4	4.1
		50	3.6	1.6	2.9	9.8	28.7	32.3	29.3	24.7
	250	14.6	9.1	13.2	53.1	22.9	24.1	23.2	14.8	14.7
Two Norm	500	28.1	17.5	26.1	71.5	20.4	22.4	20.7	10.3	10.5
	750	39.4	33.7	38.2	81.7	19.0	19.9	19.9	8.3	7.6
	1000	51.9	41.7	53.9	83.4	18.1	19.8	19.0	7.3	6.1
	1500	74.5	56.4	77.9	88.7	17.4	19.5	18.5	6.4	4.1
		50	9	2	2	2	48.0	36.0	29.5	25.6
	250	22	10	4	25	26.4	28.4	26.3	21.0	21.9
Wave form	500	43	13	12	24	33.4	28.8	23.9	19.9	19.2
	1000	66	50	26	44	26.1	26.3	25.4	18.0	17.8
	1500	113	21	35	38	25.9	26.3	21.6	17.5	16.8
	2000	135	25	52	50	24.0	28.0	23.4	17.9	16.6

# Rozmyte drzewa decyzyjne - wyniki

SDT accuracy and complexity (before and after pruning)							
DB	GS size	Error rate (%)				Complexity	
		G	P	R	B	Before	After
Omib	50	12.5	12.9	12.2	11.6	61.0	20.2
	250	7.5	8.1	4.9	4.7	110.0	52.1
	500	6.7	7.1	4.6	4.2	113.1	52.5
	1000	6.0	6.7	4.2	4.0	121.7	54.9
	1500	5.5	6.3	4.3	3.9	124.2	60.2
	2000	5.5	6.3	4.4	4.1	126.2	59.2
	50	23.6	25.6	24.7	24.8	41.9	9.8
	250	16.1	17.1	14.8	14.7	159.0	53.1
Twonorm	500	13.2	13.6	10.3	10.5	229.4	71.5
	750	11.2	11.2	8.3	7.6	256.7	81.7
	1000	9.7	10.0	7.3	6.1	254.1	83.4
	1500	8.9	9.1	6.4	4.1	245.3	88.7
	50	30.7	25.5	25.6	28.2	58	2
	250	23.2	22.4	21.0	21.9	104	25
Waveform	500	22.5	21.0	19.9	19.2	170	24
	1000	19.6	19.6	18.0	17.8	192	44
	1500	20.9	18.8	17.5	16.8	176	38
	2000	19.7	18.5	17.9	16.6	201	50

# Rozmyte drzewa decyzyjne - wyniki

Comparing CPU times						
DB	GS size	CPU time CART	ULG	SDT(R)	SDT(B)	cy
Omib	50	1.4 s	0.8 s	9.0 s	10.0 s	18
	250	2.5 s	1.7 s	25.0 s	2.2 m	32
	500	6.2 s	2.5 s	40.0 s	5.9 m	50
	1000	10.2 s	4.3 s	1.3 m	6.8 m	23
	1500	16.3 s	6.3 s	2.0 m	6.6 m	28
	2000	26.0 s	7.6 s	3.3 m	13.1 m	31
	50	0.5 s	1.0 s	8.7 s	25.0 s	51
	250	1.1 s	2.5 s	36.6 s	2.6 m	51
Twonorm	500	2.0 s	4.4 s	1.7 m	9.6 m	50
	750	2.5 s	6.6 s	3.4 m	10.0 m	50
	1000	5.4 s	11.6 s	4.2 m	14.6 m	50
	1500	5.6 s	16.0 s	9.8 m	22.2 m	46
	50	1.5 s	1.0 s	12.0 s	16.9 s	22
	250	3.0 s	2.5 s	44.5 s	3.3 m	41
Waveform	500	10.0 s	5.3 s	2.0 m	3.1 m	18
	1000	10.8 s	9.2 s	4.1 m	14.2 m	31
	1500	15.4 s	14.0 s	6.4 m	12.6 m	22
	2000	29.8 s	18.7 s	9.6 m	28.1 m	29

# Rozmyte drzewa decyzyjne - wyniki

Table 8  
Datasets—Part II

Dataset	# Attributes	# Classes	# Samples
Balance	4	3	625
Glass	9	6	214
Heart	13	2	270
Ionosphere	34	2	351
Iris	4	3	150
Monks-1	6	2	432
Monks-2	6	2	432
Monks-3	6	2	432
Pima	8	2	768
Sonar	60	2	208
Wine	13	3	178

# Rozmyte drzewa decyzyjne - wyniki

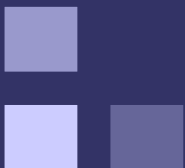
Table 9

Classification error rates and standard deviations for a 10 times 10-fold cross-validation

Database	C4.5	ULG	CART	SDT (R)
Balance	(-)21.74 ± 0.8	(-)22.29 ± 0.7	(-)21.01 ± 0.8	14.33 ± 0.5
Glass	(-)31.42 ± 2.0	31.22 ± 2.4	(-)32.27 ± 1.6	28.91 ± 1.5
Heart	(+)21.98 ± 2.1	(-)27.78 ± 2.0	(-)27.19 ± 2.0	25.81 ± 2.0
Ionosphere	10.65 ± 1.3	9.96 ± 0.8	11.05 ± 0.9	10.56 ± 1.0
Iris	4.94 ± 0.6	(-)6.13 ± 0.8	(-)6.47 ± 0.8	4.73 ± 0.9
Monks-1	(-)25.01 ± 0.0	(-)5.75 ± 2.0	(-)11.65 ± 3.2	17.56 ± 4.3
Monks-2	(-)10.56 ± 1.5	(-)3.01 ± 1.3	(-)3.15 ± 0.8	2.61 ± 1.4
Monks-3	0.00 ± 0.0	0.00 ± 0.0	0.00 ± 0.0	0.00 ± 0.0
Pima	25.68 ± 1.7	(-)30.22 ± 0.9	(-)29.89 ± 1.2	26.43 ± 1.3
Sonar	26.45 ± 2.3	(+)25.27 ± 2.9	(-)29.19 ± 2.0	26.72 ± 1.9
Wine	(-)7.27 ± 1.1	(-)6.58 ± 1.1	(-)10.66 ± 1.4	3.52 ± 0.8

# Rozmyte drzewa decyzyjne – dalsze prace

- Przedstawienie klas decyzyjnych, danych wejściowych w postaci zbiorów rozmytych
- Wprowadzono miary rozmycia każdej obserwacji po funkcjach przynależności do atrybutów (dla każdego termu w atrybucie jest określona funkcja przynależności)
- Miara jendoznaczności wartości atrybutu (w przypadku atrybutów które są zbiorami rozmytymi jest to istotna informacja)



# Rozmyte drzewa decyzyjne – dalsze prace

- Okazują się bardzo dobrym narzędziem do tłumaczenia decyzji podejmowanej przez sieci neuronowe
- Sieci neuronowe są w nich wykorzystywane do określania funkcji podziału w węzłach
- Otrzymywano lepsze wyniki w uzupełnianiu braków danych niż w przypadku zwykłych drzew
- Stworzono efektywne metody rozmywania

■ zwykłych drzew



# Moje prace

- Dostarczyć pełny formalny matematyczny opis rozmytych drzew decyzyjnych
- Wykazać lepszą jakość klasyfikacji (mniejszy błąd średniokwadratowy) dla atrybutów ciągłych niż w przypadku zwykłych drzew
- Dodatkowo, skonstruowanie hybrydowego algorytmu budowy rozmytego drzewa decyzyjnego opartego o:
  - Algorytm C4.5
  - Miarą rozróżnienia klas jest entropia
  - Funkcja podziału przedstawiana za pomocą regresji logistycznej, której parametry są wyznaczone na podstawie obserwacji znajdujących się w danym węźle





# Wyniki

Testy były prowadzone dla podziału: 40:30:30, wyniki to średnia dla 10 drzew. Przeprowadziłem pierwsze testy na zbiorach: INSURANCE - zbiór z 49 cechami określającymi klienta, oraz binarną zmienną decyzyjną - czy zakupi dany produkt ubezpieczeniowy, ponad 19000 obserwacji

Wyniki:

	DT	FDT
%Error	26,29	22,84

WINE - zbiór z informacjami na temat oceny jakości wina, 13 atrybutów opisujących, trzy klasy decyzyjne, 178 obserwacji

Wyniki:

	DT	FDT
%Error	8,11	8,02



# Wyniki

GLASS - zbór z informacjami o typach szkła, 10 atrybutów opisujących, sześć klas decyzyjnych, 214 obserwacji

Wyniki:

	DT	FDT
%Error	34,67	31,86



# Bibiografia

- [1] Cristina Olaru, Louis Wehenkel, *A complete fuzzy decision tree technique*, Fuzzy Sets and Systems, 138 (2003) 221-254
- [2] I-Jen Chiang, Jane Yung-jen Hsu, *Fuzzy classification trees for data analysis*, Fuzzy Sets and Systems 130 (2002) 87 – 99
- [3] Eyke Hullermeier, *Fuzzy sets in machine learning and data mining*, (2008) preprint Applied Soft Computing
- [4] Motohide Umano, Hirotaka Okamoto, Itsuo Hatono, *Fuzzy Decision Trees by Fuzzy ID3 Algorithm and Its Application to Diagnosis Systems*, Proceedings of The Third IEEE Conference on Fuzzy Systems, Vol. 3, June 26–29, Orlando, FL, 1994, pp. 2113–2118.
- [5] E. C. C. Tsang, X. Z. Wang, and D. S. Yeung, *Improving Learning Accuracy of Fuzzy Decision Trees by Hybrid Neural Networks*, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 8, NO. 5, OCTOBER 2000, 601-613



# Bibiografia

- [6] Cezary Z. Janikow, *Fuzzy Decision Trees: Issues and Methods*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS —PART B: CYBERNETICS, VOL. 28, NO. 1, FEBRUARY 1998, 1-14
- [7] Alberto Suarez, James F. Lutsko, *Globally Optimal Fuzzy Decision Trees for Classification and Regression*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 21, NO. 12, DECEMBER 1999, 1297 – 1311
- [8] Yufei Yuan, Michael J. Shaw, *Induction of fuzzy decisions trees*, Fuzzy Sets and Systems 69 (1995) 125-139

