

Neuronové sítě

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky

Matematicko-fyzikální fakulta

Univerzity Karlovy v Praze

Neuronové sítě

– Interní reprezentace znalostí –

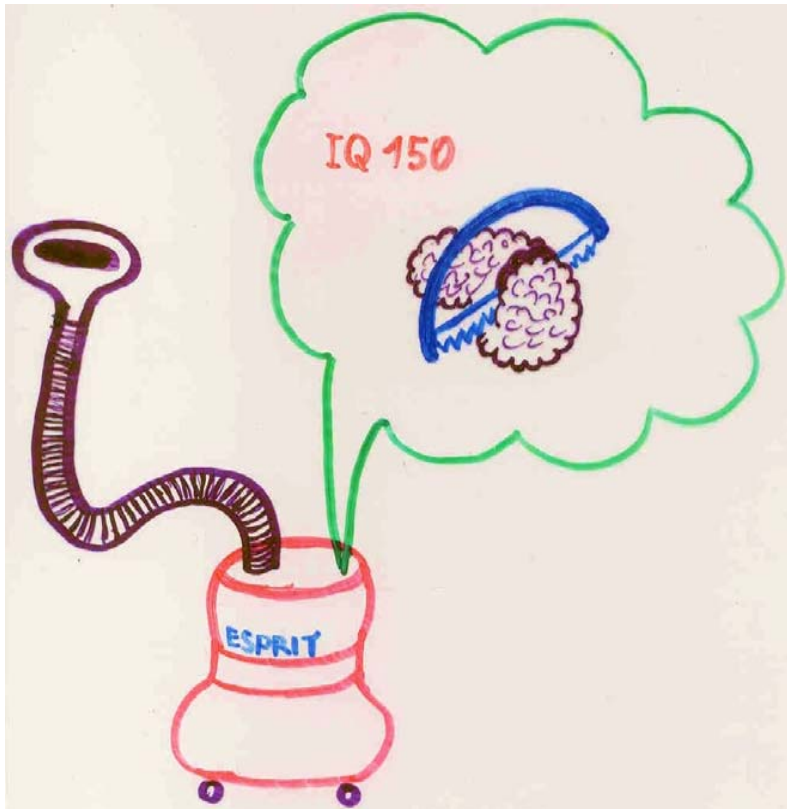
Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky

Matematicko-fyzikální fakulta

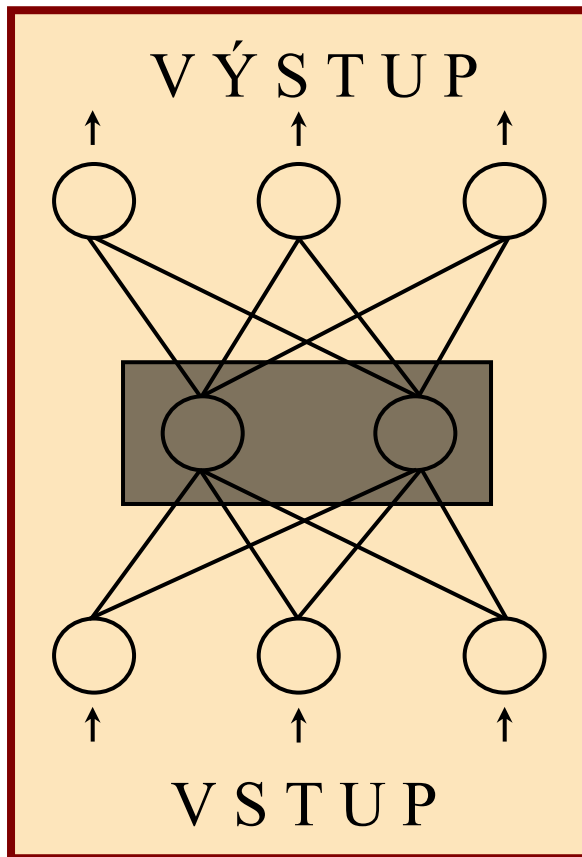
Univerzity Karlovy v Praze

Interní reprezentace znalostí

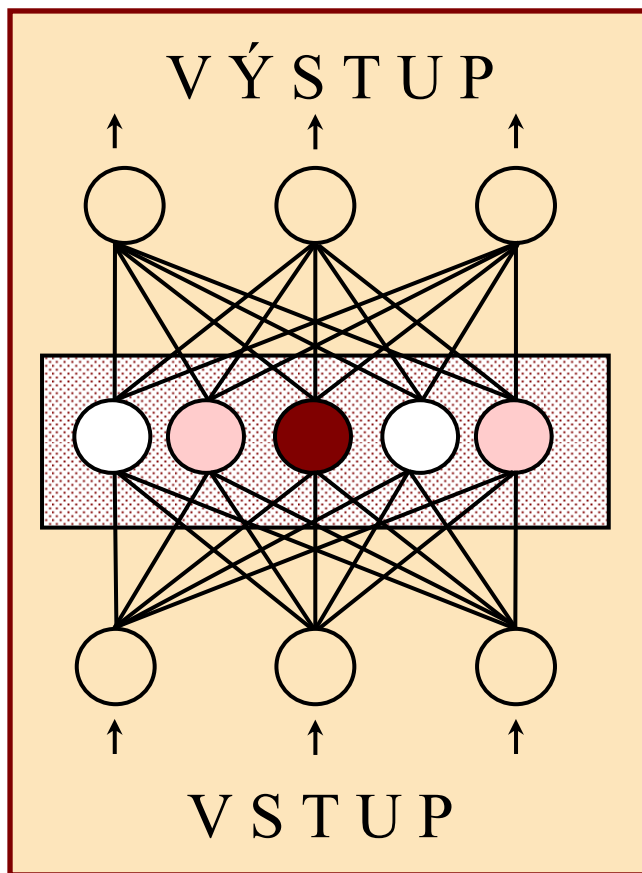


- počet neuronů a generalizační schopnosti sítě
→ **prořezávání a doučování**

Interní reprezentace znalostí



Kondenzovaná interní reprezentace



- ♦ interpretace aktivity skrytých neuronů:

●	1	↔	aktivní	↔	ANO
○	0	↔	pasivní	↔	NE
●	$\frac{1}{2}$	↔	tichý	↔	
		↔	„nelze rozhodnout“		

- ♦ průhledná struktura sítě
- ♦ detekce nadbytečných neuronů a prořezávání
- ♦ lepší generalizace

Kondenzovaná interní reprezentace

D: Pro vrstevnatou síť B zpracovávající vstupní vzor \vec{x} :

- Skrytý neuron s vahami (w_1, \dots, w_n) , prahem ϑ , vstupním vzorem \vec{z} a přenosovou funkcí $f[\vec{w}, \vartheta](\vec{z})$ vytváří **reprezentaci** $r : r = y = f[\vec{w}, \vartheta](\vec{z})$
- Vektor \vec{r} reprezentací vytvořených vrstvou skrytých neuronů se nazývá **interní reprezentace** \vec{x}

Kondenzovaná interní reprezentace

D: Pro vrstevnatou síť B :

- Interní reprezentace $\vec{r} = (r_1, \dots, r_m)$ je **binární**, jestliže $r_i \in \{0, 1\}$; $1 \leq i \leq m$
- Interní reprezentace $\vec{r} = (r_1, \dots, r_m)$ je **kondenzovaná**, jestliže $r_i \in \{0, 0.5, 1\}$; $1 \leq i \leq m$

Požadavky na vynucování kondenzované interní reprezentace

- ♦ formulace „požadovaných vlastností“ ve formě cílové funkce:

$$\mathbf{G} = \mathbf{E} + c_s \mathbf{F}$$

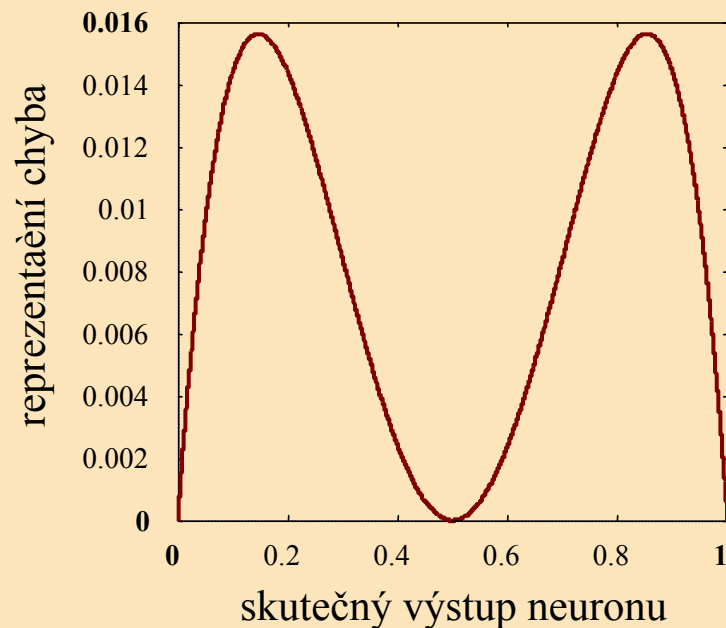
standardní chybová funkce
reprezentační chybová funkce
velikost vlivu F na G

- ♦ lokální minima reprezentační chybové funkce odpovídají aktivním, pasivním a tichým stavům:

$$F = \sum_p \sum_h y_{h,p}^s (1 - y_{h,p})^s (y_{h,p} - 0.5)^2$$

vzory p skryté neurony h pasivní stav $y_{h,p}^s$ aktivní stav $(1 - y_{h,p})^s$ tvar F $(y_{h,p} - 0.5)^2$ tichý stav

Vliv parametrů na vytváření kondenzované interní reprezentace



$$w_{ij}(t+1) = w_{ij}(t) + \alpha \delta_j y_i + \alpha_r \rho_j y_i + \alpha_m (w_{ij}(t) - w_{ij}(t-1))$$

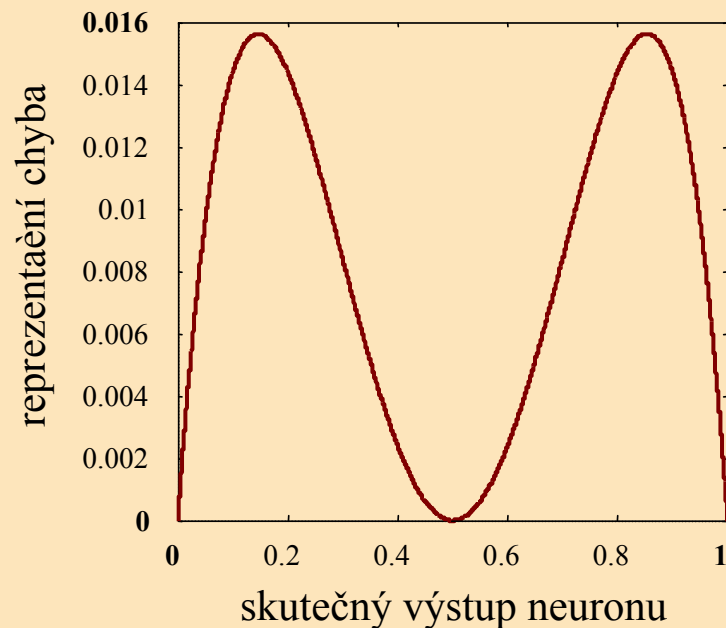
- ◆ pomalejší vytváření interní reprezentace a požadovaná funkce sítě
- ◆ stabilita vytvářené interní reprezentace a optimální architektura sítě
- ◆ tvar reprezentační chybové funkce, rychlost vytváření interní reprezentace a její forma
- ◆ časová náročnost při adaptaci vah

Chybový člen pro posilování kondenzované interní reprezentace

Kondenzovaná interní reprezentace ($y_j^s (1 - y_j)^s (y_j - 0.5)^2$):

$$\rho_j = \begin{cases} 0 & \text{pro výstupní neurony} \\ - \left[2(s+1)y_j(1-y_j) - \frac{s}{2} \right] \cdot y_j^s (1-y_j)^s (y_j - 0.5) & \text{pro neurony z nejvyšší skryté vrstvy} \\ \left(\sum_k \rho_k w_{jk} \right) y_j(1-y_j) & \text{pro ostatní skryté neurony} \end{cases}$$

Vliv parametrů na vytváření kondenzované interní reprezentace

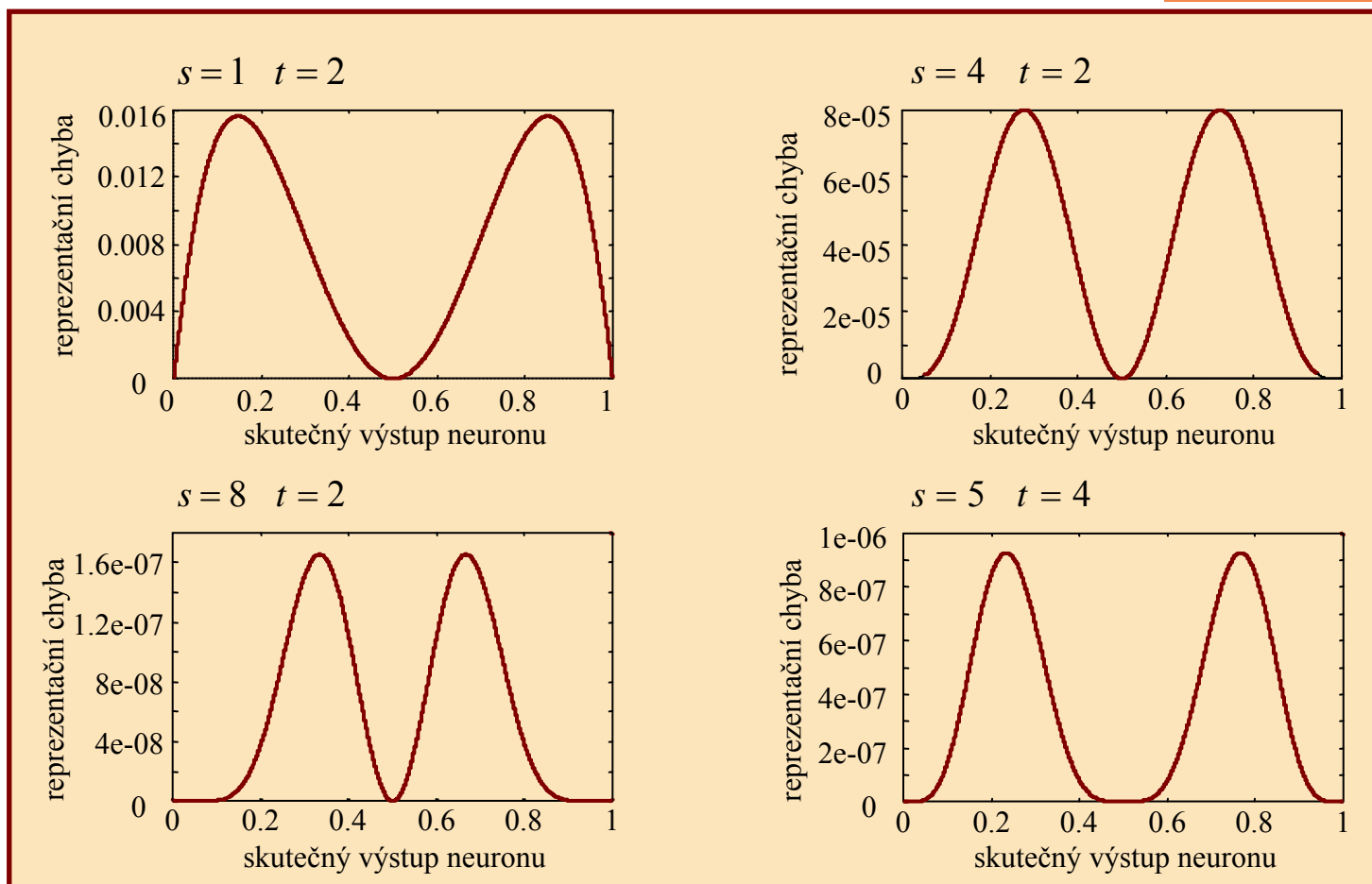


$$w_{ij}(t+1) = w_{ij}(t) + \alpha \delta_j y_i + \alpha_r \rho_j y_i + \alpha_m (w_{ij}(t) - w_{ij}(t-1))$$

- ◆ pomalejší vytváření interní reprezentace a požadovaná funkce sítě
- ◆ stabilita vytvářené interní reprezentace a optimální architektura sítě
- ◆ tvar reprezentační chybové funkce, rychlost vytváření interní reprezentace a její forma
- ◆ časová náročnost při adaptaci vah

Tvar reprezentační chybové funkce

$$F = y^s (1 - y)^s (y - 0.5)^t$$



Chybový člen pro posilování binární interní reprezentace

Binární interní reprezentace ($y_j (1 - y_j)$):

$$\rho_j = \begin{cases} 0 & \text{pro výstupní neurony} \\ - (1 - 2 y_j) y_j (1 - y_j) & \text{pro neurony z nejvyšší skryté vrstvy} \\ \left(\sum_k \rho_k w_{jk} \right) y_j (1 - y_j) & \text{pro ostatní skryté neurony} \end{cases}$$

Jednoznačná interní prezentace

- ◆ Hodně odlišným výstupům by měly odpovídat hodně odlišné interní reprezentace
- ◆ Formulace požadavků ve formě modifikované cílové funkce: $G = E + F + H$
- ◆ Kritérium pro jednoznačnost IR:

$$H = -\frac{1}{2} \sum_p \sum_{q \neq p} \sum_j \sum_o (d_{o,p} - d_{o,q})^2 (y_{j,p} - y_{j,q})^2$$

vzory skryté neurony p $q \neq p$ j o výst. neurony $d_{o,p}$ $d_{o,q}$ $y_{j,p}$ $y_{j,q}$

= konst. pro dané p = konst. pro dané p = konst. pro dané p

Prořezávání podle interní reprezentace (1)

D: Pro danou vrstevnatou síť B a množinu S vstupních vzorů určujících vstupní vektory \vec{z} :

- Skrytý neuron s vahami (w_1, \dots, w_n) , prahem ϑ a přenosovou funkcí $f[\vec{w}, \vartheta](\vec{z})$ vytváří **uniformní reprezentaci** r , jestliže:

$$r = f[\vec{w}, \vartheta](\vec{z}) = \text{const} \quad \text{pro všechny vstupní vzory } \vec{x} \in S$$

Prořezávání podle interní reprezentace (2)

D: Pro danou vrstevnatou síť B a množinu S vstupních vzorů určujících vstupní vektory \vec{z} :

- Skrytý neuron $i \in N$ s vahami (w_{i1}, \dots, w_{in}) , prahem ϑ_i a přenosovou funkcí $f_i[\vec{w}_i, \vartheta_i](\vec{z})$ vytváří reprezentaci r_i **identickou** k reprezentaci r_j vytvářené skrytým neuronem $j \in N$ s vahami (w_{j1}, \dots, w_{jn}) , prahem ϑ_j a přenosovou funkcí $f_j[\vec{w}_j, \vartheta_j](\vec{z})$, jestliže:

$$f_i[\vec{w}_i, \vartheta_i](\vec{z}) = f_j[\vec{w}_j, \vartheta_j](\vec{z}) \quad \text{pro všechny vstupní vzory } \vec{x} \in S$$

Prořezávání podle interní reprezentace (3)

D: Pro danou vrstevnatou síť B a množinu S vstupních vzorů určujících vstupní vektory \vec{z} :

- Skrytý neuron $i \in N$ s vahami (w_{i1}, \dots, w_{in}) , prahem ϑ_i a přenosovou funkcí $f_i[\vec{w}_i, \vartheta_i](\vec{z})$ vytváří reprezentaci r_i **inverzní** k reprezentaci r_j vytvářené skrytým neuronem $j \in N$ s vahami (w_{j1}, \dots, w_{jn}) , prahem ϑ_j a přenosovou funkcí $f_j[\vec{w}_j, \vartheta_j](\vec{z})$, jestliže:

$$f_i[\vec{w}_i, \vartheta_i](\vec{z}) = 1 - f_j[\vec{w}_j, \vartheta_j](\vec{z}) \quad \text{pro všechny vstupní vzory } \vec{x} \in S$$

Prořezávání podle interní reprezentace (4)

D: Pro danou vrstevnatou síť B a množinu vstupních vzorů S :

redukovaná vrstva je vrstva, pro kterou platí, že:

- žádný neuron nevytváří uniformní reprezentaci,
- žádný neuron i nevytváří reprezentaci identickou k reprezentaci vytvářené jiným neuronem j a
- žádný neuron i nevytváří reprezentaci inverzní k reprezentaci vytvářené jiným neuronem j .

Interní reprezentace vytvářené redukovanou vrstvou se nazývá **redukovaná**.

Prořezávání podle interní reprezentace (5)

D: Pro danou množinu vstupních vzorů S :

- vrstevnatá síť B je **redukovaná**, jestliže jsou všechny její skryté vrstvy redukovány.
- vrstevnatá síť B je **ekvivalentní** k vrstevnaté síti B' , jestliže je pro libovolný vstupní vzor $\vec{x} \in S$ skutečný výstup \vec{y}_B sítě B roven skutečnému výstupu $\vec{y}_{B'}$ sítě B' : $\vec{y}_B = \vec{y}_{B'}$

Prořezávání podle interní reprezentace (6)

V: Ke každé vrstevnaté síti B a množině vstupních vzorů S existuje ekvivalentní redukovaná vrstevnatá síť B' .

Důkaz (idea):

Popis konstrukce redukované vrstevnaté sítě B' :

Necht' $B = (N, C, I, O, w, t)$ je původní vrstevnatá síť.

1. Postupná eliminace všech takových neuronů j , které vytvářejí uniformní reprezentaci r_j^k a přičtení součinu $w_{ij} r_j^k$ ke všem prahům ϑ_j v následující vrstvě.

Prořezávání podle interní reprezentace (7)

Důkaz (pokračování):

2. Postupná eliminace všech takových neuronů j , které vytvářejí reprezentaci r_j^{id} identickou k reprezentaci r_k vytvářené jiným neuronem k a přičtení vah w_{ij} ke každé váze w_{ik} , kde i je neuron v následující vrstvě.
3. Postupná eliminace všech takových neuronů j , které vytvářejí reprezentaci r_j^{in} inverzní k reprezentaci r_k vytvářené jiným neuronem k a nahrazení všech vah w_{ik} , kde i označuje neuron z následující vrstvy, rozdílem $w_{ik} - w_{ij}$ a přičtení váhy w_{ij} k prahu ϑ_j každého neuronu i .

Prořezávání podle interní reprezentace (8)

Důkaz (pokračování):

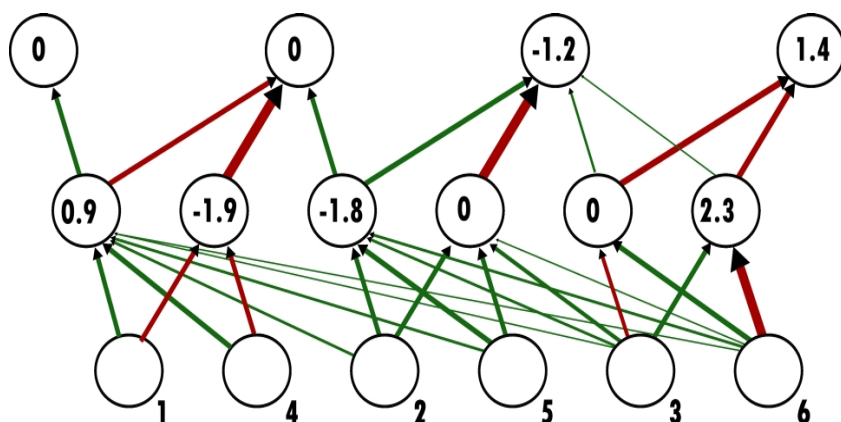
Potom bude pro libovolný vstupní vzor \vec{x} skutečný výstup $\vec{y}_{B'}$ vrstevnaté sítě B' roven skutečnému výstupu \vec{y}_B vrstevnaté sítě B .

Vrstevnatá síť B' konstruovaná ze sítě B popsáním způsobem je redukováná a ekvivalentní k B .

QED

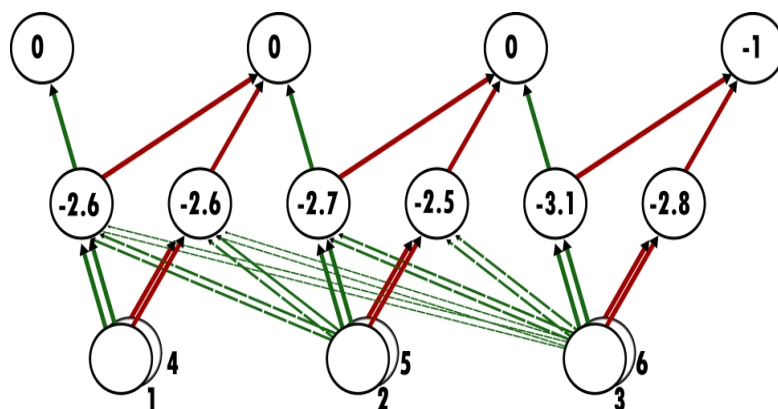
Výsledky experimentů: binární sčítání

[$5(\approx(1,-1,1)) + 3(\approx(-1,1,1)) = 8(\approx(1,-1,-1,-1))$]



◆ SCG-s náповědou (přenos na 2. výstupní neuron)

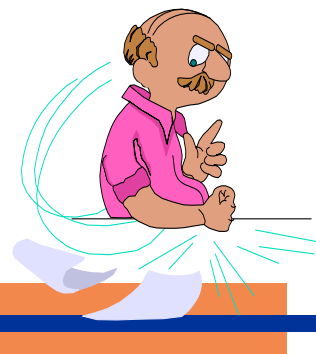
- ‘přenos’ první a druhý výstupní bit – skryté neurony 1 a 3
- funkce ostatních skrytých neuronů není tak zřejmá



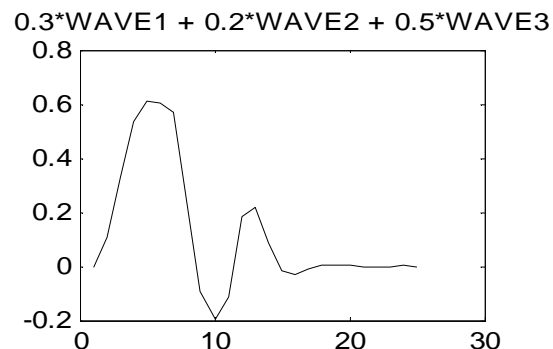
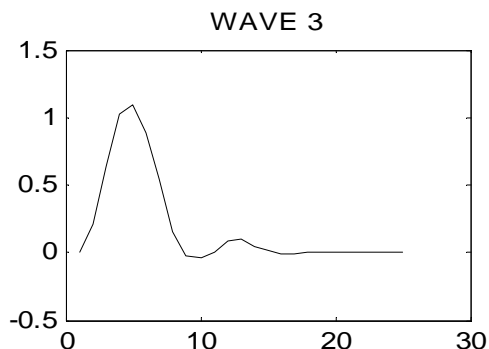
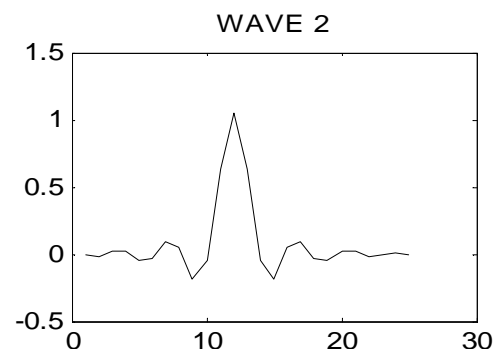
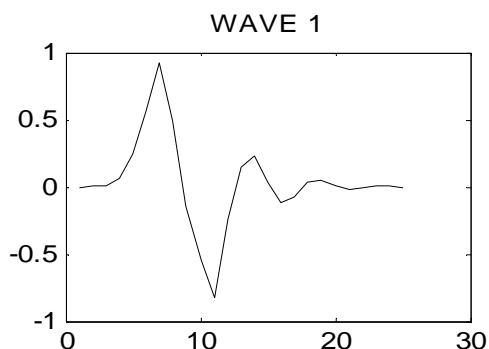
◆ SCGIR-s náповědou (přenos na 2. výstupní neuron)

- ‘přenos’ pro vyšší výstupní bity – skryté neurony 1, 3, 5
- podobná funkce je zřejmá pro jednotlivé výstupní neurony

Akustická emise: simulace (s M. Chladou a Z. Převorovským)

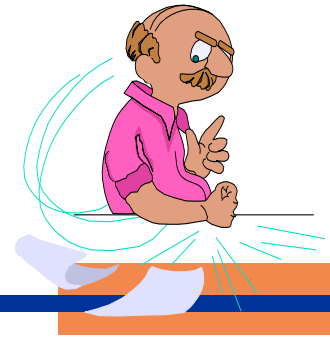


MODELOVANÝ SIGNÁL

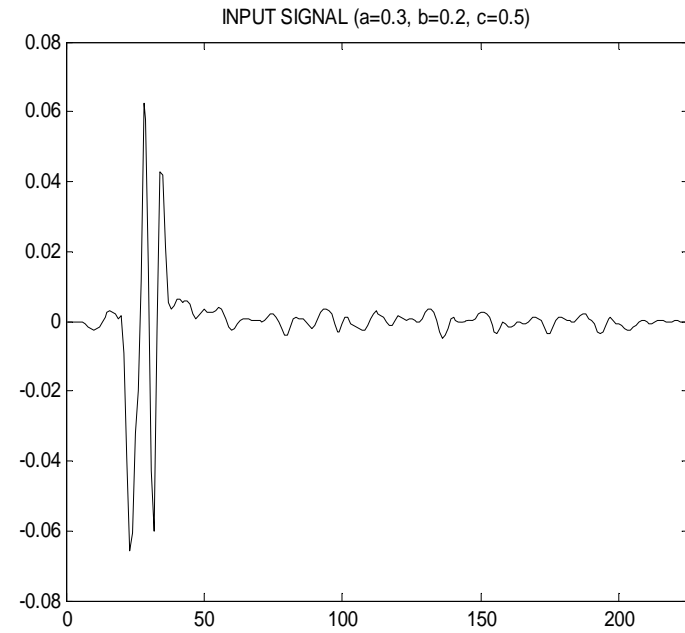
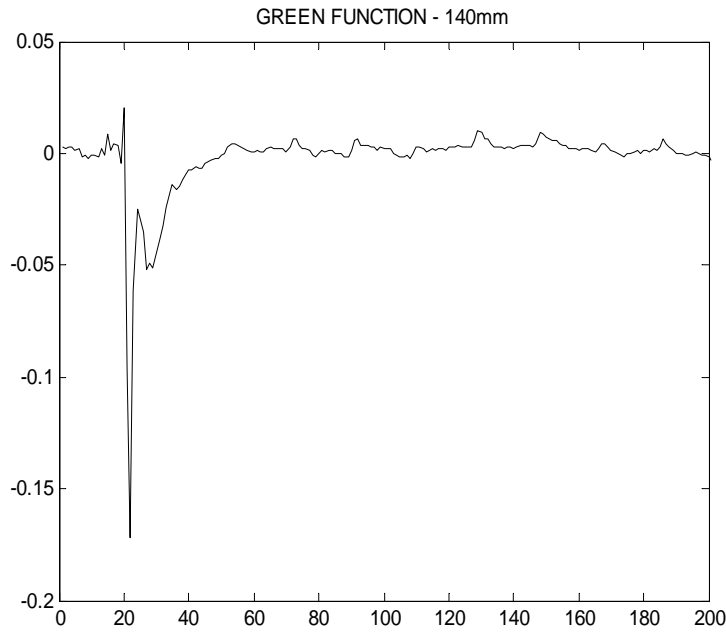


Simulovaná AE-data

(s M. Chladou a Z. Převorovským)

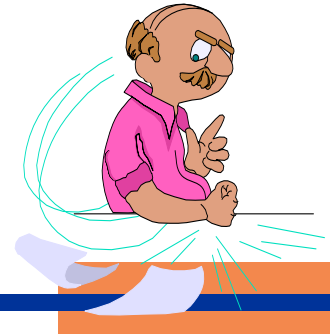


KONVOLUCE S GREENOVOU FUNKCÍ



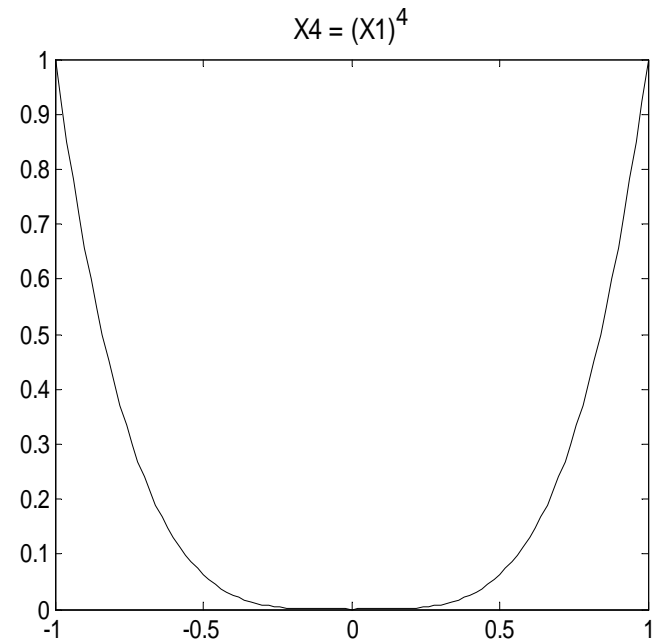
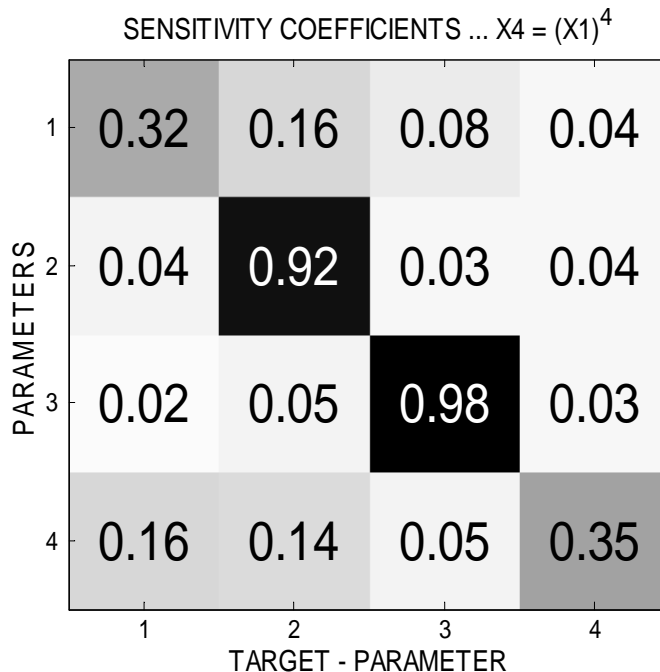
Model závislosti

(s M. Chladou a Z. Převorovským)



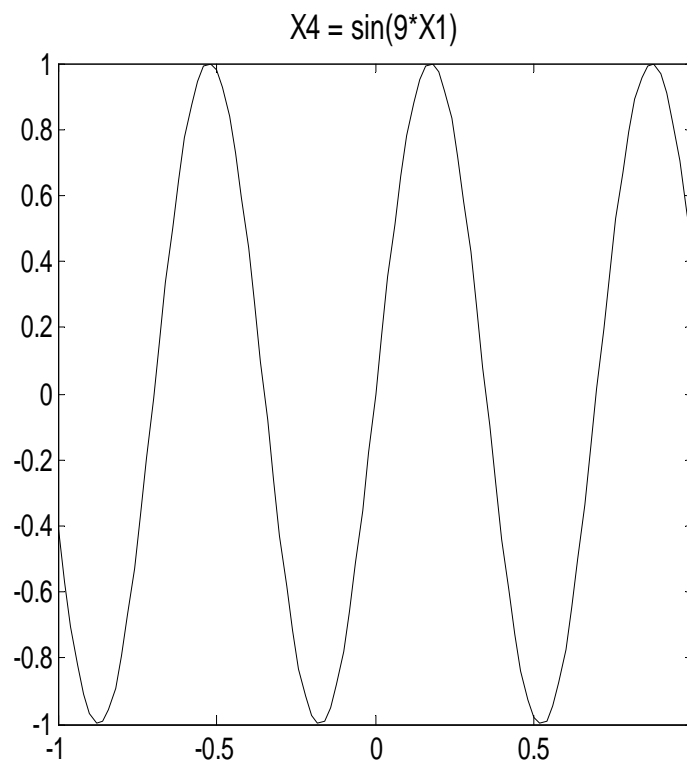
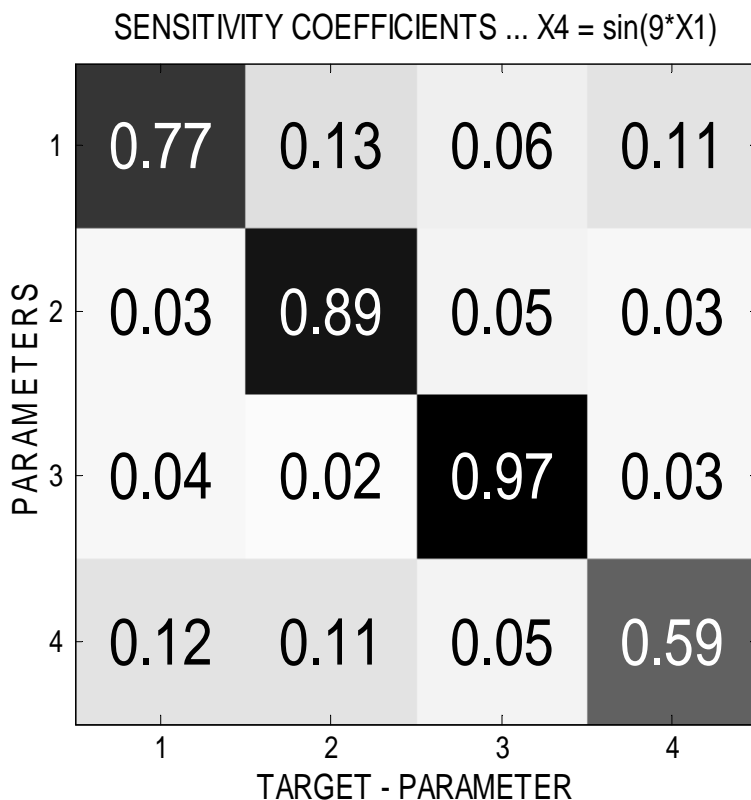
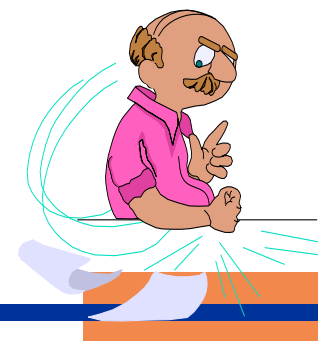
Celková **citlivost sítě** (přes Q vzorů) s -tého výstupu na r -tý vstup:

$$sens_r = 1/Q \sum_q \sum_s \left| \partial y_{q,s} / \partial y_{q,r} \right|$$



Model závislosti

(s M. Chladou a Z. Převorovským)



Faktorová vs. citlivostní analýza vstupních parametrů (s M. Chladou a Z. Převorovským)



vstupní parametry

1	0.04	0.91	0.16	0.10	0.01	0.07	0.06	0.01	0.23
2	0.09	0.02	0.01	0.19	0.03	0.03	0.95	0.03	0.16
3	0.10	0.96	0.15	0.00	0.03	0.00	0.02	0.05	0.07
4	0.13	0.91	0.02	0.03	0.05	0.06	0.04	0.06	0.20
5	0.30	0.05	0.04	0.41	0.06	0.49	0.17	0.07	0.66
6	0.26	0.00	0.04	0.03	0.08	0.93	0.02	0.10	0.20
7	0.29	0.06	0.03	0.27	0.03	0.17	0.16	0.04	0.86
8	0.12	0.06	0.01	0.88	0.02	0.02	0.24	0.03	0.36
9	0.90	0.15	0.09	0.08	0.15	0.17	0.06	0.21	0.18
10	0.93	0.14	0.06	0.08	0.10	0.17	0.06	0.09	0.19
11	0.25	0.10	0.12	0.03	0.25	0.11	0.03	0.90	0.05
12	0.20	0.07	0.14	0.02	0.93	0.09	0.03	0.23	0.04
13	0.04	0.09	0.97	0.00	0.05	0.00	0.00	0.06	0.02
14	0.08	0.18	0.95	0.02	0.09	0.04	0.01	0.06	0.02
	1	2	3	4	5	6	7	8	9

vybrané faktory

- ♦ Vybráno 9 faktorů
("vysvětlují" 98.4% proměnných)
- ♦ redukce lineárně závislých vstupních parametrů

SENSITIVITY COEFFICIENTS

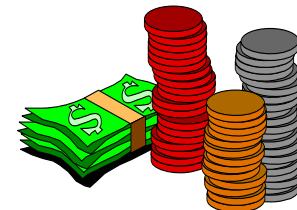
1	0.173	0.266	0.149
2	0.093	0.068	0.047
3	0.320	0.193	0.184
4	0.301	0.178	0.196
5	0.564	0.250	0.206
6	0.196	0.322	0.158
7	0.099	0.063	0.043
8	0.065	0.015	0.030
9	0.022	0.014	0.016
10	0.053	0.020	0.012
11	0.035	0.012	0.032
12	0.039	0.050	0.022
13	0.081	0.134	0.082
14	0.260	0.172	0.109
	1	2	3

INPUTS

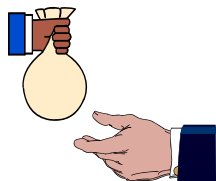
OUTPUTS

- ♦ Vybráno 7 příznaků
- ♦ detekce nelineární závislosti vstupních parametrů (1, 3, 4, 5, 6, 13, 14)

Analýza dat ze Světové banky



WDI-indikátory (indikátory vývoje ve světě)



- každoročně zveřejňovány Světovou bankou
 - pomoc rozvojovým zemím při půjčkách / investicích
 - odhad stavu ekonomik a jejich vývoje v jednotlivých zemích
- původ údajů - neúplné a nepřesné údaje

♦ používané techniky

- regresní analýza - lineární závislosti
- kategorizace států používaná v rozvinutých zemích (G. Ip, Wall Street Journal)
- kategorizace zemí podle HDP (Světová banka)
- Kohonenovy mapy (T. Kohonen, S. Kaski, G. Deboeck)

Analýza dat ze Světové banky: použité WDI-indikátory



- ◆ Implicitní deflace HDP
 - ◆ Vnější zadluženost (% HNP)
 - ◆ Celkové náklady na zadlužení (% z exportu zboží a služeb)
 - ◆ Export high-tech technologií (% z vyvážených výrobků)
 - ◆ Výdaje na armádu a zbrojení (% HNP)
 - ◆ Výdaje na výzk. a výv. (% HNP)
 - ◆ Celk. výd. na zdrav. (% HDP)
 - ◆ Veř. výd. na školst. (% HNP)
 - ◆ Očekávaná délka života u mužů
 - ◆ Plodnost
 - ◆ GINI-index (rozdělení příjmů a spotřeby)
 - ◆ Uživ. internetu na 10000 obyvatel
 - ◆ Počet mobilních telefonů na 1000 obyvatel
-
- ◆ HNP na obyvatele podle parity kupní síly (PPP)
 - ◆ HNP na obyvatele (v USD)
 - ◆ Růst HDP (% na obyvatele)

Analýza dat ze Světové banky: předzpracování



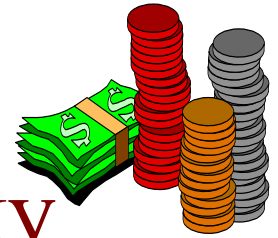
- ◆ 99 států se 16 WDI-indikátory
- ◆ po složkách transformace vzorů do intervalu (0,1) pomocí:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{a} \quad x'' = \frac{1}{1 + e^{-4(x' - 1/2)}}$$

↑
maximum přes všechny vzory minimum přes všechny vzory

- ◆ FCM-klastrování: 7 shluků, $s = 1.4$
- ◆ řízené učení a iterativní rozpoznávání:
 - 99 (90+9) států s 14 (13+1) WDI-indikátory
 - GREN-sít' 14-12-1, BP-sít' 13-10-1; 500-600 cyklů učení

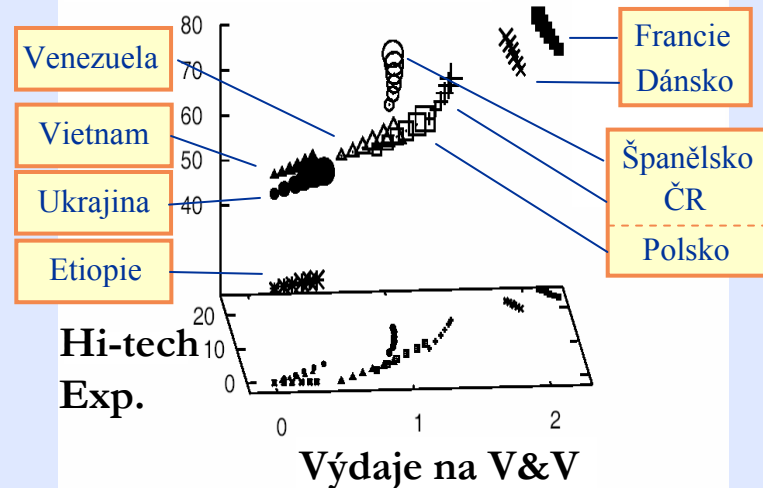
Analýza dat ze Světové banky: vliv indikátorů na stav ekonomiky



Indikátor	Síť 1	Síť 2
GDP defl.	0.0	0.0
Vněj. dluh	5.6	10.9
Celk. nákl. na dluh	5.5	8.1
Export high-tech	12.2	6.6
Vojenské výdaje	5.4	6.1
Výdaje na výzk. a výv.	16.0	12.0
Uživ. internetu	11.1	12.4
Mobily	8.3	10.0
GINI-index	7.1	3.9
Oček. délka života	12.3	7.6
Plodnost	4.4	5.0
Výdaje na zdrav.	6.1	10.9
Veř. výd. na školstv.	6.1	6.1

Relativní citlivost GREN-sítí

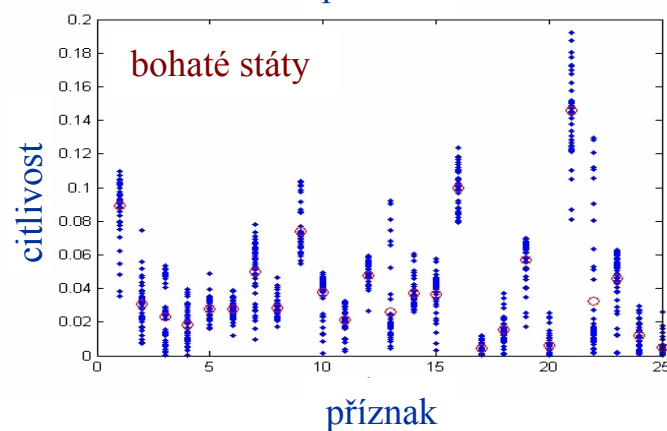
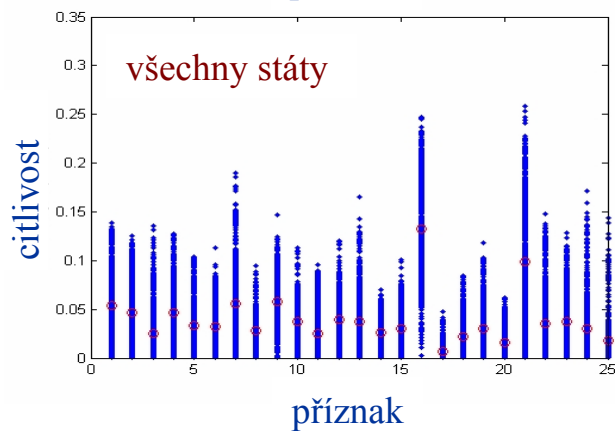
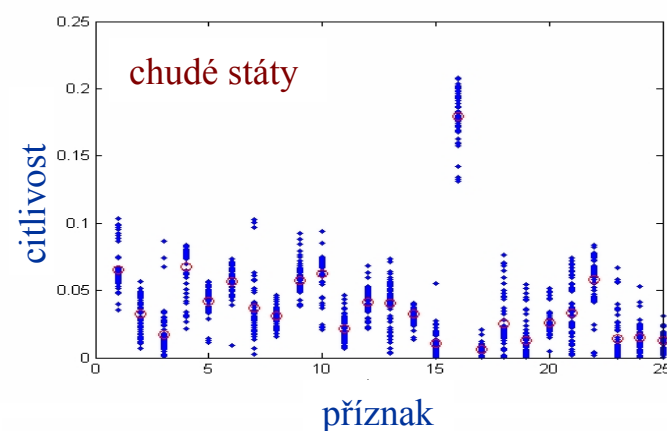
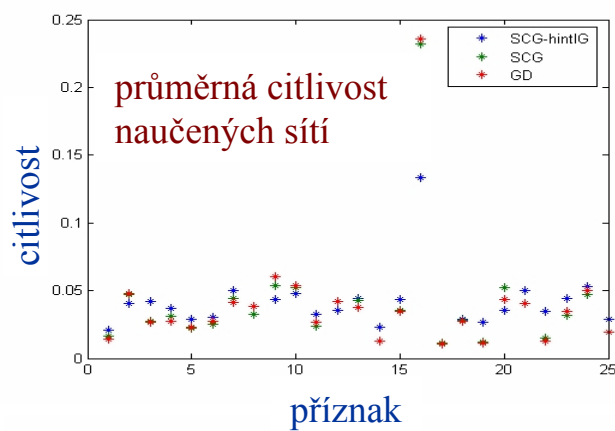
Očekávaná délka života



Iterativní rozpoznávání – vyšší
HNP podle PPP (Síť 1)

Citlivost na vstupní příznaky

(se Z. Reitermanovou)



Vzájemná závislost parametrů

(se Z. Reitermanovou)

