

Deep Neural Network

Uživatelská příručka



Verze 1.0

©2023



Výsledek „Deep Neural Network“ vznikl v rámci projektu „Užití umělé inteligence při modernizaci diagnostiky systémových prvků a optimalizaci systémových činností energetického sektoru s cílem zvýšení kvality jeho řízení“ za finanční podpory Technologické agentury ČR.

Obsah:

1. Architektura	5
2. Data.....	6
2.1. Vstup/Výstup	6
2.2. Konfigurace.....	8
2.3. Omezení.....	9
3. Multi-Layer Perceptron.....	10
3.1. Deep Learning	11
4. Chyby.....	12
5. Demonstrační příklad	14
6. Dodatek	15

1. Architektura

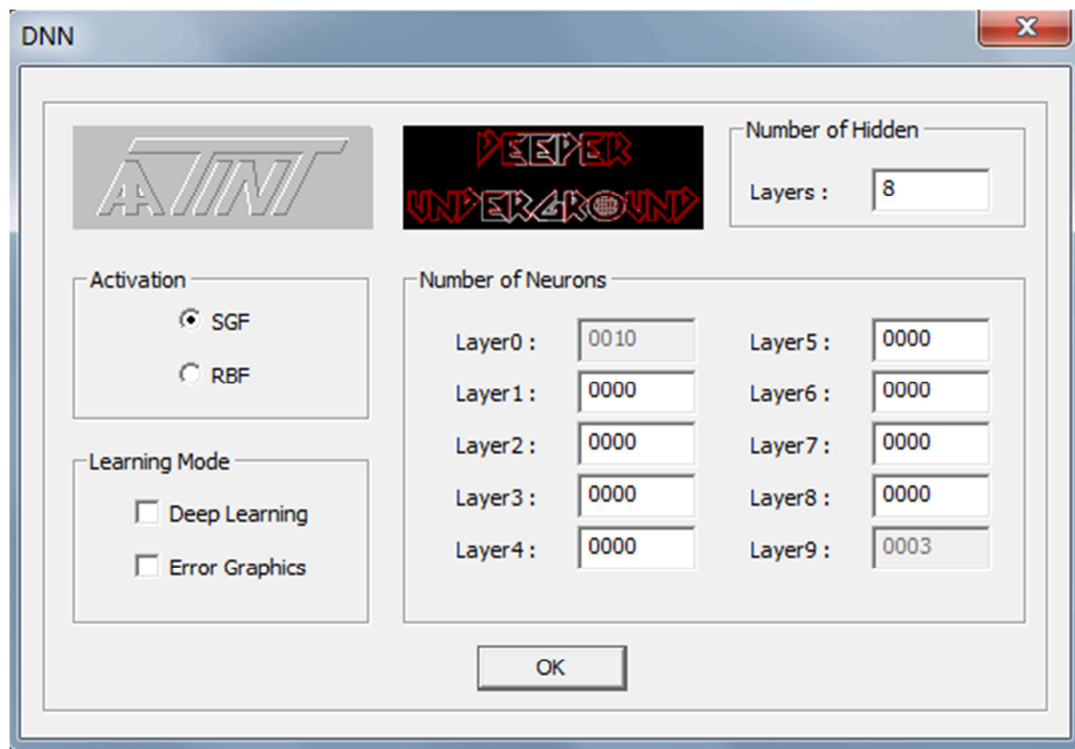
Neuronová síť je tvořena neurony, řadícími se do jednotlivých, nad sebou řazených vrstev, a jejich vzájemnými vazbami, propojujícími každý neuron nižší vrstvy s každým neuronem vyšší vrstvy. Nejnižší resp. nejvyšší vrstvu označíme jako vstupní resp. výstupní, mezilehlé vrstvy jako skryté.

Parametrizaci architektury DNN lze provést v následujících krocích:

První krok: Volba počtu skrytých vrstev včetně počtu jejich neuronů, u vstupní a výstupní vrstvy se počet jejich neuronů nastaví automaticky v závislosti předložených datech.

Druhý krok: Volba transformačních (aktivačních) funkcí (SGF, RBF) neuronů vstupní vrstvy včetně automatického nastavení jejich parametrů (prahů resp. strmostí) v závislosti na předložených datech.

Třetí krok: Volba možnosti dvoufázová adaptace (deep learning), tj. první fáze adaptace proběhne adaptací nad sebou řazených autokodérů, a volba možnosti grafického zobrazení průběhu chyby funkce sítě během druhé fáze adaptace (backpropagation).



The screenshot shows a window titled "DNN" with a close button (X) in the top right corner. The window contains several configuration options:

- Logos:** "AINT" and "DEEPER UNDERGROUND".
- Number of Hidden Layers:** A text box containing the value "8".
- Activation:** Two radio buttons: "SGF" (selected) and "RBF".
- Learning Mode:** Two checkboxes: "Deep Learning" (unchecked) and "Error Graphics" (unchecked).
- Number of Neurons:** A grid of text boxes for layers 0 through 9:
 - Layer0: 0010
 - Layer1: 0000
 - Layer2: 0000
 - Layer3: 0000
 - Layer4: 0000
 - Layer5: 0000
 - Layer6: 0000
 - Layer7: 0000
 - Layer8: 0000
 - Layer9: 0003
- OK Button:** A button at the bottom center.

Parametrizace architektury DNN

2. Data

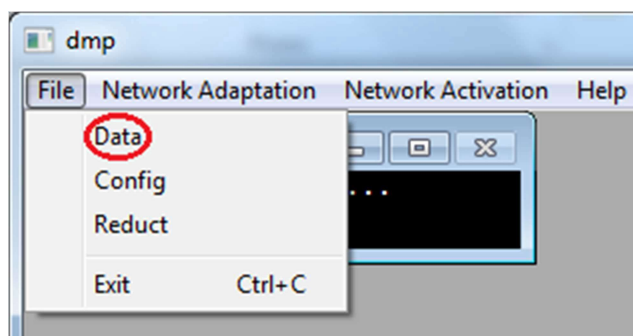
V adresáři aplikace lze editací souboru PRJ.INI parametrizovat volbu pracovního adresáře PRJ1÷PRJ9 (zadáním čísla 0÷9) jakožto podadresáře adresáře aplikace, který je pak nutno v adresáři aplikace vytvořit, zadáním čísla nula je pak pracovní adresář přímo adresář aplikace.

2.1. Vstup/Výstup

V následující tabulce je uveden přehled jmen souborů daného jména a formátu, s nimiž jednotlivé dynamiky v pracovním adresáři pracují:

	vstup	výstup	tr. mn. vstup	tr. mn. výstup	ts. mn. vstup	ts. mn. výstup
soubor	INPUT.ADI	OUTPUT.ADO	INPUT.ATI	OUTPUT.ATO	INPUT.ASI	OUTPUT.ASO
mód	aktivní	aktivní	adaptivní	adaptivní	adaptivní	adaptivní

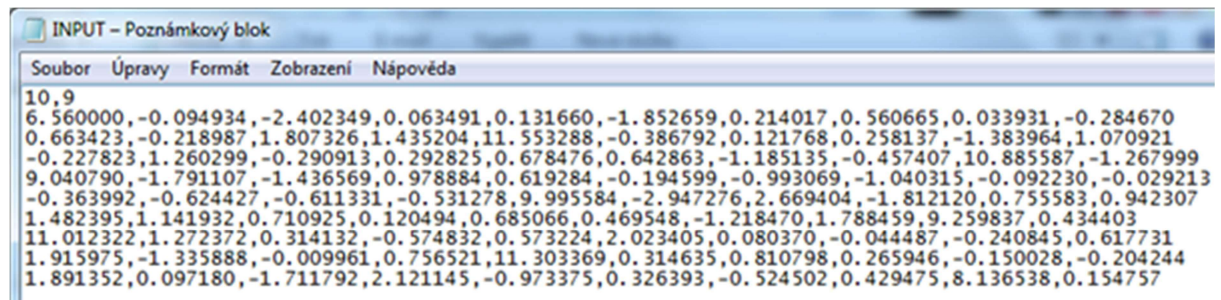
Všechny výše uvedené soubory jsou binární soubory daného formátu a lze je vytvořit (vyjma souboru OUTPUT.ADO) pomocí aplikační funkce *Data* volané v menu položce *File*:



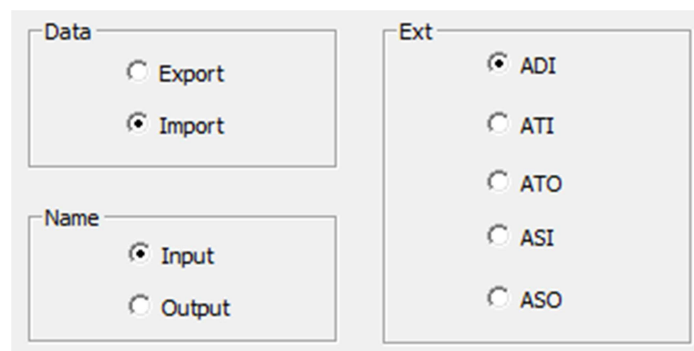
následujícím postupem:

- data pořídíme např. v tabulkovém procesoru Excel tak, aby jednotlivé řádky odpovídaly jednotlivým vzorům resp. obrazům a jednotlivé sloupce jednotlivým neuronům vstupní resp. výstupní vrstvy sítě
- naplněnou tabulku exportujeme jako text oddělený tabulátory do souboru jménem INPUT.TXT resp. OUTPUT.TXT

- v exportovaném textovém souboru nejprve hromadně nahradíme všechny desetinné čárky za tečky (je-li potřeba) a po té hromadně nahradíme všechny tabulátory odděluující jednotlivé položky za čárky
- do exportovaného textového souboru vložíme na pozici prvního řádku čárkou oddělený počet sloupců a řádků tabulky:



- v případě, že takto vytvořený textový soubor máme v pracovním adresáři, klikneme na funkci *Data* a v otevřeném dialogu zvolíme *Import* včetně příslušného jména a přípony cílového souboru, tak se nám soubor importuje do požadovaného formátu:

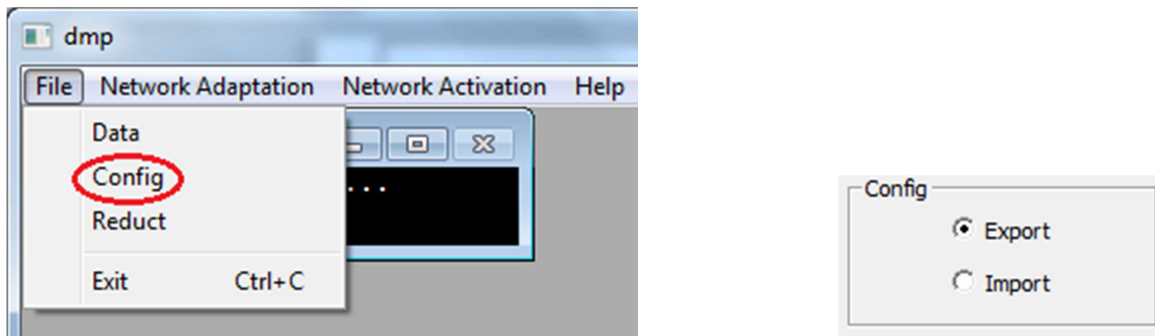


pozn.: Výše uvedeným způsobem se vytvoří soubor INPUT.ADI.

Soubor OUTPUT.ADO se automaticky vytvoří po spuštění dané funkcionality v aktivním módu, a volbou *Export* (viz výše - ostatní volby se neakceptují) jej lze exportovat do textového souboru OUTPUT.TXT, který lze následně importovat jako text o pevné šířce do tabulky „Excel“.

2.2. Konfigurace

Po spuštění vybrané funkcionality v adaptivním módu se vytvořená konfigurace sítě uloží do binárního souboru CONFIG.BIN, jehož existence v pracovním adresáři je nezbytná při spuštění dané funkcionality v aktivním módu. Uvedený soubor lze pomocí aplikační funkce *Config* volané v menu položce *File* exportovat do textových souborů vytvořených v pracovním adresáři CONFIGPAR.TXT a CONFIGWGT.TXT o pevné šířce a formátu:



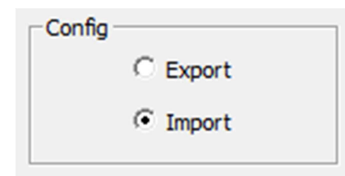
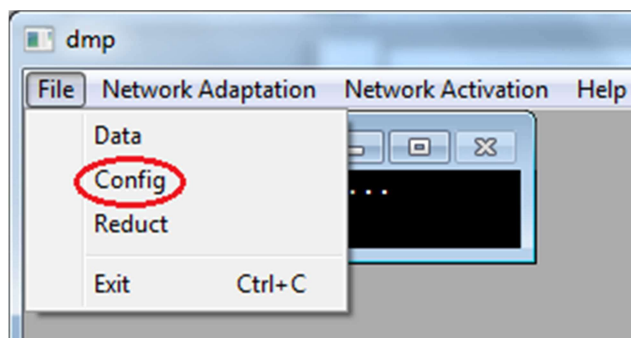
Soubor CONFIGPAR.TXT obsahuje v prvním řádku informaci o počtu neuronů ve vrstvách sítě v pořadí od nejnižší k nejvyšší a dále v pořadí sloupců: index, vrstvu, koncový potenciál a stav neuronu, a dále práh a strmost aktivační funkce neuronu:

Soubor	Úpravy	Formát	Zobrazení	Nápověda			
10	8	6	4	3	0	0	
1	1		-0.180000		0.331992		3.195614
2	1		-0.340621		0.410603		-0.007841
3	1		0.069360		0.550073		-0.139510
4	1		0.296919		0.558895		0.068412
5	1		-0.587831		0.309353		3.340922
6	1		-1.058475		0.216856		0.055455
7	1		-0.683083		0.354708		-0.128971
8	1		2.137512		0.903344		0.048677
9	1		11.243932		0.826063		3.341950
10	1		0.604973		0.609013		0.169626
1	2		0.690995		0.693858		0.789117
2	2		-0.766564		0.173951		-0.419100
3	2		0.600469		0.792069		0.284145
4	2		-0.240922		0.437790		-0.879417
5	2		0.594875		0.686793		-0.286892
6	2		-0.155451		0.459859		0.086968
7	2		-0.288061		0.425797		-0.360959
8	2		0.165016		0.554149		0.394784
1	3		0.030341		0.507887		0.808648
2	3		0.268042		0.647292		-0.460351
3	3		1.592895		0.865360		0.673033
4	3		-0.332241		0.369484		0.526037
5	3		-0.454355		0.412091		-0.865338
6	3		-0.513052		0.198473		-0.984765
1	4		-1.098196		0.236286		0.202078
2	4		0.918595		0.782248		0.361846
3	4		-1.211556		0.032711		0.510913
4	4		-0.943385		0.185909		-0.763642
1	5		-0.011332		0.326164		0.333333
2	5		-0.026929		0.318989		0.333333
3	5		1.039233		0.815485		0.333333

Soubor CONFIGWGT.TXT obsahuje v prvním řádku informaci o počtu neuronů ve vrstvách sítě v pořadí od nejnižší k nejvyšší a dále v pořadí řádků váhové vektory neuronů řazených dle indexu neuronu vyšší vrstvy, jejichž složky jsou řazené dle indexu neuronu nižší vrstvy:

Soubor	Úpravy	Formát	Zobrazení	Nápověda							
10	8	6	4	3							
-0.224834	0.042171	0.215513	-1.080658	0.762325	0.278655	-0.320248	-0.327495	0.573868	0.137536		
-0.509311	-0.438609	0.045588	0.624610	1.178758	0.196215	0.260376	-0.246486	-1.349559	0.209023		
-2.101004	-0.006834	0.369661	-0.120401	1.031054	-0.454196	0.602526	0.458600	0.269314	-0.313594		
0.240484	0.614256	0.742603	-0.321210	-0.613548	0.754302	0.964487	0.546343	-0.348962	-0.728517		
-0.941985	-0.091360	0.126855	0.882487	-0.786774	-0.049825	0.205153	-0.161257	1.142984	0.085276		
-1.066893	-0.270864	-0.673488	0.028737	-0.386066	-0.218406	0.736307	0.290968	0.220612	0.062299		
0.428065	-0.795928	0.477607	-0.580822	-0.920072	0.226210	0.081434	-0.120607	0.724571	0.060080		
0.693792	-0.448384	0.664172	-0.123311	-0.587263	-1.058454	0.437919	0.550273	-1.009183	0.032866		
0.254478	1.077702	-1.052826	-0.398255	0.215391	0.480121	-0.968477	-0.163176				
-0.185108	-1.492399	-0.064606	0.026456	1.197083	0.798116	0.397610	-0.365321				
-0.375754	-0.759889	0.915381	-0.399287	0.951256	-0.332062	0.035806	0.445953				
-1.005513	-0.308148	-1.217698	0.897667	0.187020	-0.598838	-0.008900	1.110254				
-0.461409	0.416250	0.468517	-0.151828	-0.148770	-0.538608	0.654980	0.767063				
0.840961	0.838292	1.595581	-0.421019	-0.734864	0.060892	-1.180930	-0.645268				
-0.674874	-0.797901	-0.323474	-1.119899	0.365369	0.514440						
0.442975	1.189480	0.279487	-0.722775	-0.541965	-0.955488						
0.332501	-1.231415	-1.400142	1.241491	-0.300692	-1.095508						
1.053606	-0.985515	-0.321883	0.184480	-0.183018	1.052351						
-0.245989	-0.380009	1.340109	-0.178132								
0.718475	-0.927225	-0.624465	1.160347								
0.198913	1.189444	-0.612776	-1.352795								

Výše uvedené textové soubory lze např. v poznámkovém bloku při dodržení pevného formátu ručně upravit a zpětně importovat do souboru CONFIG.BIN:



pozn.: Nulové váhy se neadaptují.

upozornění: Úpravy konfigurace sítě se doporučuje provádět v režimu přepisování, aby nebyl narušen pevný formát textového souboru CONFIGPAR.TXT resp. CONFIGWGT.TXT.

2.3. Omezení

Rozsah dat je omezen rozmezím počtu neuronů ($2 \div 1000$) v každé vrstvě sítě a maximálním počtem vzorů resp. obrazů (1 000 000) obsažených v jednom souboru. Rozsah datové položky je omezen formátem „xxxxxxx,xxxxxx“, tj. 7 znaků před a 6 znaků za desetinou čárkou resp. tečkou, kde znakem rozumíme číslici resp. na první pozici znaménko.

3. Multi-Layer Perceptron

Při spuštění funkcionality v adaptivním módu proběhne adaptace síly vazeb mezi jednotlivými neurony a adaptace prahů a strmostí aktivačních funkcí neuronů skrytých vrstev. Prahy a strmosti aktivačních funkcí neuronů vstupní a výstupní vrstvy se nastaví v závislosti na předložených datech:

Před spuštěním funkcionality zvolí uživatel řídicí parametry učení (předvoleny doporučené hodnoty):

Params		Precision	
A :	0.1	LIM :	1000
B :	0.7	PA :	0.1
C :	1.0	PR :	0.1

kde A, B a C je rychlost, setrvačnost a zrychlení učení, LIM představuje limit počtu předložení tréninkových dat a PA a PR nastavuje kritérium přesnosti funkce sítě složené z absolutní a relativní přesnosti, při jehož splnění se iterační cyklus učení přeruší ještě před dosažením zadaného limitu počtu předložení tréninkových dat.

pozn.: Při parametrizaci architektury sítě lze pro účely adaptivního módu funkcionality zvolit zaškrtnutím zobrazení průběhu chyby funkce sítě během adaptace.

Po spuštění funkcionality se v pracovním adresáři vytvoří soubor ERR.LOG, obsahující informaci o průběhu učení, tj. velikost globální chyby funkce sítě určené na tréninkové a testovací množině v každé iteraci, a dále soubor ERR.TXT, obsahující informaci o velikosti lokální chyby funkce sítě každého prvku tréninkové množiny, a to maximální chybu potenciálu neuronu výstupní vrstvy včetně jeho indexu a průměrnou chybu stavů neuronů výstupní vrstvy.

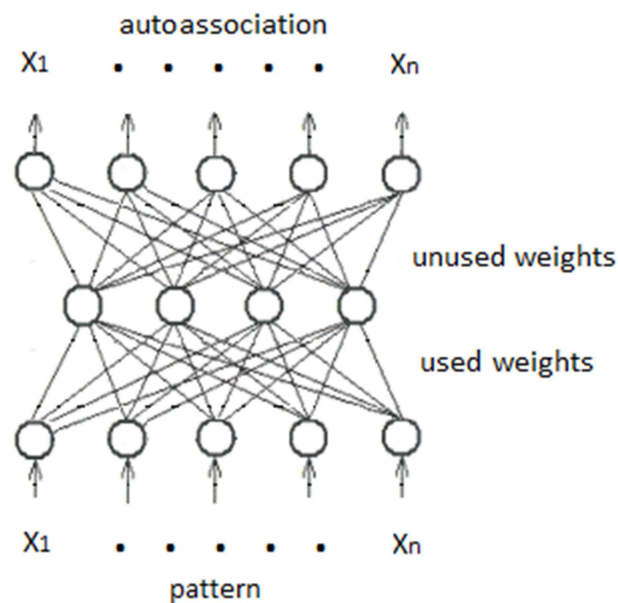
Při spuštění funkcionality v aktivním módu proběhne po řadě vyhodnocení předložených vzorů obsažených ve vstupních datech, které se uloží ve formě jejich obrazů do výstupních dat.

3.1. Deep Learning

Při parametrizaci architektury sítě lze pro účely adaptivního módu funkcionality zvolit zaškrtnutím hloubkové učení sítě, kdy se počáteční aproximace konfigurace sítě neurčí náhodně, ale adaptací autokodérů po jednotlivých vrstvách sítě ve směru od druhé nejnížší vrstvy k vrstvě nejvyšší.

V souboru INIT.INI v pracovním adresáři lze parametrizovat adaptaci autokodérů nastavením parametrů v pořadí A, B, C, NV, LIM, SEED oddělených čárkami, kde NV značí počet vrstev užitých hloubkovým učením, a to v rozmezí $2 \div 5$, a SEED je startovací hodnota generátoru náhodných čísel.

Během hloubkového učení se v pracovním adresáři vytvoří soubory ERR<x>.LOG, obsahující informaci o průběhu učení jednotlivých autokodérů, tj. velikost globální chyby funkce autokodéru v každé iteraci ($x=2 \div 5$).



Autokodér na druhé nejnížší (první skryté) vrstvě DNN

4. Chyby

Skončí-li funkcionální síť zprávou „Action aborted“, popis chyby se zapíše do souboru ERROR.LOG vytvořeného v pracovním adresáři:

Data consistency error

Nebyly nalezeny požadované soubory, nebo si vzájemně neodpovídají počty v nich zadaných řádků resp. sloupců.

Main dialog error

Nebyl dodržen pevný formát zadané položky hlavního dialogu (architektura sítě).

Dimensions error

Nebyly dodrženy dané limity počtu neuronů resp. počtu vzorů resp. obrazů obsažených v jednom souboru.

I/O error

Vyskytla se chyba čtení resp. zápisu při práci s předmětnými soubory.

Initialization error

Vyskytla se chyba inicializace konfigurace sítě, tj. chyba čtení souboru DL.INI.

pozn.: Soubor ERROR.LOG obsahuje pořadí lineárně závislých vzorů ve vstupních datech.

Save configuration error

Vyskytla se chyba zápisu konfigurace sítě.

Parameters dialog error

Nebyl dodržen pevný formát či meze zadané položky řídicích parametrů:

Parametr	minimum	maximum
A	0.0	0.9
B	0.0	0.9
C	1.0	1.9
LIM	0	999999
PA	0.0	100.0
PR	0.0	1.0

Adaptation error

Vyskytla se chyba adaptace sítě.

Activation error

Vyskytla se chyba aktivace sítě.

Action aborted

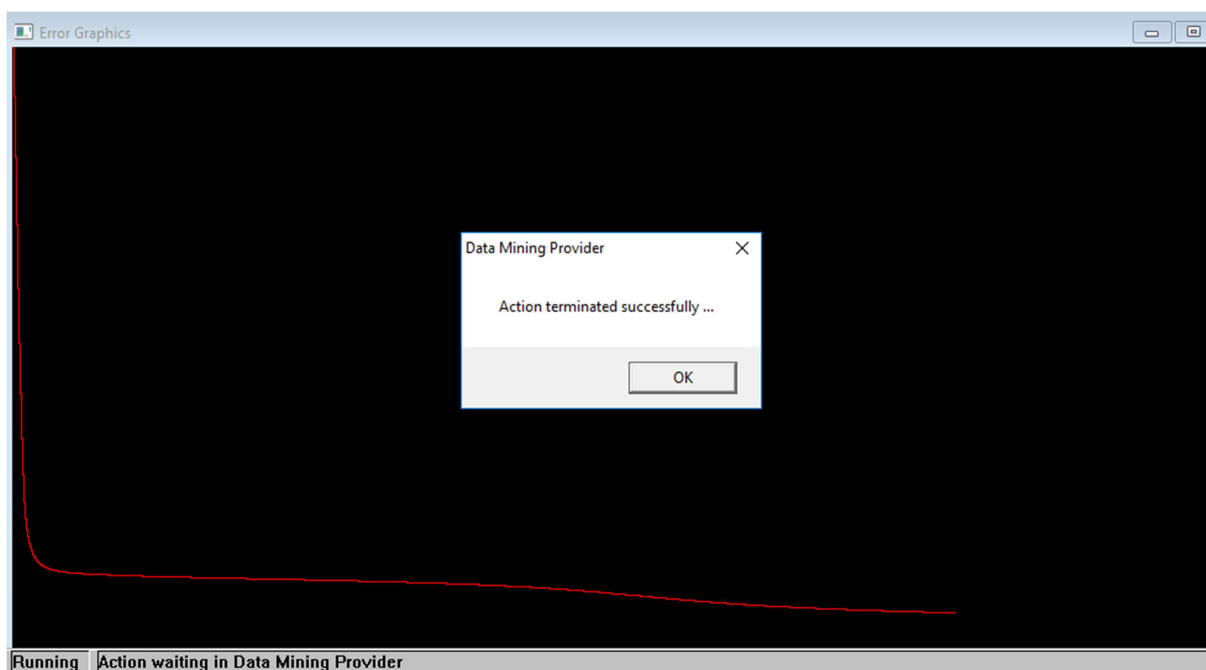
Vyskytla se neočekávaná chyba.

pozn.: Skončí-li aplikační funkce (*Data*, *Config*, *Reduct*) zprávou „Program aborted“, vyskytla se chyba čtení resp. zápisu při práci s předmětnými soubory.

5. Demonstrační příklad

V adresáři aplikace je uveden příklad užití MLP funkcionality pro úlohu klasifikace deseti složkových objektů z hlediska tří kategorií vždy dle dominantní složky, a to jednak v původní formě „Excel“ souborů a jednak ve formě binárních souborů aplikace DNN.

Soubory INPUT.ATI resp. OUTPUT.ATO tvoří tréninková data sloužící pro adaptaci sítě, soubory INPUT.ASI resp. OUTPUT.ASO tvoří na tréninkových datech nezávislá testovací data sloužící k validaci funkce sítě během její adaptace (nejsou-li k dispozici, mohou být nahrazena tréninkovými daty) a v souboru INPUT.ADI jsou obsažena data pro aktivaci sítě, tj. objekty určené ke klasifikaci pomocí již adaptované sítě. Po průběhu adaptace sítě:



jejíž výsledek lze ověřit kontrolou vygenerovaných souborů ERR.LOG a ERR.TXT, a po spuštění aktivace sítě se vytvoří soubor OUTPUT.ADO, obsahující výsledky klasifikace objektů obsažených v souboru INPUT.ADI, který lze exportovat do souboru OUTPUT.TXT.

6. Dodatek

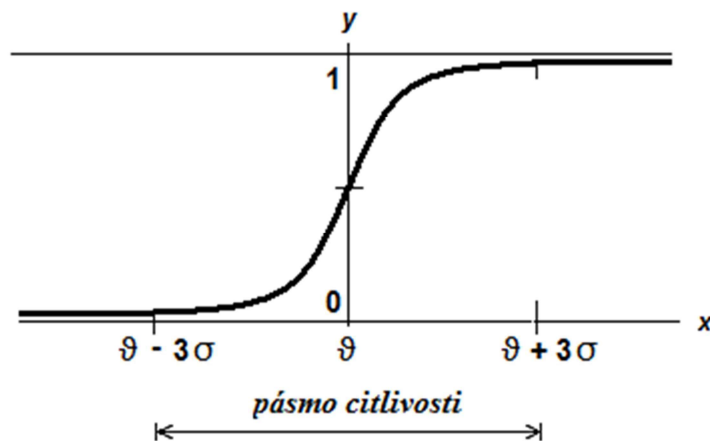
Dále jsou uvedeny alternativní způsoby transformace dat včetně nastavení parametrů aktivačních funkcí neuronů, kde σ představuje směrodatnou odchylku od střední hodnoty ϑ dat přiváděných vždy na daný neuron.

SGF transformace: $f(x) = (1 + e^{-p(x-\vartheta)})^{-1}$

$$0,95 = (1 + e^{-p(\vartheta+3\sigma-\vartheta)})^{-1} \Rightarrow e^{-3p\sigma} = \frac{0,05}{0,95}$$

$$\Rightarrow p = \frac{\ln 0,95 - \ln 0,05}{3\sigma} \cong \frac{1}{\sigma}$$

$$0,05 = (1 + e^{-p(\vartheta-3\sigma-\vartheta)})^{-1} \Rightarrow e^{3p\sigma} = \frac{0,95}{0,05}$$



RBF transformace: $f(x) = e^{-p(x-\vartheta)^2}$

$$0,05 = e^{-p(\vartheta \pm \sqrt{6}\sigma - \vartheta)^2} \Rightarrow p = -\frac{1}{6\sigma^2} \ln 0,05 \cong \frac{1}{2\sigma^2}$$

