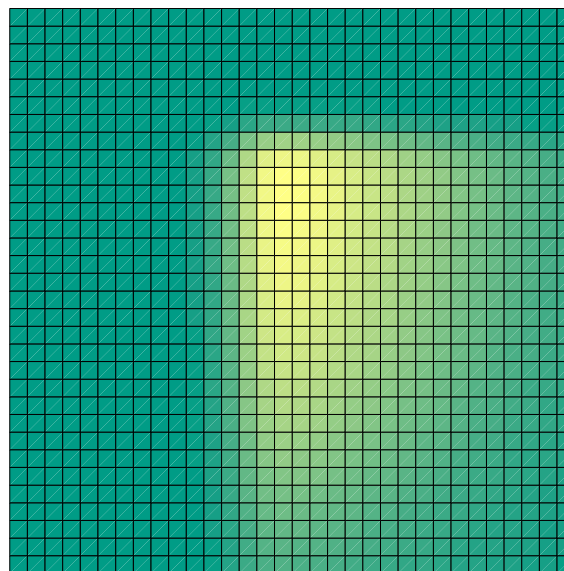


N o r b e r t J a n k o w s k i

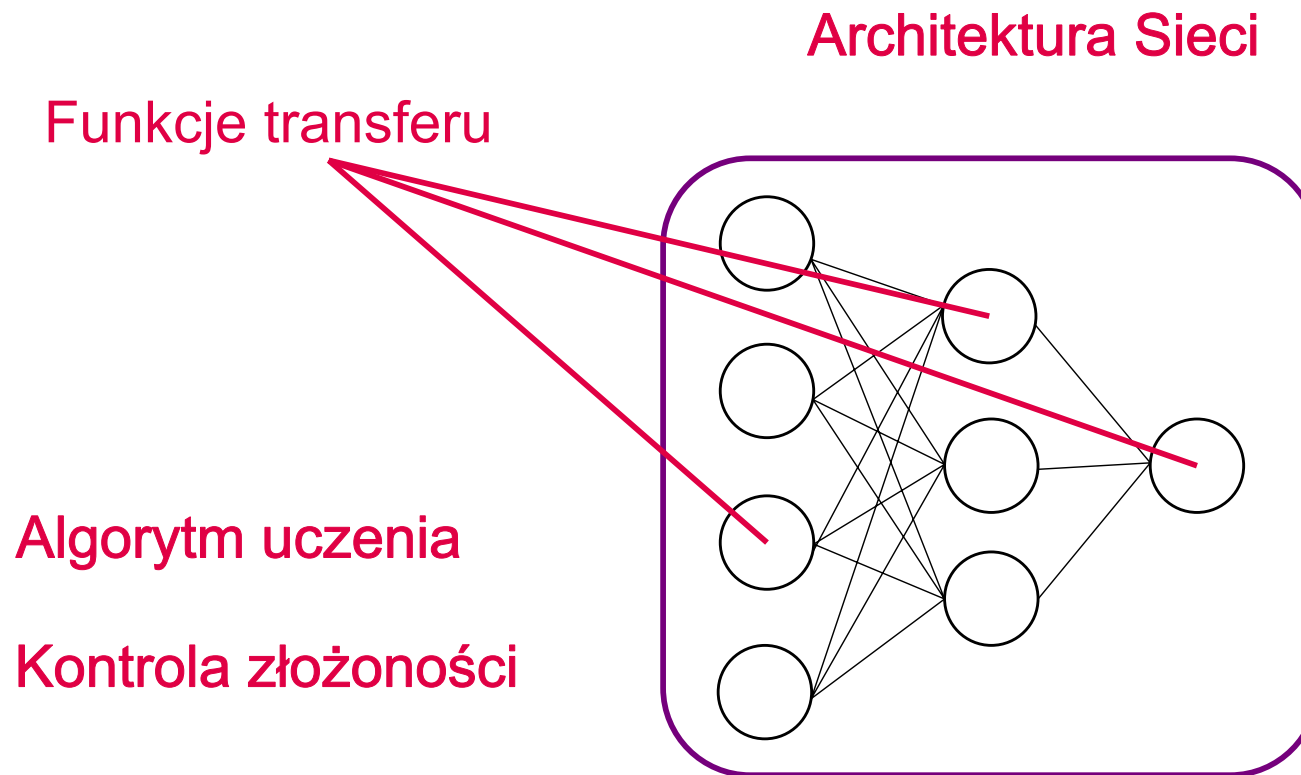
Ontogeniczne sieci neuronowe
w zastosowaniu do
klasyfikacji danych medycznych



Cele pracy

- Zbudowanie ontogenicznej sieci neuronowej
- Interpretacja sposobu działania sieci neuronowych
- Zastosowanie do klasyfikacji i analizy danych medycznych

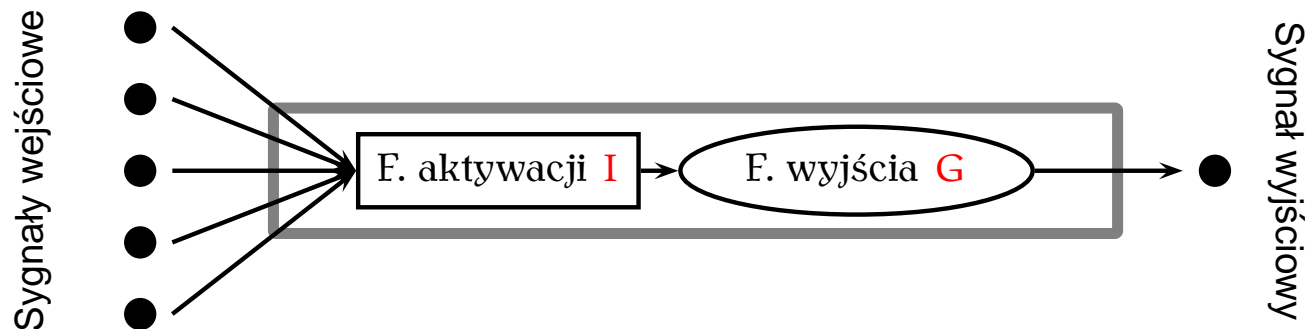
Sieci neuronowe — Modele Adaptacyjne



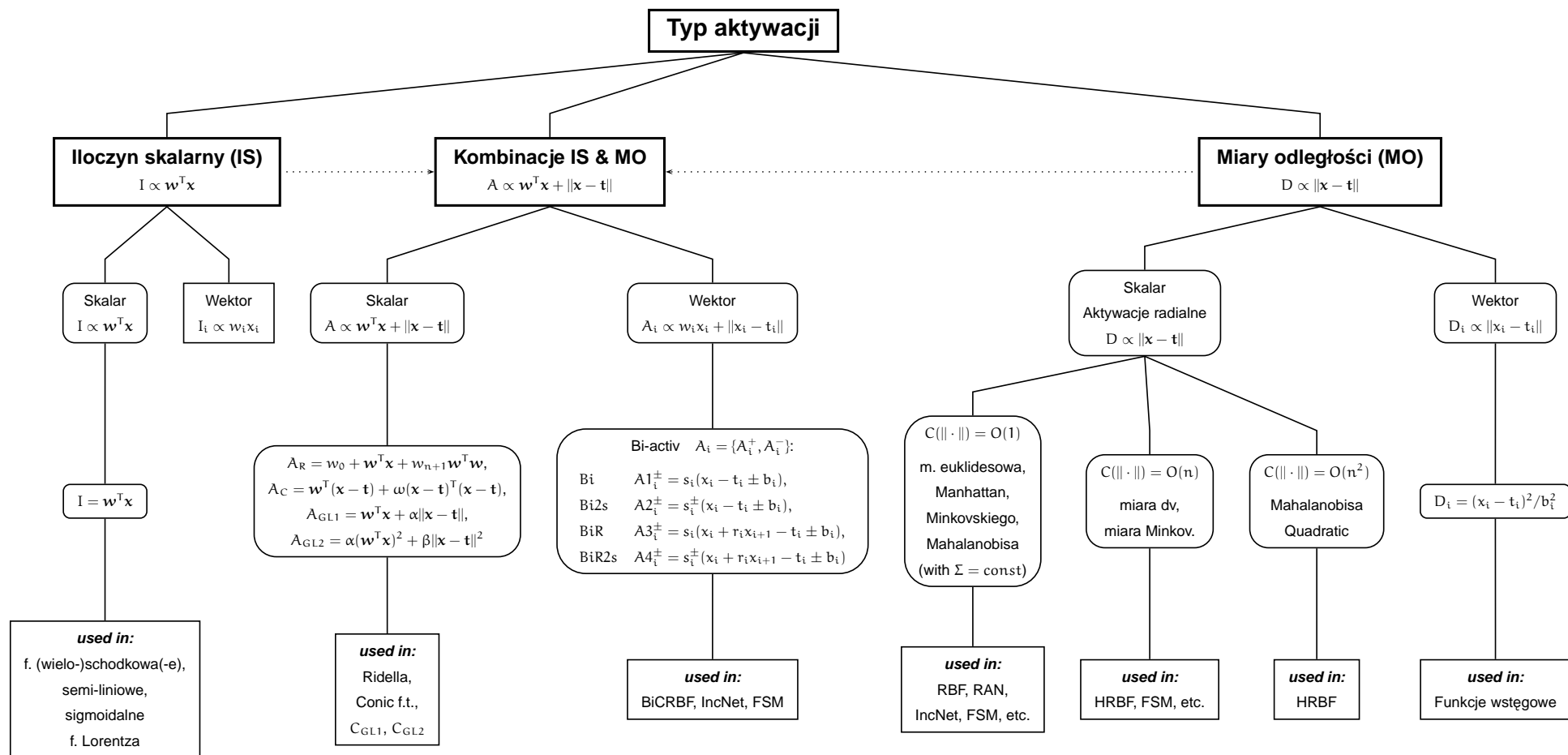
= **Możliwości modelu**

Funkcje transferu

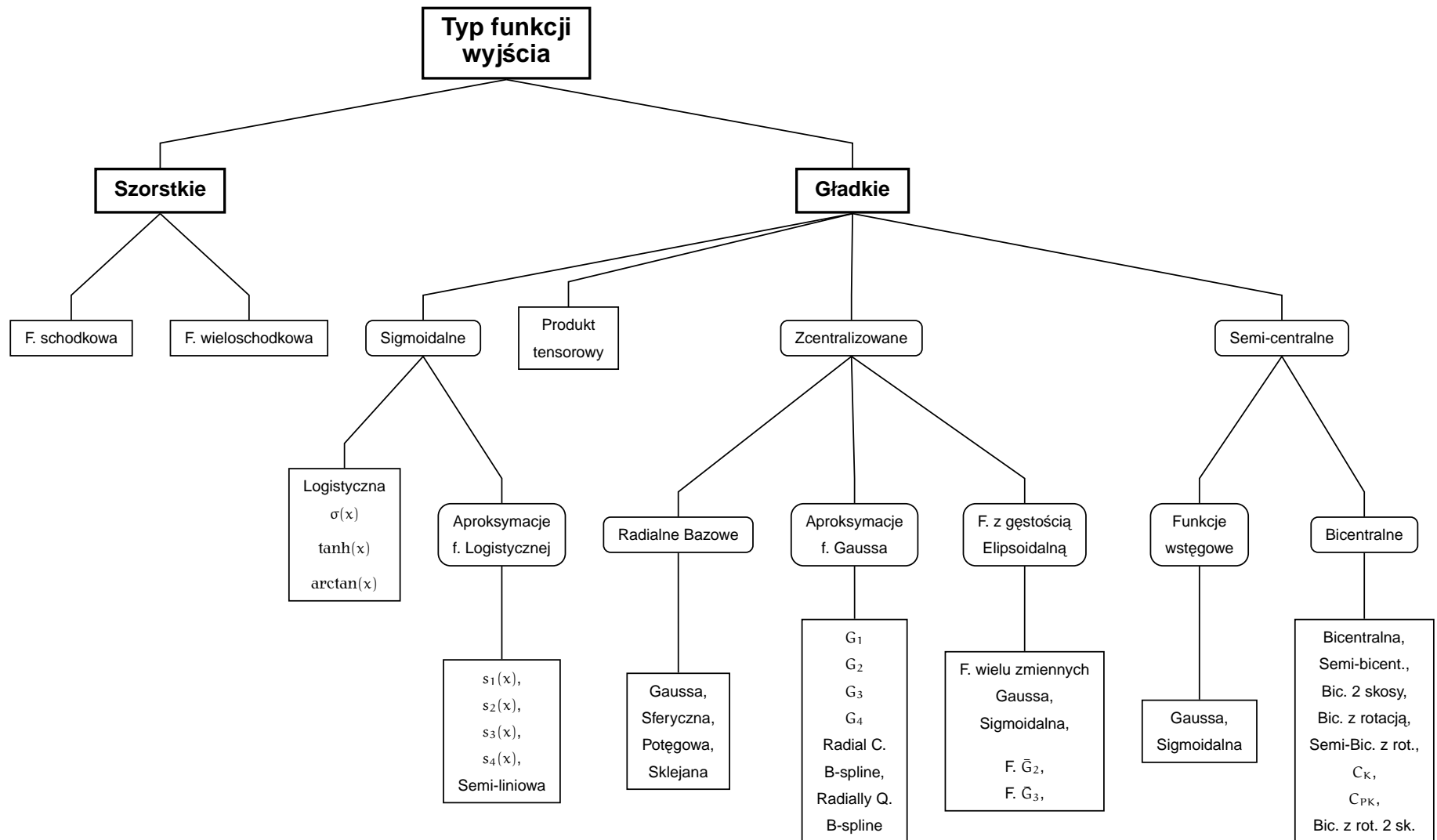
- W sieciach MLP — funkcje **sigmoidalne** $1/(1 + e^{-\mathbf{x}^T \mathbf{w}})$
- W sieciach RBF — funkcje **gaussowskie** $e^{-\|\mathbf{x} - \mathbf{t}\|^2}$



$$G(I(\mathbf{x}))$$



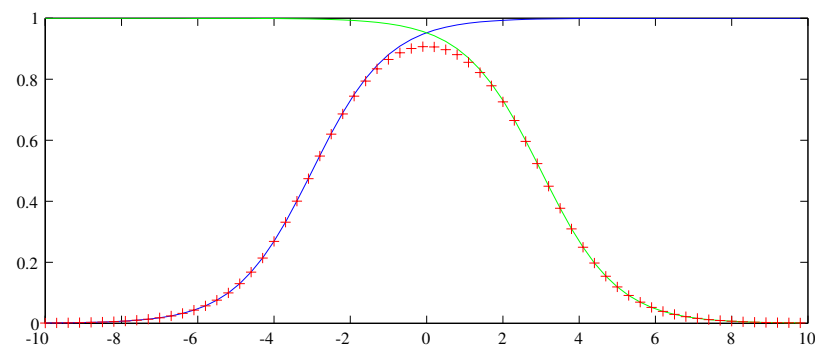
Rysunek 1: Taksonomia funkcji aktywacji.



Rysunek 2: Taksonomia funkcji wyjścia.

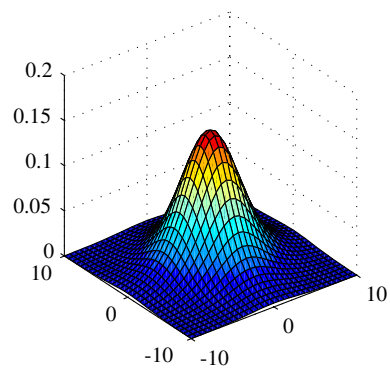
Funkcje bicentralne

$$\begin{aligned}
 \text{Bi}(\mathbf{x}; \mathbf{t}, \mathbf{b}, \mathbf{s}) &= \prod_{i=1}^N \sigma(A1_i^+) (1 - \sigma(A1_i^-)) \\
 &= \prod_{i=1}^N \sigma(e^{s_i} \cdot (x_i - t_i + e^{b_i})) (1 - \sigma(e^{s_i} \cdot (x_i - t_i - e^{b_i})))
 \end{aligned} \tag{1}$$

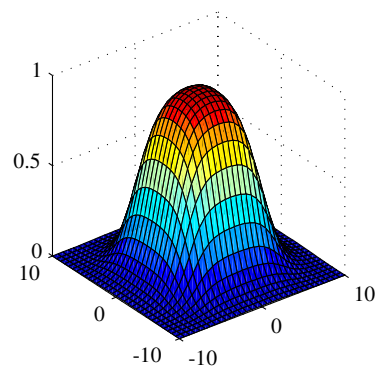


Funkcje Bicentralne

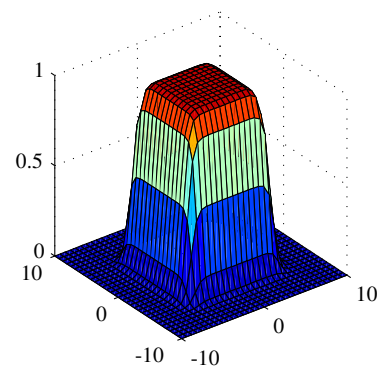
Rozmycie 1 1 Skos .5 .5



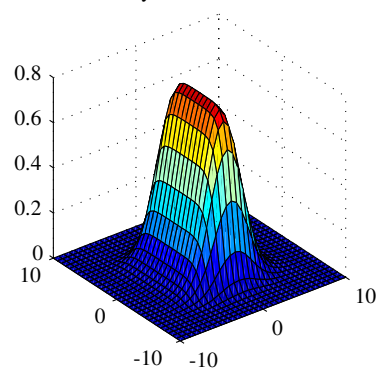
Rozmycie 5 5 Skos 1 1



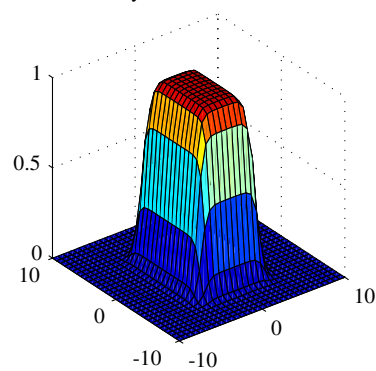
Rozmycie 5 5 Skos 3 3



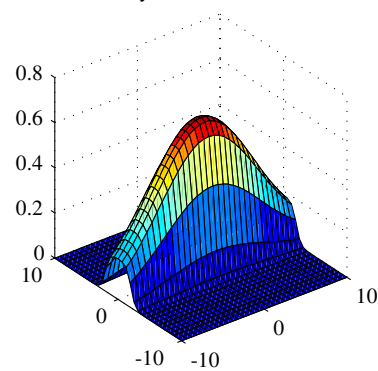
Rozmycie 2 5 Skos 1 2



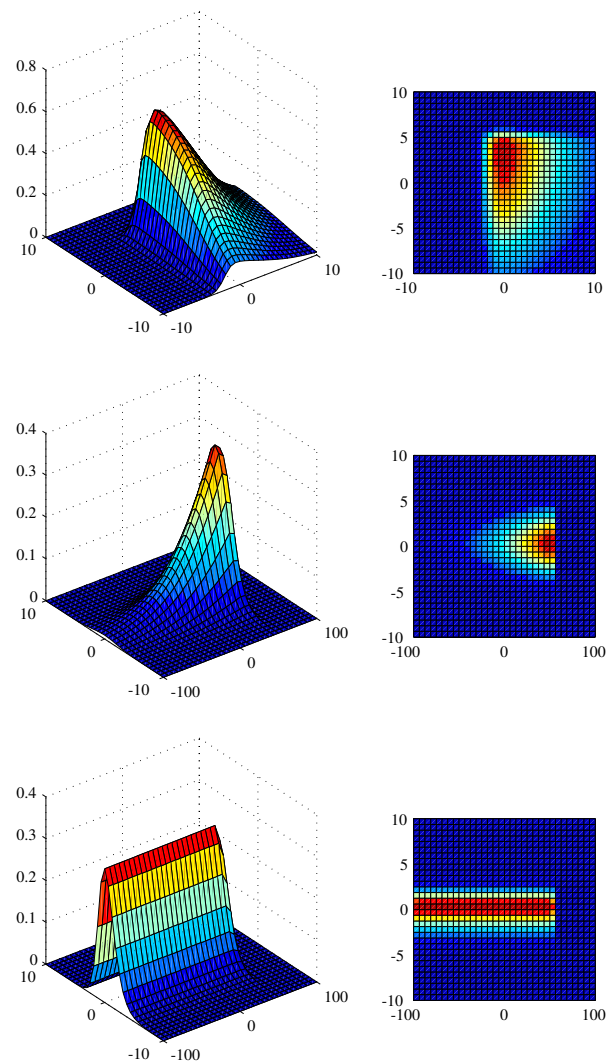
Rozmycie 3.5 5 Skos 3 3



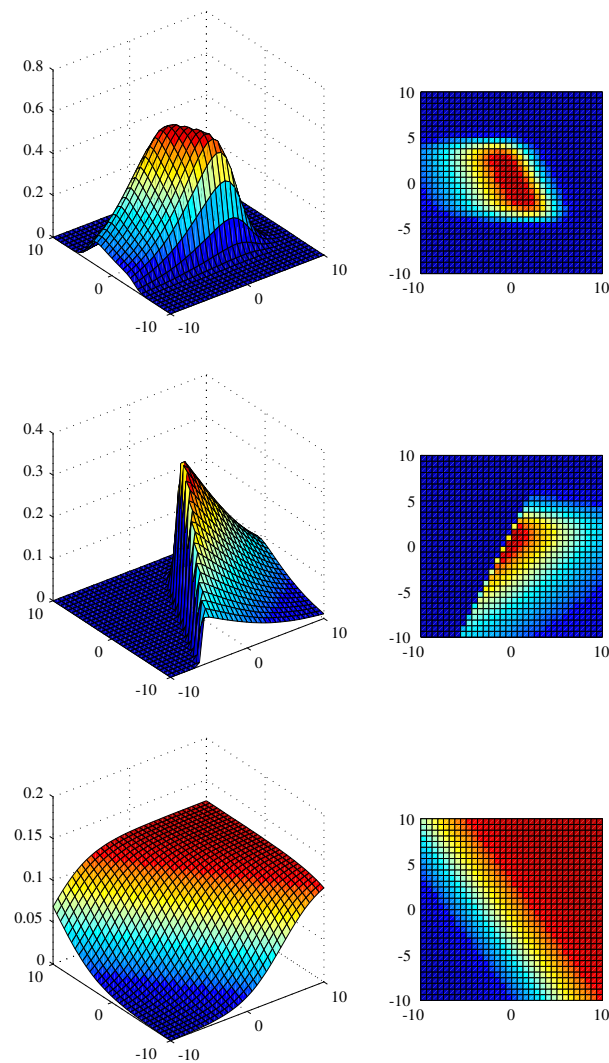
Rozmycie 5 2 Skos .3 3



Funkcje Bicentralne z dwoma skosami



Funkcje Bicentralne z rotacją i dwoma skosami



Model Adaptacyjny

Bicentralne f. transferu

+

Architektura sieci

+

Algorytm uczenia

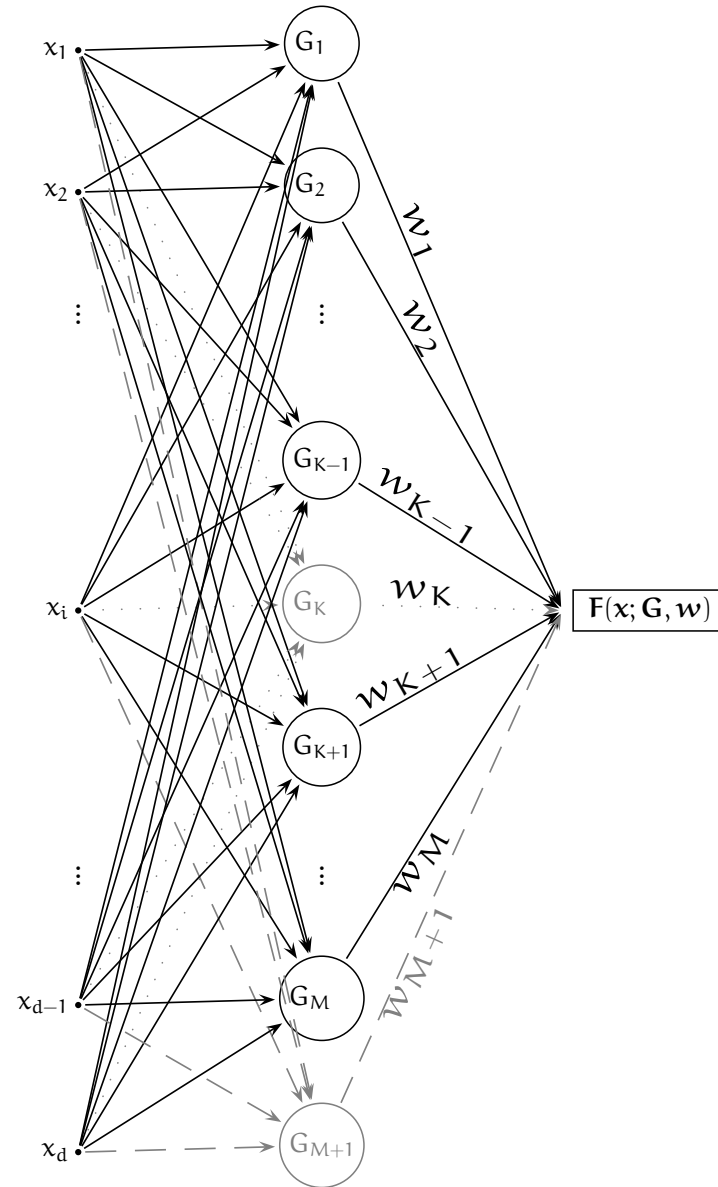
+

Kontrola złożoności

Architektura sieci

Radial Basis Function

$$f(\mathbf{x}; \mathbf{w}, \mathbf{p}) = \sum_{i=1}^M w_i G_i(\mathbf{x}, \mathbf{p}_i).$$



Model Adaptacyjny

Bicentralne f. transferu

+

Architektura RBF

+

Algorytm uczenia

+

Kontrola złożoności

Algorytm uczenia

Uczenie sekwencyjne

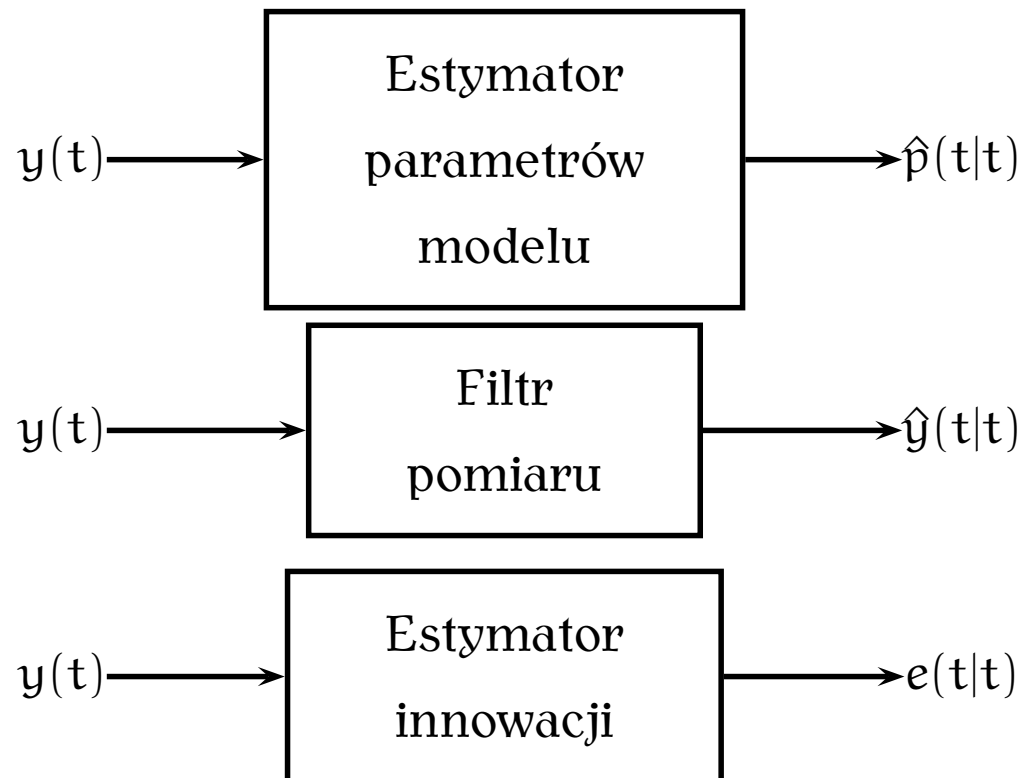
Minimalizacja funkcji:

$$E_{sq} = \int_{\mathcal{D}} |F^{(n)}(\mathbf{x}) - F^{(n-1)}(\mathbf{x})|^2 p^{(n-1)} d\mathbf{x} + \frac{1}{n-1} |y_n - F^{(n)}(\mathbf{x}_n)|^2$$

gdzie $p^{(n-1)}$ jest rozkładem prawdopodobieństwa ostatnich $n-1$ obserwacji (wektorów wejściowych) [Kadirkamanathan].

Rozszerzony Filtr Kalmana (EKF)

Minimalizacja funkcji E_{sq} błędu może być aproksymowana przez użycie rozszerzonej wersji algorytmu filtru Kalmana (EKF)



Plusy EKF

- każdy parametr uczony jest niezależnie
- uczenie jest szybkie (l. epok)
- oszczędność neuronów
- regularyzacja
- estymowane parametry EKF są przydatne w kontroli złożoności

Stworzono szybką wersję EKF: złożoność P_n : $O(M^2)$ redukuje się do $O(M)$

Model Adaptacyjny

Bicentralne f. transferu

+

Struktura RBF

+

Filtr Kalmana

+

Dodawanie/Usuwanie/Łączenie
neuronów

Cel uczenia vs. złożoność modelu

CEL: Jak najlepsza GENERALIZACJA

Złożoność Modelu \approx Złożoność Problemu (Danych)

Złożoność Modelu = Architektura + Funkcje Transferu

Kontrola złożoności

- Dodawanie nowych neuronów
- Usuwanie zbędnych neuronów
- Łączenie neuronów

Kryterium wystarczalności modelu

Kiedy sieć jest wystarczalna, aby akumulować nową informację?

- Popęłniony błąd vs. niepewność modelu

$$\frac{\text{błąd}}{\text{niepewność modelu}} < \alpha(M)$$

Hipotezę \mathcal{H}_0 , iż aktualny model jest wystarczający, można zapisać jako:

$$\mathcal{H}_0 : \frac{e^2}{\text{Var}[f(\mathbf{x}; \mathbf{p}) + \eta]} = \frac{e^2}{\sigma_y^2(\mathbf{x}) + \sigma_{n_s}^2} = \frac{e_n^2}{R_y} < \chi_{M, \theta}^2$$

gdzie $\chi_{M, \theta}^2$ jest rozkładem chi-kwadrat z progiem ufności $\theta\%$ i M stopniami swobody

Usuwanie neuronów

Faworyzowanie silnych wag (wagi o istotnym wpływie), których zmian wartości są małe (wagi nauczone)

$$\frac{\text{waga}_i}{\text{zmienność wagi}_i} \quad L_1 = \min_i \frac{w_i^2}{\sigma_{w_i}} = \min_i \frac{w_i^2}{[\mathbf{P}_w]_{ii}}$$

Czy w ogóle usuwać? — najczęściej nieistotne neurony występują w modelach niestabilnych, a z kolei wartościowe neurony mamy w dobrze nauczonych, stabilnych sieciach:

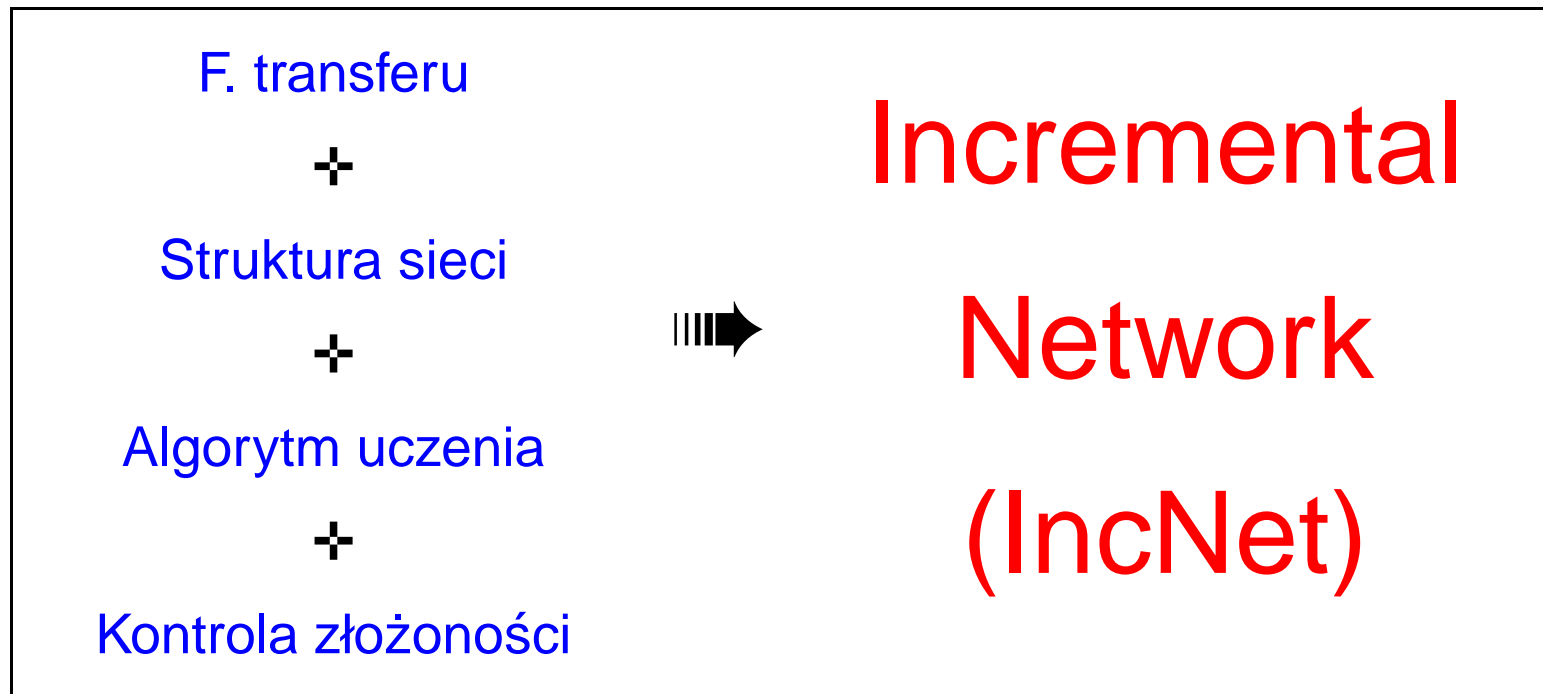
$$\frac{L_1}{R_y} < \chi^2_{1,\vartheta} \quad (2)$$

gdzie $\chi^2_{n,\vartheta}$ jest rozkładem chi-kwadrat z poziomem ufności $\vartheta\%$ z jednym stopniem swobody.

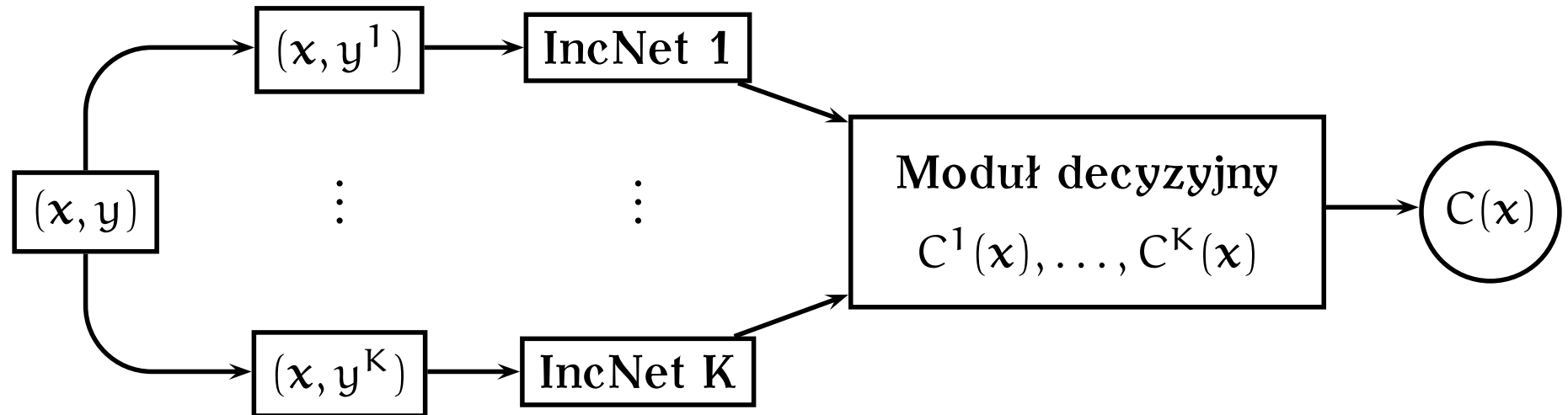
Łączenie neuronów

$$\frac{\int_{\mathbf{d} \subseteq \mathcal{D}^n} |\bar{\Phi}_i(\mathbf{x}) + \bar{\Phi}_j(\mathbf{x}) - \bar{\Phi}_{new}(\mathbf{x})| \, d\mathbf{x}}{\int_{\mathbf{d} \subseteq \mathcal{D}^n} |\bar{\Phi}_i(\mathbf{x}) + \bar{\Phi}_j(\mathbf{x})| \, d\mathbf{x}} < \alpha \quad (3)$$

Model Adaptacyjny



Sieć IncNet w klasyfikacji



Klasa zwycięska:

$$C(\mathbf{x}) = \arg \max_i C^i(\mathbf{x})$$

Prawdopodobieństwo:

$$p(C^i|\mathbf{x}) = \frac{\sigma(C^i(\mathbf{x}) - \frac{1}{2})}{\sum_{j=1}^K \sigma(C^j(\mathbf{x}) - \frac{1}{2})}$$

Interpretacja — Przedziały ufności

Dla wektora \mathbf{x} przedział $[x_{\min}^r, x_{\max}^r]$ dla cechy r wyznaczony jest przez:

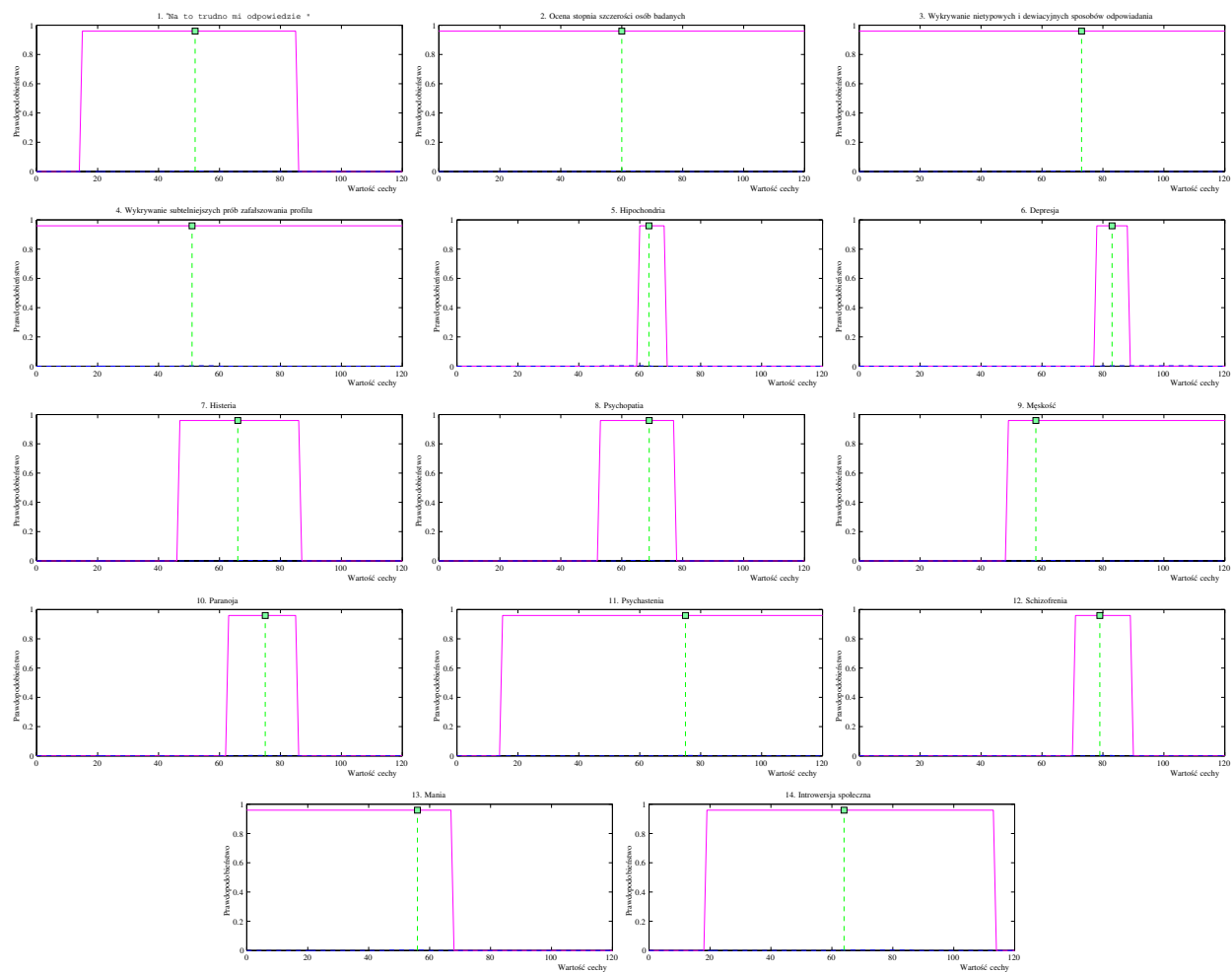
$$x_{\min}^r = \min_{\bar{\mathbf{x}}} \{C(\bar{\mathbf{x}}) = k \wedge \forall_{\mathbf{x}_r > \hat{\mathbf{x}} > \bar{\mathbf{x}}} C(\hat{\mathbf{x}}) = k\}$$

$$x_{\max}^r = \max_{\bar{\mathbf{x}}} \{C(\bar{\mathbf{x}}) = k \wedge \forall_{\mathbf{x}_r < \hat{\mathbf{x}} < \bar{\mathbf{x}}} C(\hat{\mathbf{x}}) = k\}$$

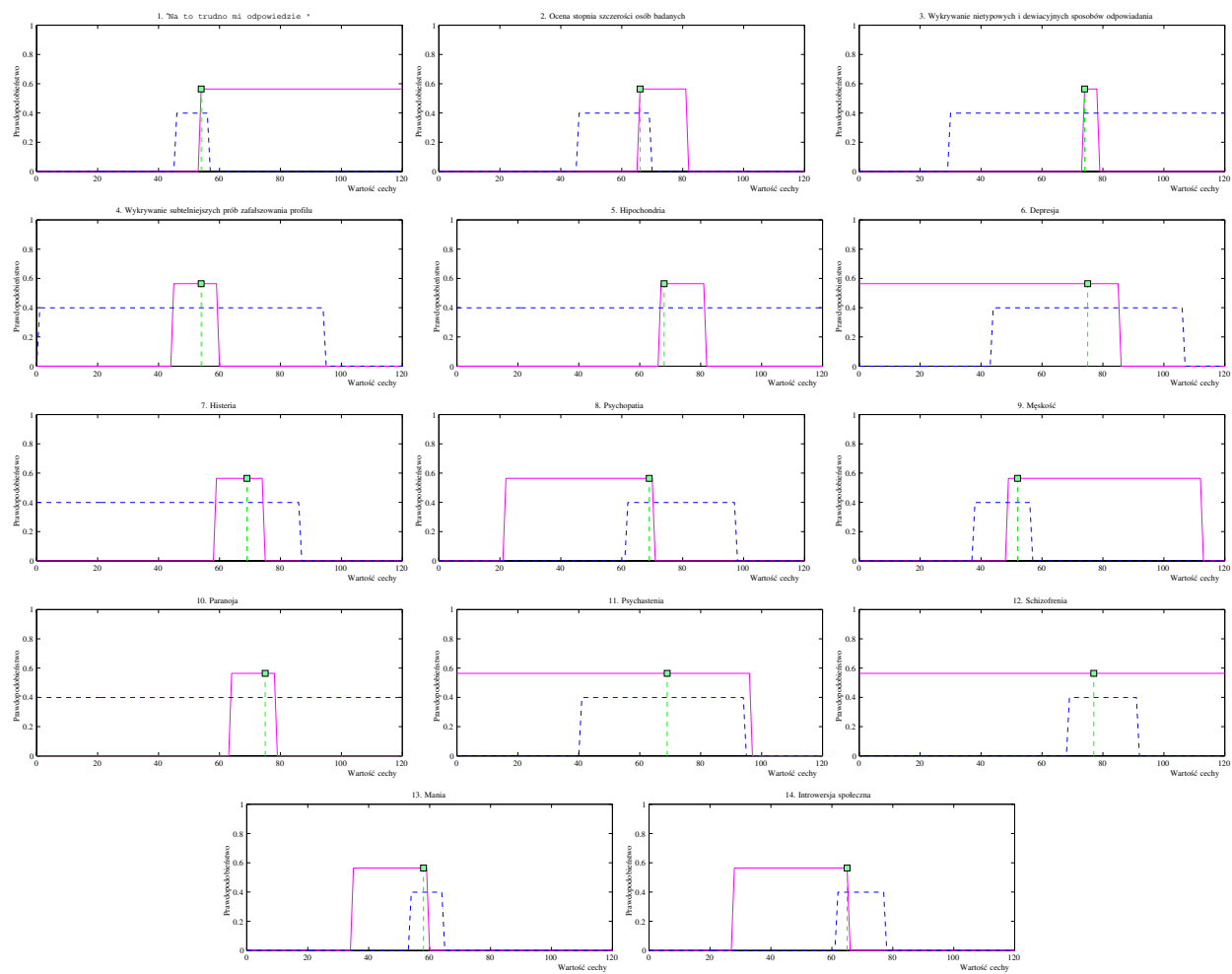
$\bar{\mathbf{x}} = [x_1, \dots, x_{r-1}, \bar{x}, x_{r+1}, \dots, x_N]$, a

$\hat{\mathbf{x}} = [x_1, \dots, x_{r-1}, \hat{x}, x_{r+1}, \dots, x_N]$.

- Uogólnienie o próg ufności.



Rysunek 3: Psychoza reaktywna.



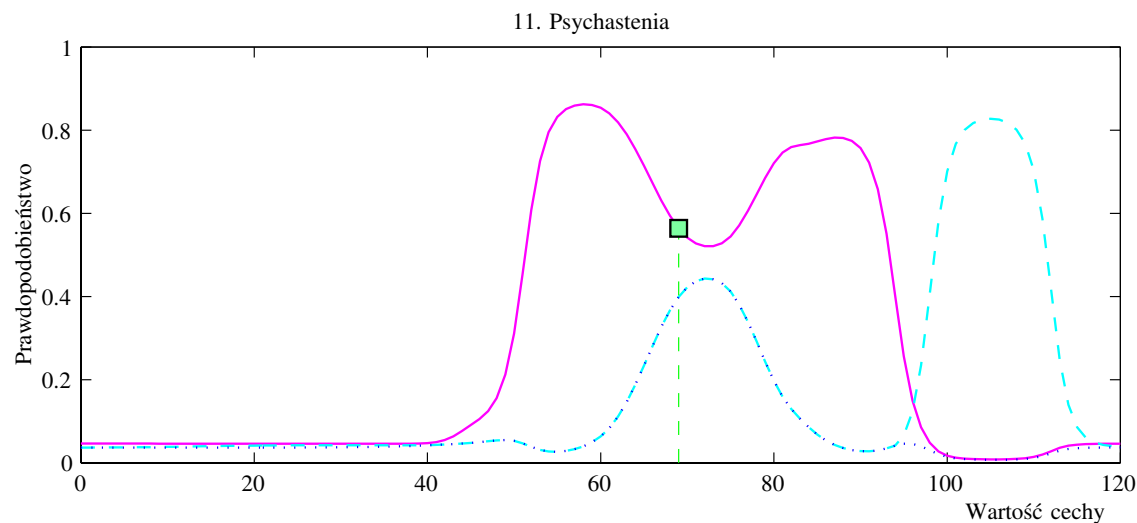
Rysunek 4: Zmiany organiczne – schizofrenia.

Interpretacja — Probabilistyczne p. ufności

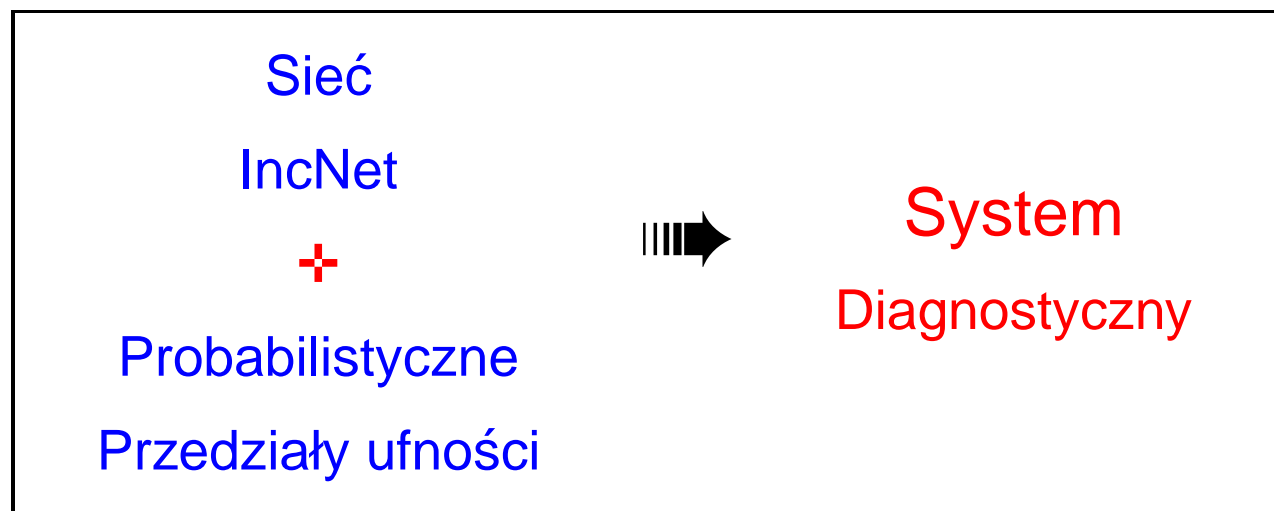
Dla wektora \mathbf{x} i każdego wymiaru r :

Klasa	Prawdopodobieństwo.	# klasy
zwycięska	$p(C(\mathbf{x}) \bar{\mathbf{x}}(z))$	$C(\mathbf{x})$
alternatywna I	$p(C^{k_2} \bar{\mathbf{x}}(z))$	$k_2 = \arg \max_i \{p(C^i \mathbf{x}), C^i \neq C(\mathbf{x})\}$
alternatywna II	$p(C^{k_M} \bar{\mathbf{x}}(z))$	$k_M = \arg \max_i \{p(C^i \bar{\mathbf{x}}(z)), C^i \neq C(\mathbf{x})\}$

$$\bar{\mathbf{x}}(z) = [x_1, \dots, x_{r-1}, z, x_{r+1}, \dots, x_N]$$



Interpretacja wyników System diagnostyczny



Zastosowania sieci IncNet do klasyfikacji i analizy danych medycznych

- Klasyfikacja danych psychometrycznych
- Zapalenie wyrostka robaczkowego
- Dane dotyczące raka piersi
- Dane dotyczące zapalenia wątroby
- Dane dotyczące cukrzycy
- Choroby tarczycy

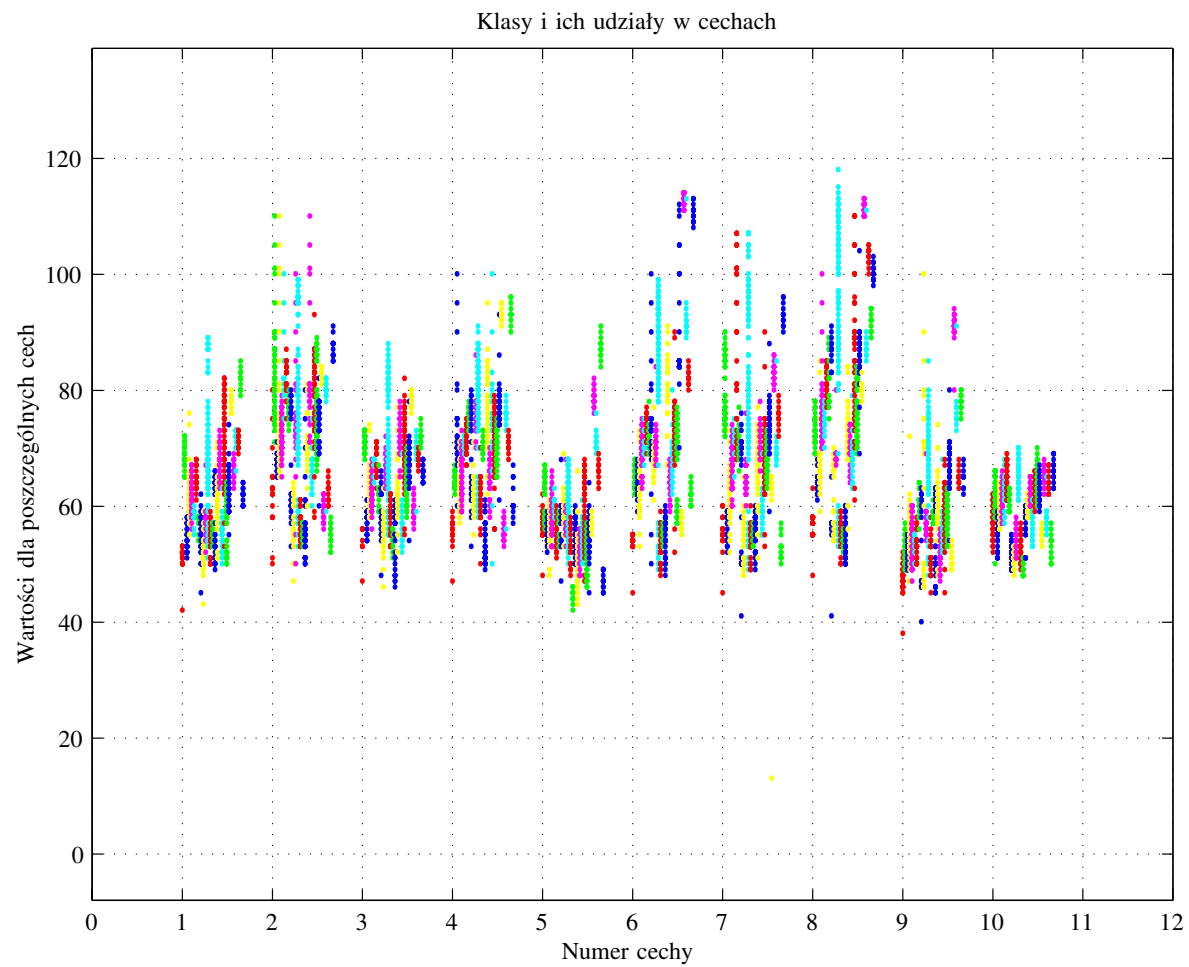
- Problemy aproksymacyjne

Klasyfikacja danych psychometrycznych

- Test psychometryczny Minnesota Multiphasic Personality Inventory
- Test składa się z ponad 550 pytań
- 550 pytań ➡ 14 skal (kontrolne i kliniczne)
hipochondria, depresja, histeria, psychopatia męskość, paranoja, psychastenia, schizofrenia, mania, introwersja społeczna
- 27–28 klas nozologicznych
norma, nerwica, psychopatia, schizofrenia, z. urojeniowy, psychozy, paranoja, alkoholizm, etc.

Bazy danych

Baza	l. wektorów	rozmiar wej.	l. klas
Pierwsza baza	1027	14	27
Druga baza	1167	14	28



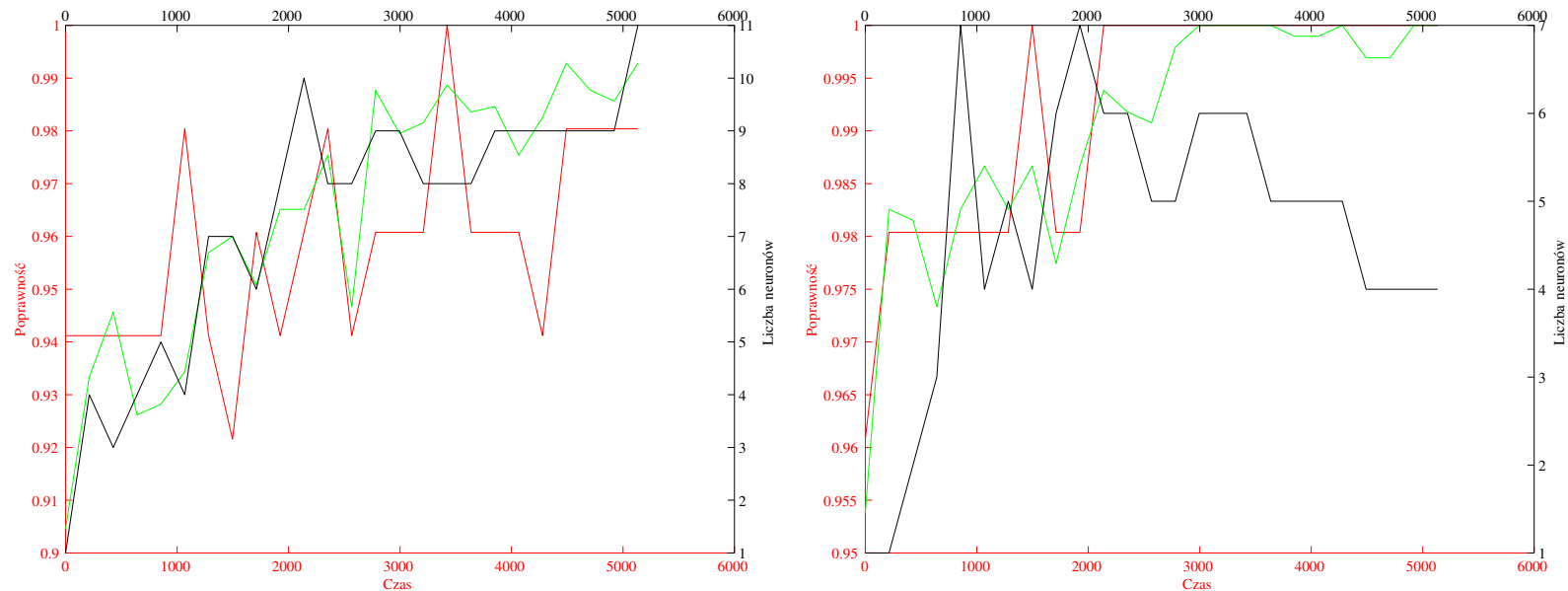
Rysunek 5: Pierwsza baza — skale kliniczne.

Rozkład złożoności podsieci

Liczby neuronów w poszczególnych podsięciach																										
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
5	1	4	9	9	6	4	3	8	8	3	11	4	1	8	4	5	8	12	8	1	2	2	1	1	1	1
Całkowita liczba neuronów: 130																										

Tabela 1: Rozkład złożoności sieci IncNet dla zbioru 27 klasowego.

Sieci były uczone przez 5 epok.



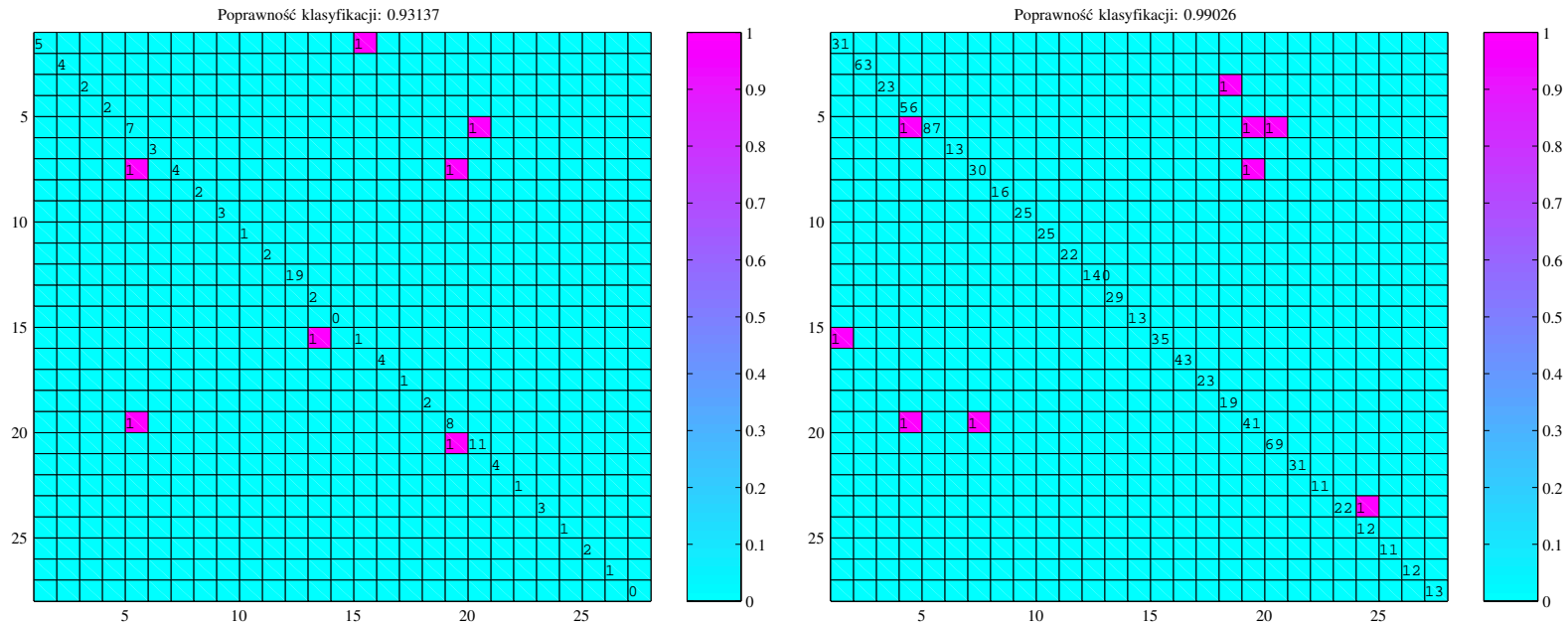
Rysunek 6: Poprawność klasyfikacji dla zbioru treningowego (kolor zielony) i zbioru testowego (kolor czerwony), jak i liczbę neuronów (kolor czarny). Dane dla klasy 5-tej i 16-tej

Porównanie wyników

Model	Uczenie na całym zbiorze	
	27 klasowym	28 klasowym
IncNet	99.22	99.23
C 4.5	93.67	93.06
FSM+R Opt.	97.57	96.91

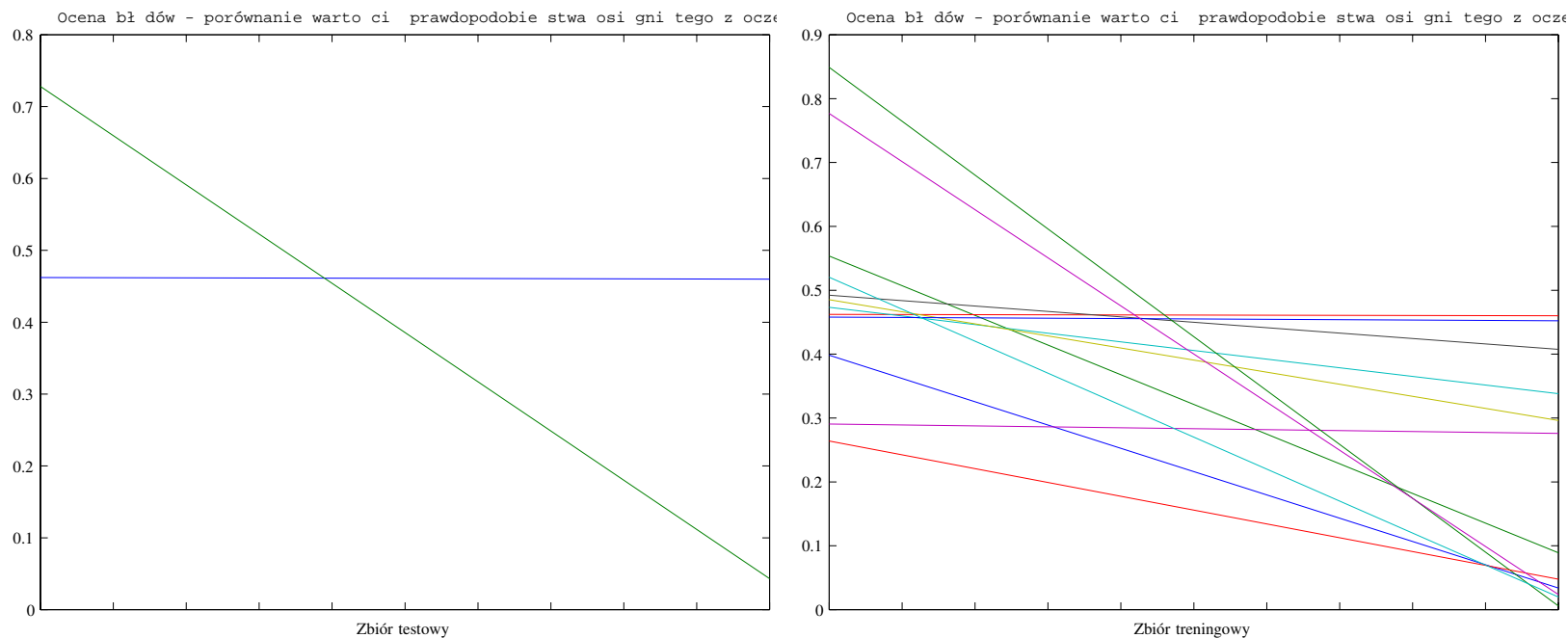
Model	zbiór 27-klasowy				zbiór 28-klasowy			
	10%		5%		10%		5%	
	TRS	TES	TRS	TES	TRS	TES	TRS	TES
IncNet	99.03	93.14	98.77	96.08	98.95	93.10	98.29	94.83
FSM + G	97.65	92.16			97.08	90.20		
FSM + R	96.97	84.70			96.66	82.76		
C 4.5	93.22	83.70			93.13	78.90		

Macierze rozrzutu

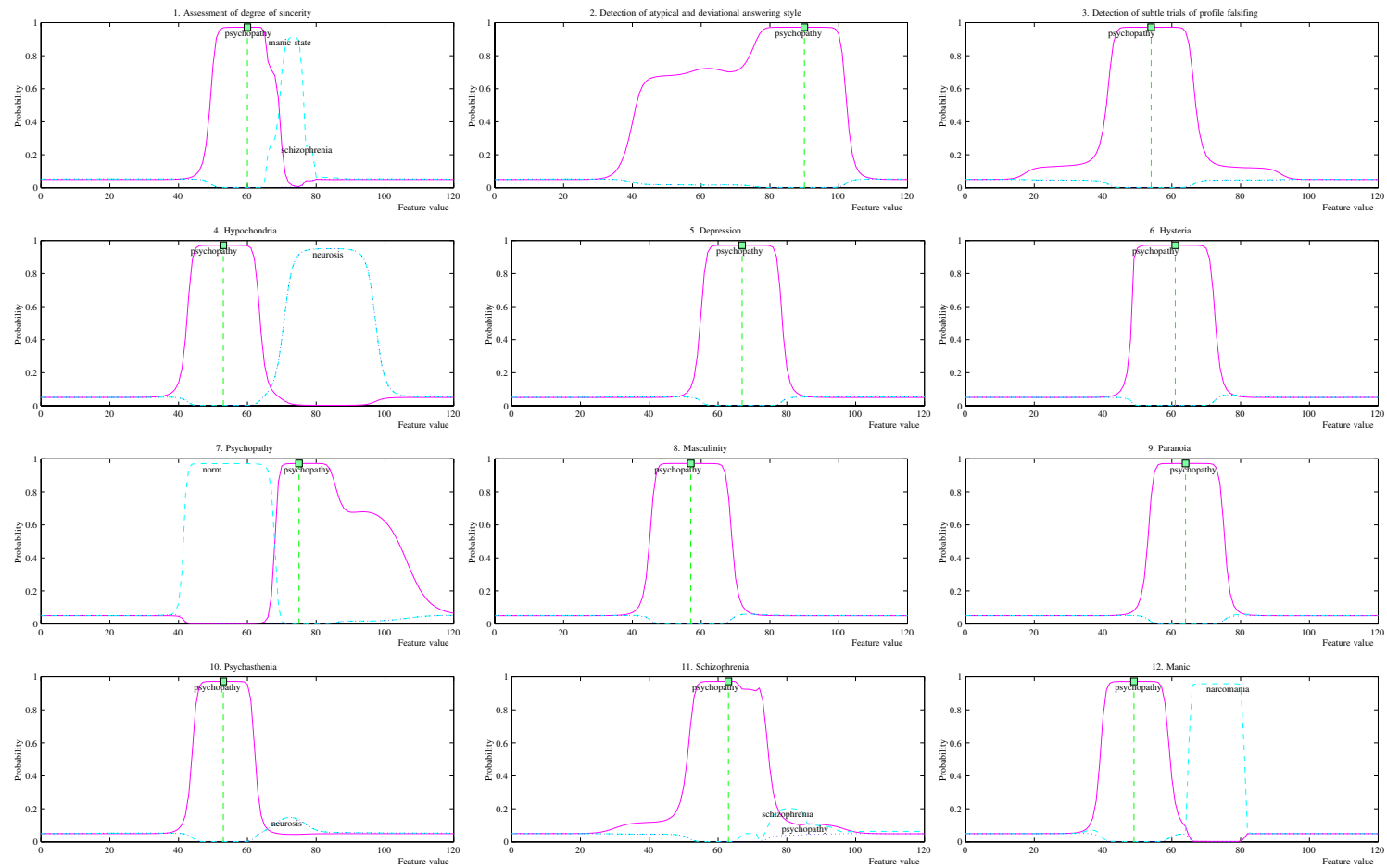


Rysunek 7: uczenie na 90%-owej części. Lewa dla zbioru testowego, prawa treningowego.

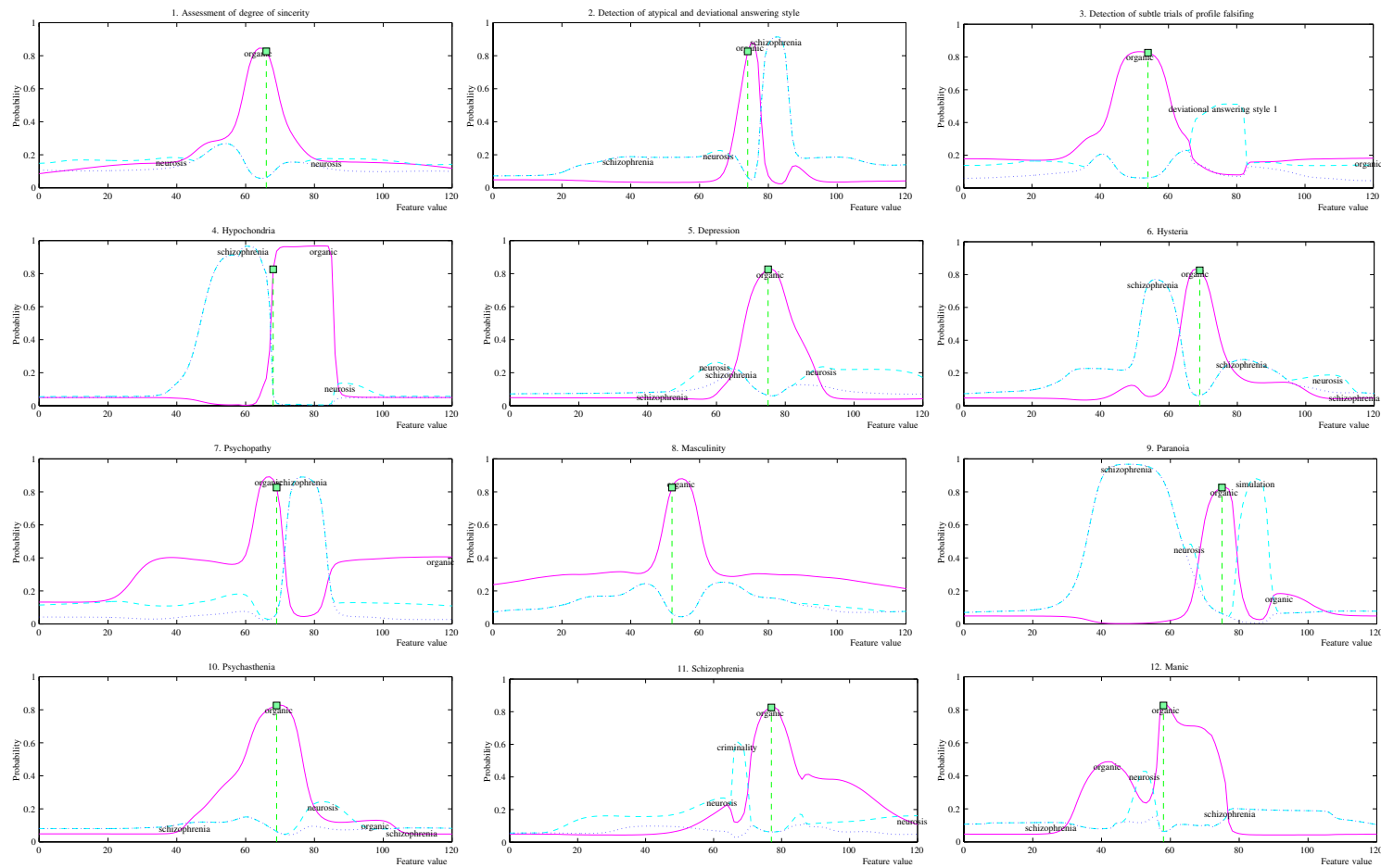
Wartości uzyskane i oczekiwane dla błędnie sklasyfikowanych wektorów



Rysunek 8: 95+5 dla 27-klasowej bazy.



Rysunek 9: Klasa: Psychopatia (0.97); alt. nerwica (0.002).



Rysunek 10: Organika (0.83), schizofrenia (0.062)

ANIM

Norma — alkoholizm

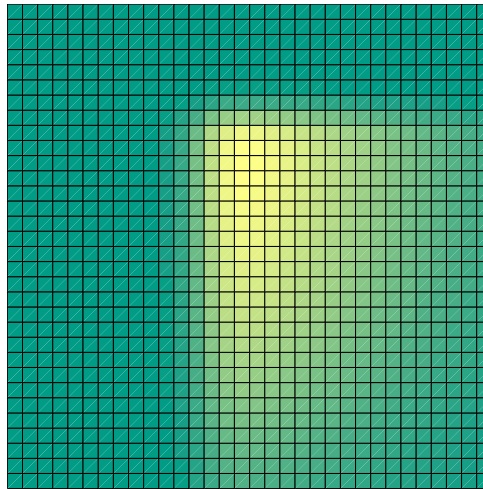
Norma — nerwica

Podsumowanie

- Taksonomia f. aktywacji i wyjścia
- Bicentralne funkcje transferu
- Szybka wersja filtra EKF
- statystyczne kryteria kontroli złożoności
- Łączenie neuronów
- Program: IncNet (C++)
- Interpretacja wyników — Prob. przedz. ufności
- Wyniki — zastosowania medyczne

N o r b e r t J a n k o w s k i

Ontogeniczne sieci neuronowe
w zastosowaniu do
klasyfikacji danych medycznych



Norbert.Jankowski@phys.uni.torun.pl

<http://www.phys.uni.torun.pl/~norbert>

XX