

# ROBUSTNÍ ARCHITEKTURA VÍCEVRSTVÝCH NEURONOVÝCH SÍTÍ

ZUZANA PETŘÍČKOVÁ

reitezuz@fjfi.cvut.cz

Katedra softwarového inženýrství,  
Fakulta jaderná a fyzikálně inženýrská,  
ČVUT v Praze

- Na základě dizertační práce:** Umělé neuronové sítě a jejich využití při extrakci znalostí  
**Školitel:** doc. RNDr. Iveta Mrázová, CSc.  
**Školící pracoviště:** Katedra teoretické informatiky a matematické logiky  
MFF UK.  
**Obhájeno:** 22. září 2015.

# Obsah prezentace

## 1 Úvod

- Motivace
- Vrstevnaté neuronové sítě
- Optimalizace struktury

## 2 Navržený framework

- Rychlá extrakce znalostí (SCGIR)
- Zjednodušení vnitřní struktury BP-sítě (SCGS)
- Rychlé učení s vytvořením jednodušší a jasnější vnitřní struktury sítě (SCGSA)

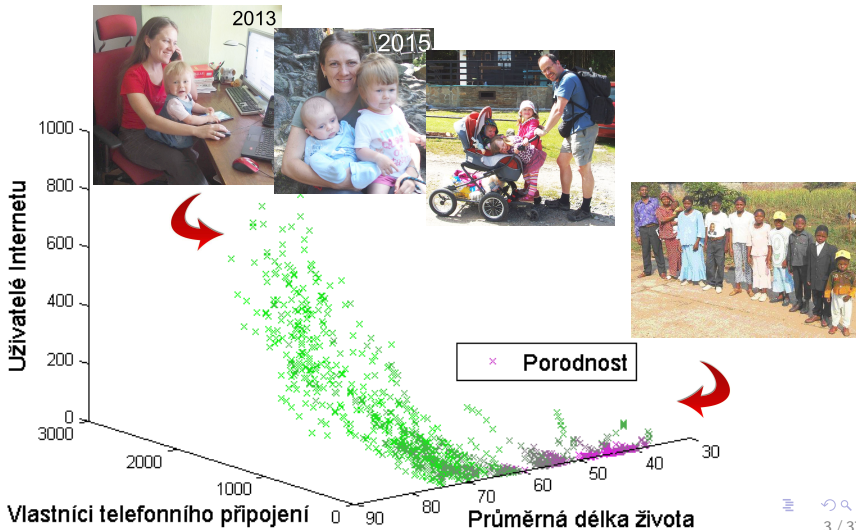
## 3 Experimenty

- Rychlost učení a schopnost zobecňovat
- Stabilita metod
- Vytvoření jednoduché a transparentní struktury během učení
- Citlivostní analýza

## 4 Závěr

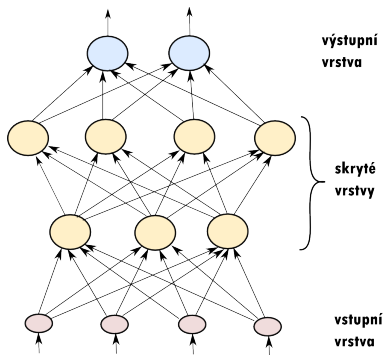
# Motivace: Data ze Světové Banky

- Indikátory světového vývoje, 162 zemí, 2001-2006, 956 příkladů



# Vrstevnaté neuronové sítě typu zpětného šíření (BP-sítě)

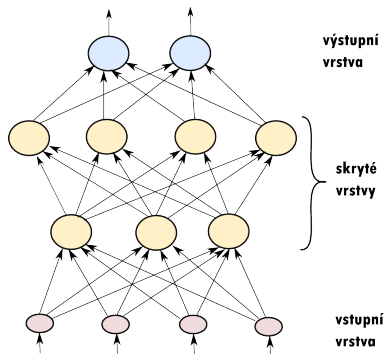
- Klasický výpočetní model  
(*Werbos, 1975*), (*Parker, 1985*),  
(*LeCun, 1985*), (*Rumelhart, 1986*).



- Gradientní metody učení (např. back-propagation):
  - Učení na základě množiny trénovacích příkladů.
  - Adaptace vah a prahů
    - proti směru gradientu chybové funkce.
    - zpětně: směrem od výstupní vrstvy ke vstupní.

# Vrstevnaté neuronové sítě typu zpětného šíření (BP-sítě)

- Klasický výpočetní model  
(*Werbos, 1975*), (*Parker, 1985*),  
(*LeCun, 1985*), (*Rumelhart, 1986*).

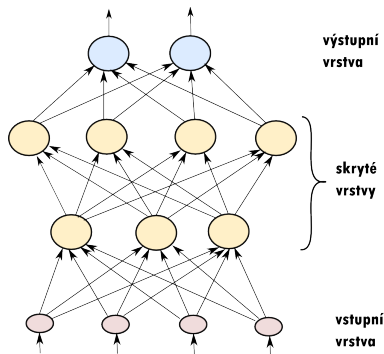


## Hlavní klady

- Jednoduchý, výpočetně poměrně nenáročný model.
- Jsou schopné dobře aproximovat neznámou funkci.
- Dobře zobecňují.
- Zdroj poznatků pro složitější modely neuronových sítí - např. hluboké neuronové sítě (*LeCun, 1998*), (*Hinton, 2009*), rekurentní modely.

# Vrstevnaté neuronové sítě typu zpětného šíření (BP-sítě)

- Klasický výpočetní model  
(*Werbos, 1975*), (*Parker, 1985*),  
(*LeCun, 1985*), (*Rumelhart, 1986*).



## Hlavní nevýhody:

- Učení je poměrně pomalé.
- Nebezpečí přeučení.
- Mají tendenci vytvářet složitou a netransparentní vnitřní strukturu.
- Vysoká citlivost k volbě parametrů (architektura, algoritmus učení, trénovací data,...)

# Optimalizace architektury a detekce důležitých vstupních příznaků

- Často řešeno odděleně – odlišná motivace.

## Proč hledat důležité vstupní příznaky?

- Součást předzpracování dat.
- Efektivnější výpočet modelu.
- Zvýšení přesnosti modelu např. v případě nedostatku trénovacích dat.
- Porozumění datům.

→ **BP-sítě umožňují řešit obě úlohy zároveň.**

## Proč optimalizovat architekturu?

- Součást algoritmu učení.
- Lepší predikce a zobecňování.
- Jednodušší a transparentnější struktura modelu.
- Porozumění tomu, jak model počítá.

# Optimalizace architektury a detekce důležitých vstupních příznaků

## Jak optimalizovat strukturu BP-sítě?

- Metody hrubé síly
  - Vyzkoušet různé architektury a vybrat tu s nejmenší chybou po naučení.
  - **Nevýhody:** Časová náročnost, obtížné učení a stabilita minimálního modelu, citlivost k lokálním minimům chybové funkce.
- **Prořezávání**
- **Regularizace**
- Konstrukční metody
- Pravděpodobnostně-optimalizační metody (genetické algoritmy, simulované žíhání,...)
- ...



# Optimalizace architektury a detekce důležitých vstupních příznaků

## Prořezávání - základní algoritmus:

- 1 Nauč BP-síť s dostatečně velkou počáteční topologií.
- 2 Dokud je chyba sítě větší určená mez nebo dokud klesá:
  - 1 Spočítej relevanci skrytých neuronů nebo hran.
  - 2 Odstraň z BP-sítě nejméně relevantní část(i).
  - 3 Přeuč model.

## Otázky:

- Jak vyhodnotit důležitost jednotlivých neuronů/hran → **míra relevance**.
- Jak rozpoznat, které části sítě prořezat → **heuristiky**.

# Optimalizace architektury a detekce důležitých vstupních příznaků

## Regularizační techniky

- Přidávají k chybové funkci další členy:

$$E = c_{mse} E_{mse} + c_F F + c_G G + \dots$$

## Weight decay (*Werbos, 1988*)

- Jednoduchý penalizační člen:  $E = E_{mse} + \beta \sum_i w_i^2$
- Vynucuje snížení absolutní hodnoty všech vah.
- **Cíle:**
  - Usnadnit odstranění hran s "malými" vahami.
  - Zlepšit schopnost modelu zobecňovat.

- 
- $E$  ... chybová funkce,
  - $E_{mse}$  ... střední kvadratická odchylka mezi skutečným a požadovaným výstupem,
  - $i$  ... index přes všechny váhy  $w$  v BP-síti,
  - $0 < \beta \ll 1$  ... konstantní parametr.

# Cíle mé dizertační práce

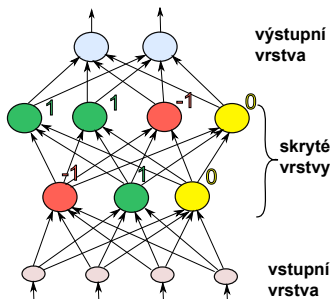
## Vytvořit obecný framework pro učení BP-sítí s důrazem na:

- 1 Rychlou extrakci znalostí (**SCGIR**).
- 2 Zjednodušení vnitřní struktury BP-sítě (**SCGS**).
- 3 Oba předchozí cíle zároveň (**SCGSA**).  
(Rychlé učení s vytvořením jednodušší a jasnější vnitřní struktury sítě)

# I. Rychlá extrakce znalostí (SCGIR)

## Učící algoritmus by měl:

- Být **rychlý** a **robustní** k volbě parametrů  
→ **Metoda škálovaných konjugovaných gradientů (SCG)**

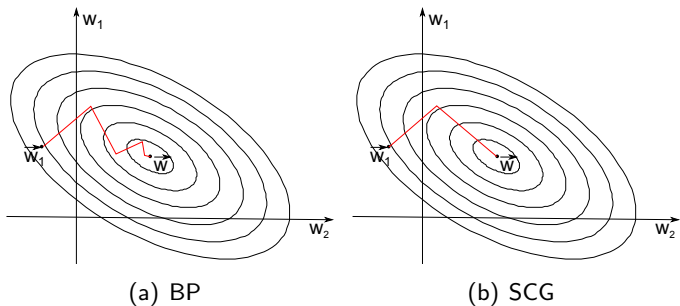


- Vytvořit **transparentní strukturu sítě** a umožnit tak **jednoduchou interpretaci** získaných znalostí  
→ **Metoda vynucované kondenzované interní reprezentace (IR)**
- Vytvořit jednoduchý model, který dobře **zobecňuje**  
→ **Učení s nápovědou**  
→ **Prořezávání**

# I. Rychlá extrakce znalostí (SCGIR)

## Metoda škálovaných konjugovaných gradientů

(Moller, 1993)



Velmi rychlý a robustní algoritmus, s malým počtem volitelných parametrů.

# I. Rychlá extrakce znalostí (SCGIR)

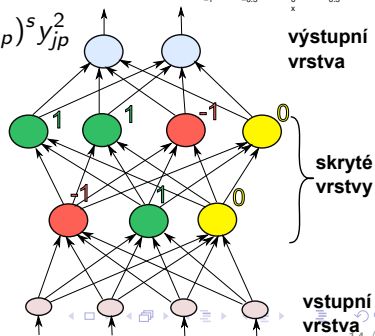
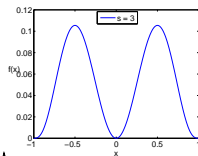
## Metoda vynucované kondenzované interní reprezentace (IR)

(Mrázová, Wang, 2007)

- **Regularizační technika** - vynucuje vytvoření transparentní a stabilní vnitřní struktury BP-sítě
- Chybová funkce:  $E = E_{mse} + c_F F$

$$F = \sum_p \sum_j (1 + y_{j,p})^s (1 - y_{j,p})^s y_{jp}^2$$

- Aktivity skrytých neuronů blízké hodnotám:
  - -1 ... NE
  - 1 ... ANO
  - 0 ... NEVÍM (bez odpovědi)
- Usnadňuje prořezání nadbytečných skrytých neuronů.

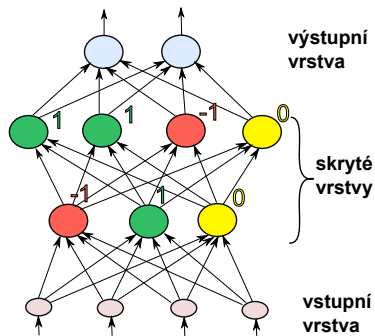


# I. Rychlá extrakce znalostí (SCGIR)

## Prořezávání skrytých neuronů založené na interní reprezentaci

(*Sietsma, Dow, 1991*)

- 1 Naučení modelu
- 2 Nalezení a odstranění / sloučení skrytých neuronů s:
  - uniformní reprezentací
  - navzájem identickou reprezentací
  - navzájem inverzní reprezentací
- 3 Přepočtení vah (doučení není třeba)



## II. Zjednodušení vnitřní struktury BP-sítě (SCGS)

### Citlivostní koeficienty jako míra relevance:

- $S_{uv,p}$  ... jak velký má malá změna  $u$ -tého vstupu síť vliv na  $v$ -tý výstup:

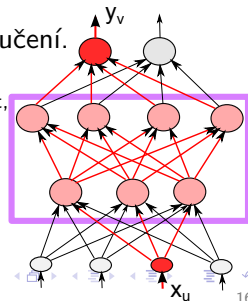
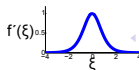
$$S_{uv,p} = \frac{\partial y_{v,p}}{\partial x_{u,p}} = \sum_k S_{kv,p} S_{uk,p} = \sum_k f'(\xi_{v,p,j}) w_{kv} S_{uk,p}$$

### Použití

- Identifikace a prořezání nadbytečných vstupních a skrytých neuronů.
- Oslabování celkové citlivosti sítě v průběhu učení.

- $u$  ... vstupy sítě,
- $v$  ... výstupy sítě,
- $k$  ... neurony v poslední skryté vrstvě,
- $x$  ... aktuální vstup,
- $y$  ... aktuální výstup,
- $p$  ... trénovací příklady,

- $S_{uv,p}$  ... citlivostní koeficient,
- $w_{kv}$  ... váha z  $k$  do  $v$ ,
- $S_{kv,p} = f'(\xi_{v,p,j}) w_{kv}$
- $f'(\xi_{v,p})$  ... derivace přenosové funkce





## II. Zjednodušení vnitřní struktury BP-sítě (SCGS)

### Strategie prořezávání

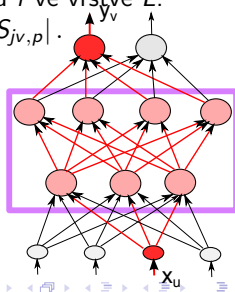
Iterativní opakování následujících kroků:

- 1 Učení (doučení) modelu
- 2 Prořezávání
  - Prořezání skrytých neuronů založené na interní reprezentaci.
  - Prořezání vstupních a skrytých neuronů na základě spočtených citlivostních koeficientů.

Heuristické pravidlo pro prořezání neuronu  $i$  ve vrstvě  $L$ :

$$\max_p \text{mean}_v |S_{iv,p}| < \delta_L = \beta \text{mean}_{\{v,j,p\}} |S_{jv,p}|.$$

- 
- $v$  ... výstupy sítě,
  - $j$  ... neurony ve vrstvě  $L$ ,
  - $p$  ... trénovací příklady,
  - $S_{iv,p} = \frac{\partial y_{v,p}}{\partial y_{i,p}}$  ... citlivostní koeficient,
  - $y_{i,p}$  ... výstup neuronu  $i$  pro trénovací příklad  $p$ ,
  - $0 < \beta \leq 1$  ... konstantní parametr.



## II. Zjednodušení vnitřní struktury BP-sítě (SCGS)

### Analytická metoda pro snižování citlivosti sítě (SC)

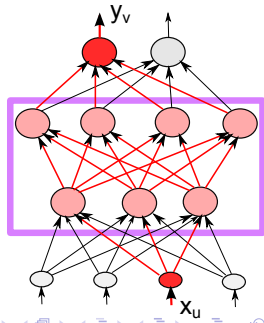
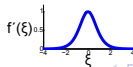
- Snižování citlivosti výstupů sítě ke vstupům.
- Chybová funkce:  $H = E_{mse} + c_G G$

$$G = \frac{1}{2} \sum_p \sum_u \sum_v S_{uv,p}^2, \quad S_{uv,p} = \frac{\partial y_{v,p}}{\partial x_{u,p}} = \sum_k f'(\xi_{v,p}) w_{kv} S_{uk,p}$$

- Minimalizace  $G$  v průběhu učení je **velmi výpočetně náročná**.

- $u$  ... vstupy sítě,
- $v$  ... výstupy sítě,
- $k$  ... neurony v poslední skryté vrstvě,
- $x$  ... aktuální vstup,
- $y$  ... aktuální výstup,
- $p$  ... trénovací příklady,

- $S_{uv,p}$  ... citlivostní koeficient,
- $w_{kv}$  ... váha z  $k$  do  $v$ ,
- $f'(\xi_{v,p})$  ... derivace přenosové funkce



## II. Zjednodušení vnitřní struktury BP-sítě (SCGS)

### Kombinace následujících technik:

- + Metoda škálovaných konjugovaných gradientů (SCG)
- + Regularizace:
  - + Metoda vynucované kondenzované interní reprezentace (IR)
  - + **Analytická metoda pro snižování citlivosti sítě (SC)**
- + Prořezávání založené na:
  - + citlivostní analýze
  - + interní reprezentaci

- 
- Chybová funkce  $H = E_{mse} + c_F F + c_G G$ 
    - $E_{mse}$  ... standardní chybová funkce (MSE),
    - $F$  ... chybová funkce regulující interní reprezentaci,
    - $G$  ... nově navržená chybová funkce pro snižování citlivosti,
    - $c_F, c_G$  ... parametry regulující váhu  $E, F$  and  $G$  v  $H$ .

# Aproximativní metoda pro snižování citlivosti sítě (SCA)

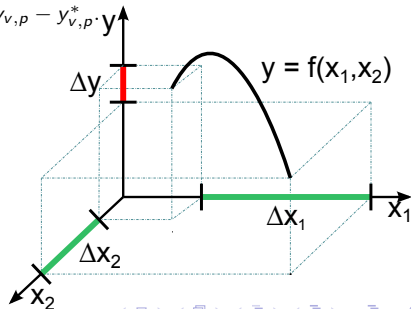
Alternativní vzorec pro citlivostní chybovou funkci:

$$G = \frac{1}{2} \sum_p \sum_u \sum_v \left( \frac{\partial y_{v,p}}{\partial x_{u,p}} \right)^2 \rightarrow G_1 = \frac{1}{2} \sum_p \sum_u \sum_v \left( \frac{\Delta y_{v,p}}{\Delta x_{u,p}} \right)^2$$

- $x_{u,p}^*$  ...  $x_{u,p}$  s přidáním 0.1-1% náhodným šumem,
- $y_{u,p}^*$  ... výstup sítě pro  $x_{u,p}^*$ ,
- $\Delta x_{u,p} = x_{u,p} - x_{u,p}^*$  a  $\Delta y_{v,p} = y_{v,p} - y_{v,p}^*$ .

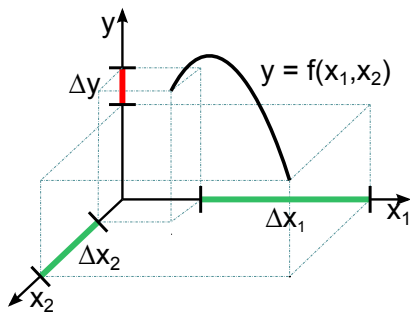
**Příklad:**

$$\begin{aligned} G_1 &\approx \left( \frac{\Delta y}{\Delta x_1} \right)^2 + \left( \frac{\Delta y}{\Delta x_2} \right)^2 = \\ &= \frac{(y - y^*)^2}{(x_1 - x_1^*)^2} + \frac{(y - y^*)^2}{(x_2 - x_2^*)^2} \end{aligned}$$

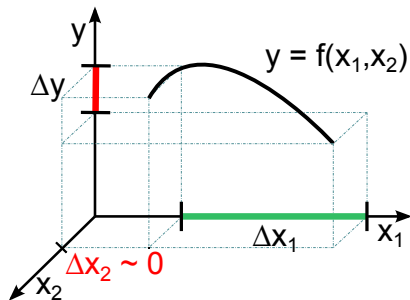


# Aproximativní metoda pro snižování citlivosti sítě (SCA)

## Příklad:



## Problém:



$$G_1 \approx \left( \frac{\Delta y}{\Delta x_1} \right)^2 + \left( \frac{\Delta y}{\Delta x_2} \right)^2 = \frac{(y - y^*)^2}{(x_1 - x_1^*)^2} + \frac{(y - y^*)^2}{(x_2 - x_2^*)^2}$$

malá hodnota jmenovatele  $\rightarrow$  velké problémy kvůli zaokrouhlení výsledku

# Aproximativní metoda pro snižování citlivosti sítě (SCA)

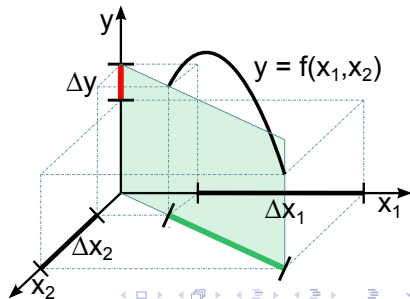
**Řešení:** Alternativní vzorec pro citlivostní chybovou funkci:

$$G_1 = \frac{1}{2} \sum_p \sum_u \sum_v \left( \frac{\Delta y_{v,p}}{\Delta x_{u,p}} \right)^2 \rightarrow G_2 = \frac{1}{2} \sum_p \frac{\sum_v \Delta^2 y_{v,p}}{\sum_u \Delta^2 x_{u,p}}$$

- $\Delta x_{u,p} = x_{u,p} - x_{u,p}^*$  a  $\Delta y_{v,p} = y_{v,p} - y_{v,p}^*$ .
- $x_{u,p}^* \dots x_{u,p}$  s přidáním 0.1-1% náhodným šumem,
- $y_{u,p}^* \dots$  výstup sítě pro  $x_{u,p}^*$ ,

**Příklad:**

$$\begin{aligned} G_2 &\approx \frac{\Delta^2 y}{\Delta^2 x_1 + \Delta^2 x_2} = \\ &= \frac{(y - y^*)^2}{(x_1 - x_1^*)^2 + (x_2 - x_2^*)^2} \end{aligned}$$



# Rychlé učení s vytvořením jednodušší a jasnější vnitřní struktury sítě (SCGSA)

## Kombinace následujících technik:

- + Metoda škálovaných konjugovaných gradientů (SCG)
- + Regularizace:
  - + Metoda vynucované kondenzované interní reprezentace (IR)
  - + **Aproximativní metoda pro snižování citlivosti sítě (SCA)**
- + Prořezávání založené na:
  - + citlivostní analýze
  - + interní reprezentaci

- 
- Chybová funkce  $H = E_{mse} + c_F F + c_G G_1$ 
    - $E_{mse}$  ... standarní chybová funkce (MSE),
    - $F$  ... chybová funkce regulující interní reprezentaci,
    - $G_1$  ... nově navržená chybová funkce pro snižování citlivosti,
    - $c_F, c_G$  ... parametry regulující váhu  $E, F$  and  $G_1$  v  $H$ .

# Experimenty

## Testované vlastnosti:

- 1 Schopnost zobecňovat
- 2 Rychlost učení
- 3 Stabilita metod
- 4 Vytvoření jednoduché a transparentní struktury modelu

## Testovací data:

- **Diskrétní data** – Binární sčítání a násobení 3-bitových čísel
- **Spojitá data** – Světová Banka (Indikátory světového vývoje)

- 
- Ke každé datové sadě jsme přidali 10-12 irelevantních (náhodně vygenerovaných) vstupních příznaků.
  - Testovány různé počáteční topologie BP-sítí s jednou či dvěma skrytými vrstvami.

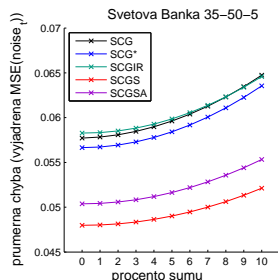
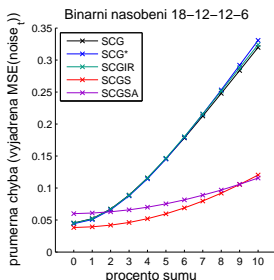
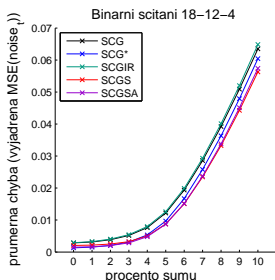


# Souhrnné výsledky experimentů

Metoda	Schopnost zobecňovat	Rychlost	Robustnost k parametrům	Transparentnost	Struktura
SCG	*** dobrá	***** rychlý	***** robustní	* není vynucovaná	* neprořezává
SCGIR	*** srovnatelná s SCG	**** nejvýše dvakrát pomalejší než SCG	*** relativně citlivý k volbě $c_F$	***** silně vynucovaná pro větší hodnoty $c_F$	*** prořezávání skrytých neuronů
SCGS	***** lepší než SCG pro diskrétní data a více skrytých vrstev	* extrémně pomalá	***** robustní k volbě $c_F$ a $c_G$	**** středně, ale stabilně vynucovaná	***** prořezávání skrytých a vstupních neuronů
SCGSA	***** lepší než SCG pro diskrétní data a více skrytých vrstev	**** nejvýše dvakrát pomalejší než SCG	***** robustní k volbě $c_F$ a $c_G$	**** středně, ale stabilně vynucovaná	***** prořezávání skrytých a vstupních neuronů

# Robustnost k šumu v testovacích datech

- Průměrná chyba MSE na testovacích datech pro různé hodnoty šumu v datech:

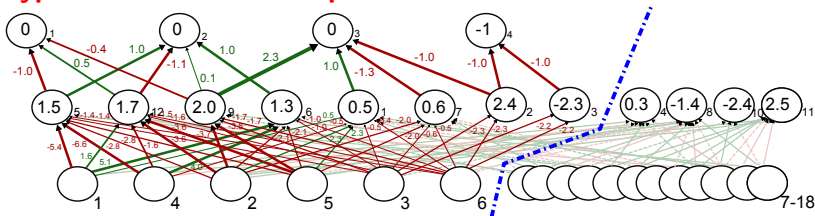


- Přidání  $p$ -procentního šumu k testovacímu příkladu  $x_q$ :

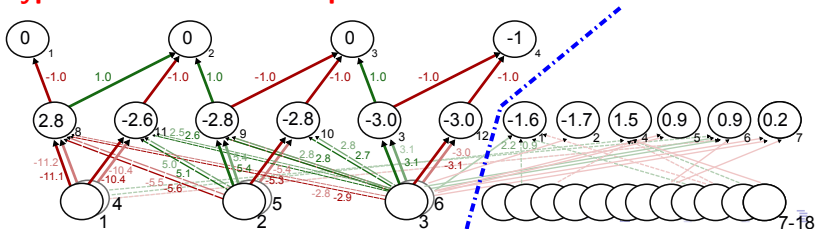
$$x_{i,q}^{noise} = x_{i,q}(1 + 0.01 p \text{ rand}(\{0, 1, -1\}))$$

# Vytvoření jednoduché a transparentní struktury během učení – Binární sčítání

## Typická BP-síť naučená pomocí SCG

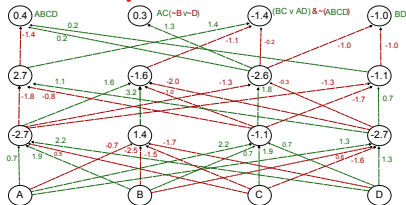


## Typická BP-síť naučená pomocí SCGSA

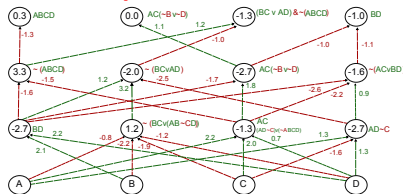


# Vytvoření jednoduché a transparentní struktury během učení – Binární násobení

## Typická BP-síť naučená pomocí SCG



## Typická BP-síť naučená pomocí SCGS



# Data ze Světové Banky - Indikátory světového vývoje

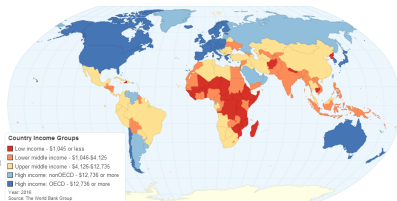
(25 indikátorů, 162 zemí, 2001-2006, 956 trénovacích vzorů)

## Vstupní příznaky:

1	PPP konverzní faktor, místní měna vzhledem k mezinárodnímu \$	17	Růst HDP, ročně, %
2	Míra PPP konverzního faktoru vzhledem k oficiálnímu směnnému kurzu	18	Vývoz vyspělých technologií, % vývozu průmyslového zboží
3	Aktuální hodnota státního dluhu, % vývozu	19	Inflace, deflace HDP (ročně, %)
4	Krátkodobý dluh, % státního dluhu	20	Uživatelé internetu, na 1000 obyvatel
5	Krátkodobý dluh, % vývozu	21	Průměrná délka života, roky
6	Celkový státní dluh, % HNP	22	Výdaje na zbrojení, % HDP
7	Giniho koeficient	23	Počet obyvatel pod hranicí národní míry chudoby, % obyvatelstva
8	Příjem státního rozpočtu z daní, % HDP	24	Aktuální hodnota státního dluhu, % HNP
9	Daně z příjmu, zisku a investic, % příjmů	25	Celkový státní dluh, % vývozu
10	Daně ze zboží a služeb, % příjmů	26..35	Náhodně generované vstupní příznaky
11	Daně z mezinárodního obchodu, % příjmů		
12	Sociální příspěvky, % příjmů		
13	Výdaje na výzkum a školství, % HDP		
14	Deflace HDP, % růstu		
15	Porodnost, celkem (počet porodů na ženu)		
16	Vlastníci pevných a mobilních telefonů, na 1000 obyvatel		

## Výstupní příznaky:

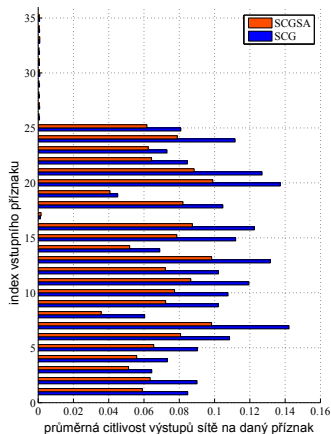
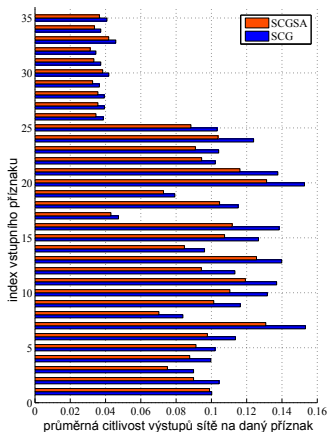
- 1..5 | Příjmová skupina (Vysoký příjem & člen OECD, Vysoký příjem, Vyšší střední příjem, Nižší střední příjem, Nízký příjem)



# Citlivostní analýza

## Průměrná citlivost naučených BP-sítí

$S_u = \text{mean}_{\{v,p\}} |S_{uv,p}|$ , průměr přes 100 různých inicializací  
**před prořezáním:** **po prořezání:**

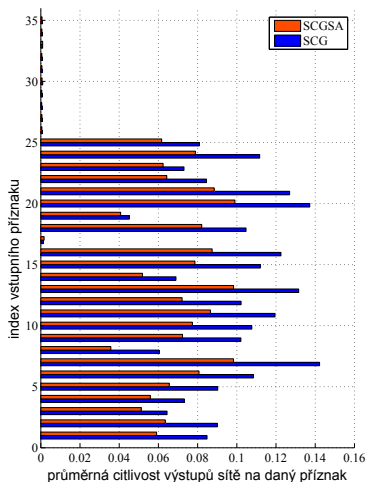


# Citlivostní analýza

## Pořadí nejdůležitějších příznaků podle citlivostních koeficientů:

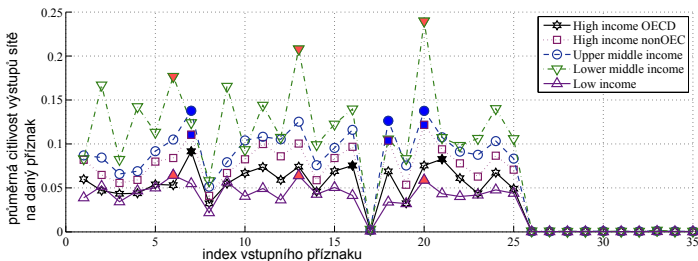
pořadí	SCG	SCGSA	význam (SCGSA)
1.	7	20	<b>Uživatelé internetu</b>
2.	20	7	Giniho koeficient
3.	13	13	Výdaje na výzkum a školství
4.	16	11	Daně z mezinárodního obchodu
5.	21	21	<b>Průměrná délka života</b>
6.	11	16	<b>Vlastníci mobilních či pevných telefonů</b>
7.	10	10	Daně ze zboží a služeb
8.	15	15	<b>Porodnost</b>
...	...	...	
23.	8	19	Inflace, deflace HDP
24.	19	8	Příjem rozpočtu z daní
25.	17	17	Růst HDP
26..35.	26..35	26..35	Náhodně generované příznaky

## Průměrná citlivost naučených BP-sítí po prožezání:



# Citlivostní analýza

## Průměrná citlivost pro jednotlivé příjmové skupiny (metoda SCGSA, po prožezání)



7. Giniho koeficient
21. Průměrná délka života
16. Vlastníci telefonního připojení

7. Giniho koeficient
20. Uživatelé internetu
18. Vývoz vospělých technologií

20. Uživatelé internetu
13. Výdaje na výzkum a školství
6. Celkový státní dluh



# Závěr

Navržený framework zlepšuje schopnost naučených BP/sítí:

- 1 Dobře **zobecňovat**.
- 2 **Rychle** vytvořit **jednodušší a transparentní strukturu** neuronové sítě.
- 3 Adekvátně **prořezat** vstupní i skryté neurony.
  - **Nižší citlivost** modelu k volbě parametrů a k šumu v datech.
  - Snadnější interpretace extrahovaných znalostí.
  - **Nejlepší výsledky** pro **diskrétní data** a pro sítě **s více skrytými vrstvami**.

# Závěr

Děkuji Vám za pozornost.

# Rychlost učení a schopnost zobecňovat

Binární sčítání – topologie 18-12-4										
method	$c_F$	$c_G$	$n_A$	$c_R$	$c_A$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCG	–	–	20	14	3	$0.035 \pm 0.028$	1.0	$0.09 \pm 0.12$	1413.7	5.9
SCG*	–	–	23	25	13	$0.031 \pm 0.035$	1.1	$0.09 \pm 0.09$	1312.9	7.1
SCGIR	$10^{-6}$	–	25	39	18	$0.036 \pm 0.036$	1.0	$0.08 \pm 0.11$	1429.2	11.2
SCGS	–	$5 \cdot 10^{-7}$	22	33	11	$0.028 \pm 0.022$	1.3	$0.07 \pm 0.07$	1636.3	420.5
SCGS	$10^{-6}$	$5 \cdot 10^{-7}$	21	39	11	$0.027 \pm 0.020$	1.3	$0.07 \pm 0.06$	1625.7	420.1
<b>SCGSA</b>	–	$10^{-5}$	22	29	12	$0.028 \pm 0.020$	1.2	$0.09 \pm 0.09$	1308.6	8.4
<b>SCGSA</b>	$10^{-6}$	$10^{-5}$	19	31	10	$0.029 \pm 0.022$	1.2	$0.08 \pm 0.08$	1378.3	13.5

Binární násobení – topologie 18-12-12-6										
method	$c_F$	$c_G$	$n_I$	$c$	$c_n$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCG	–	–	33	51	0	$0.297 \pm 0.189$	1.0	$0.28 \pm 0.22$	601.0	6.1
SCG*	–	–	38	49	0	$0.300 \pm 0.210$	1.0	$0.26 \pm 0.22$	601.0	10.4
SCGIR	$10^{-6}$	–	35	50	0	$0.301 \pm 0.191$	1.0	$0.28 \pm 0.23$	601.0	10.5
SCGS	–	$5 \cdot 10^{-7}$	34	55	11	$0.098 \pm 0.057$	3.0	$0.14 \pm 0.09$	601.0	3733.8
SCGS	$10^{-6}$	$5 \cdot 10^{-7}$	31	53	10	$0.097 \pm 0.060$	3.0	$0.14 \pm 0.09$	601.0	3620.7
<b>SCGSA</b>	–	$5 \cdot 10^{-5}$	30	38	19	$0.085 \pm 0.064$	3.5	$0.11 \pm 0.06$	601.0	8.3
<b>SCGSA</b>	$10^{-6}$	$5 \cdot 10^{-5}$	27	42	16	$0.079 \pm 0.053$	3.8	$0.10 \pm 0.06$	600.8	13.3

- Nejlepší výsledky pro binární násobení.
- Výrazné zvýšení schopnosti zobecňovat a snížení citlivosti k šumu v datech.
- Časová složitost SCGSA je srovnatelná s SCG.

# Rychlost učení a schopnost zobecňovat

Světová banka – topologie 35-50-5							
method	$c_F$	$c_G$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCG	–	–	$0.063 \pm 0.017$	1.00	$0.09 \pm 0.03$	969.0	23.6
SCG*	–	–	$0.062 \pm 0.017$	1.02	$0.09 \pm 0.03$	974.5	32.0
SCGIR	$10^{-6}$	–	$0.063 \pm 0.017$	1.00	$0.09 \pm 0.03$	968.0	48.5
SCGS	–	$10^{-5}$	$0.051 \pm 0.016$	1.23	$0.05 \pm 0.02$	964.0	2036.0
SCGS	$10^{-6}$	$10^{-5}$	$0.050 \pm 0.016$	1.25	$0.05 \pm 0.02$	968.4	1984.2
<b>SCGSA</b>	–	$2 \cdot 10^{-4}$	$0.054 \pm 0.015$	1.17	$0.06 \pm 0.03$	978.2	35.2
<b>SCGSA</b>	$10^{-6}$	$2 \cdot 10^{-4}$	$0.055 \pm 0.017$	1.14	$0.06 \pm 0.02$	945.4	59.0
Světová banka – topologie 35-15-15-5							
method	$c_F$	$c_G$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCG	–	–	$0.064 \pm 0.038$	1.00	$0.05 \pm 0.02$	305.1	15.6
SCG*	–	–	$0.057 \pm 0.025$	1.11	$0.04 \pm 0.02$	305.9	18.5
SCGIR	$10^{-6}$	–	$0.064 \pm 0.037$	0.99	$0.05 \pm 0.02$	304.5	24.2
SCGS	–	$10^{-5}$	$0.050 \pm 0.021$	1.26	$0.03 \pm 0.01$	302.7	53188.0
<b>SCGSA</b>	–	$10^{-4}$	$0.052 \pm 0.022$	1.23	$0.04 \pm 0.01$	303.4	19.8
<b>SCGSA</b>	$10^{-6}$	$10^{-4}$	$0.054 \pm 0.025$	1.19	$0.04 \pm 0.02$	304.1	28.5

- Průměrná finální topologie byla **23-23-5** a **21-12-8-5**.
- Obdobné výsledky pro obě topologie.
- Výrazné zvýšení schopnosti zobecňovat a snížení citlivosti k šumu v datech.
- Časová složitost SCGSA je srovnatelná s SCG.

Robustnost metod k volbě parametrů  $c_F$  a  $c_G$ 

## Binární sčítání – topologie 18-12-4

method	$c_F$	$c_G$	$n_I$	$c_R$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCGS	$10^{-6}$	$5 \cdot 10^{-7}$	87	39	$0.027 \pm 0.020$	1.3	$0.07 \pm 0.06$	1625.7	420.1
<b>SCGS</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>5 \cdot 10^{[-8,-6]}</math></b>	92	36	$0.022 \pm 0.011$	1.6	$0.07 \pm 0.07$	1527.2	271.8
SCGSA	$10^{-6}$	$10^{-5}$	84	31	$0.029 \pm 0.022$	1.2	$0.08 \pm 0.08$	1378.3	13.5
<b>SCGSA</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>10^{[-6,-4]}</math></b>	93	31	$0.023 \pm 0.010$	1.5	$0.08 \pm 0.06$	1341.0	12.5

## Binární násobení – topologie 18-12-12-6

method	$c_F$	$c_G$	$n_I$	$c$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCGS	$10^{-6}$	$5 \cdot 10^{-7}$	31	53	$0.097 \pm 0.060$	3.1	$0.14 \pm 0.09$	601.0	3620.7
<b>SCGS</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>5 \cdot 10^{[-8,-6]}</math></b>	35	56	$0.098 \pm 0.065$	3.0	$0.14 \pm 0.13$	601.0	3438.1
SCGSA	$10^{-6}$	$5 \cdot 10^{-5}$	27	42	$0.079 \pm 0.053$	3.8	$0.10 \pm 0.06$	600.8	13.3
<b>SCGSA</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>5 \cdot 10^{[-6,-4]}</math></b>	49	42	$0.080 \pm 0.052$	3.7	$0.11 \pm 0.07$	946.7	13.7

## Světová banka – topologie 35-50-5

method	$c_F$	$c_G$	$n_I$	$MSE(noise_t)$	imp	sensitivity	epochs	time(s)
SCGS	$10^{-6}$	$10^{-5}$	88	$0.050 \pm 0.016$	1.25	$0.05 \pm 0.02$	968.4	1984.2
<b>SCGS</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>10^{[-6,-4]}</math></b>	59	$0.054 \pm 0.016$	1.16	$0.05 \pm 0.02$	987.0	1704.0
SCGSA	$10^{-6}$	$2 \cdot 10^{-4}$	97	$0.055 \pm 0.017$	1.14	$0.06 \pm 0.02$	945.4	59.0
<b>SCGSA</b>	<b><math>10^{[-7,-5]}</math></b>	<b><math>2 \cdot 10^{[-5,-3]}</math></b>	86	$0.058 \pm 0.016$	1.09	$0.07 \pm 0.03$	990.6	48.3