Slovenská technická univerzita

Fakulta informatiky a informačných technológií

Ilkovičova 3, 842 16 Bratislava 4

**Dopredná neurónová sieť**

**Martin Čertek**

Predmet: Neurónové siete

Vedúci projektu: doc. Ing. Michal Čerňanský, PhD.

Ak. rok: 2013/2014

# Zadanie

V ľubovolnom programovacom jazyku implementujte jednoduchý simulátor doprednej neurónovej siete (vstupy a parametre simulácie sa musia dať meniť bez nutnosti kompilovania). Natrénujte ju algoritmom spätného šírenia chyby na vstupných množinách:

* XOR - 2 vstupné neuróny , 1 výstupný neuron, 4 vstupné vzory
* Binárna parita – 8 vstupných neurónov, 1 výstupný neuron, 256 vstupných vzorov
* vlastnej množine dát

Sieť trénujte na celej vstupnej množine a počas trénovania zobrazujte percentuálnu úspešnosť klasiﬁkácie (kolko percent zo vstupov je uz správne klasiﬁkovaných).

# Návrh siete

Riešenie vychádza z požiadaviek a je navrhnuté ako dopredná neurónová sieť zo spätnou propagáciou chyby.

### Definícia výpočtu váhy

edge.weight = (learningRate \* self.lastOutput \* (1 - self.lastOutput) \* self.error \* self.lastInput[i]) + (momentum \* edge.lastWeight)

, kde lastOutput sa rovná f(x)

# Implementácia

Riešenie zadanie je implementované v jazyku python. Implementácia je rozdelená do 3 samostatnývh súborov. Prvý obsahuje vytvorenie a prácu zo sieťou, druhý slúži ako pomocný pre prácu zo súbormi (načítanie), tretí ako hlavný súbor, kde prebieha volanie siete a operácie nad ňou a dátami.

Riešenie je v podobe konzolovej aplikácie s jednoduchým textovým výstupom, bez grafických prvkov.

Riešenie je založené na objektovom prístupe. Vychádza z viacerých definovaných objektov:

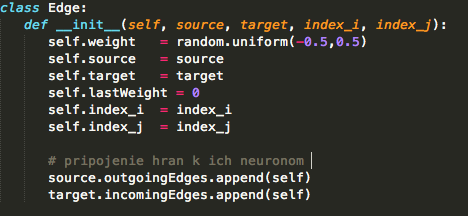
### Network (neurónová sieť)

Návrh vychádza z konceptu objektu – network, kt. uchovávat list vstupných Nodov (vstupných neurónov inputNode) a hiddenNodes/outputNode, kotré sú objektmi typu Node.

### Node (neuron)

### Macintosh HD:Users:matoc:Desktop:Snímka obrazovky 2014-03-18 o 19.08.55.png

### Edge (hrana)



Výpočet váhy:Macintosh HD:Users:matoc:Desktop:Snímka obrazovky 2014-03-18 o 20.32.40.png

# Dáta

## XOR

xorData = [

((0,0), 0),

((0,1), 1),

((1,0), 1),

((1,1), 0),

]

## Parita

Dátovú množinu tvoria všetky možné kombinácie (1/0) pri dĺžke 8 bit = 256 možností. Po každej možnosti nasleduje informácia (1/0), o nepárnosti/párnosti počtu „1“ v reťazci.

Dáta su v implementačním jazyku preprezentované ako množina *touple* :

[((0, 0, 0, 0, 0, 0, 0, 0), 0),

((0, 0, 0, 0, 0, 0, 0, 1), 1),

((0, 0, 0, 0, 0, 0, 1, 0), 1),

...

((1, 1, 1, 1, 1, 1, 1, 1), 0)

]

## Telco

Doplňujúcou množinou dát na overenie fungovania neurónovej siete je množina dát z telekomunikačnej oblasti. Táto množina zachytáva:

* demografické faktory
  + pohlavie, miesto bydliska – v rôznych rozsahoch, získané vzdelanie (SŠ/VŠ)
* časový interval, kolko rokov je osoba zákazníkom spoločnosti
* rozhodnutie o zrušení zákaznického vzťahu

Dáta sú tvorené 12 volitelnými atribútmi a 1 ktorý predstavuje výsledok. Počet záznamov je 3000.

Dáta su v implementačním jazyku uložené v množine *touple*

[((1,1,0,0,0,1,0,0,0,1,0,0),1),

((0,1,1,0,0,0,0,0,0,0,1,1),1),

((0,1,0,0,0,0,1,0,0,1,0,1),1),

...

((1,0,1,0,0,0,0,0,1,0,0,1),1)]

Z povahy dát možno očakávať niektoré závery:

na základe skúseností a komunikácie s oddelením predaja, najčastejšei odchádzajúci sú klienti, ktorí sú tesne po konci prvej viazanosti, teda v dĺžke kontraktu (1-3 roky), sú mužského pohlavia, vysokoškolsky vzdelaní a žijú vo väčších mestách a sú vo vekocom rozpätí 18-25 rokov.

Podľa informácií zo spoločnosti sú takýto klienti najnáchylnejší na odchod ku konkurencii, no nepredstavujú extrémne odlišen sa chovajúcu množinu. Rozdiel v migrovaní u nich a u množiny rovnakej, resp. dolišujúcej sa iba ženským pohlavím je niekoľko percentný, nie niekoľko násobný.

Na základe týchto postulátov budeme nazerať na dáta, ktoré budú v časti Testovanie podrobnejšie overené.

# Použitie siete

Riešenie je implementované ako konzolová aplikácia v jazyku python, spstenie siete a natrénovanie je realizované zadaním príkazu, s konfigurovatelnou množinou atribútou špecifikujúccih jednotlivé parametre pre beh danej siete.

Macintosh HD:Users:matoc:Desktop:Snímka obrazovky 2014-03-18 o 20.28.14.png

### Parametre:

*xor/parity/telco:*

definuje s akými dátami bude sieť pracovať

*learning* *rate*

učiaci faktor (hodnoty 0-1)

*momentum*

*aplikované hodnoty* (hodnoty 0-1)

*maximálny počet iterácií pre beh siete*

- celé číslo (pre rôzne vstupy je potrebné nastaviť iné parametre – příliš veľké číslo znamená presnejšie natrénovanie, no násobne rasit zložitosť výpočtu a teda aj času)

*počet vstupných neurónov*

zodpovedá počtu vstupných informácií pre každý vstupný vektor

*počet neurónov v sktyrej vrstve*

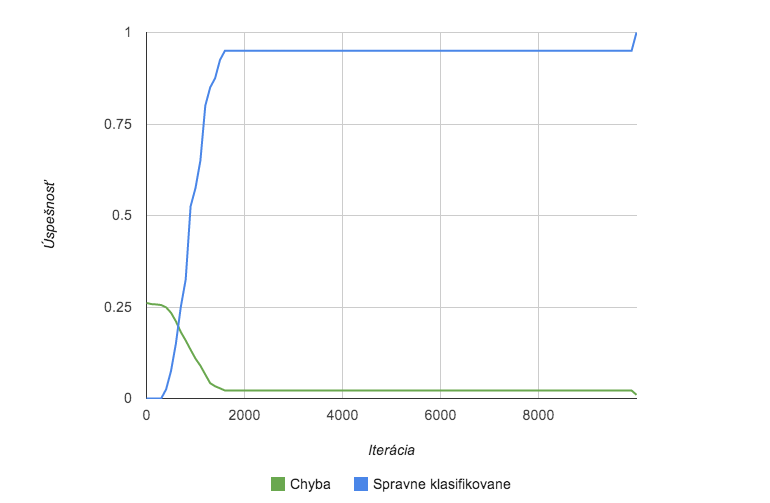
početnosť neurónov – parameter meniaci sa s počtom vstuponých neurónov v rozmedzí: počet vstupných neurónov – n (prie behu siete bolo skúšaných max 2 počet vstupných neurónov)

# Testovanie

## XOR

*parametre experimentu:*

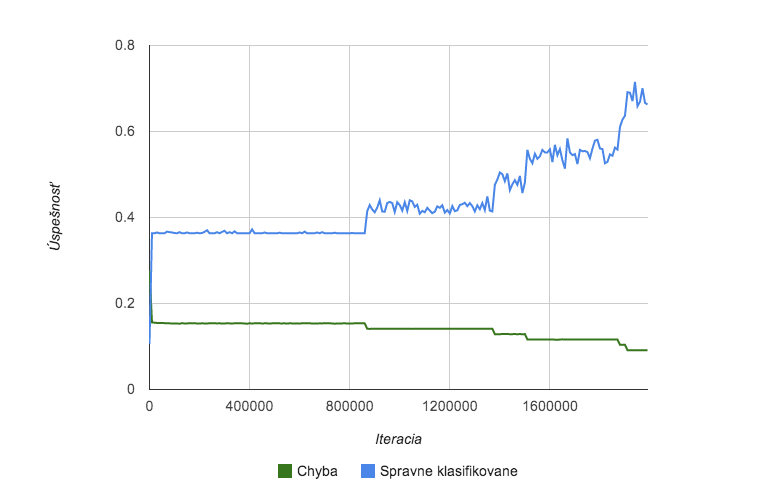
learning rate 0,3 ; momentum 0,7; 2 neurony v skrytej vrstve,



## Parita

*Experiment s parametrami:*

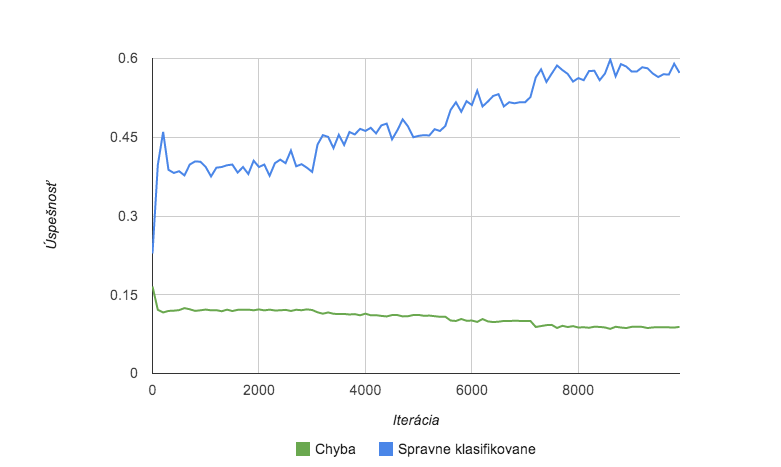
learning rate 0,3; momentum 0,7; 8 neurónov v skrytej vrstve



## Telco dáta

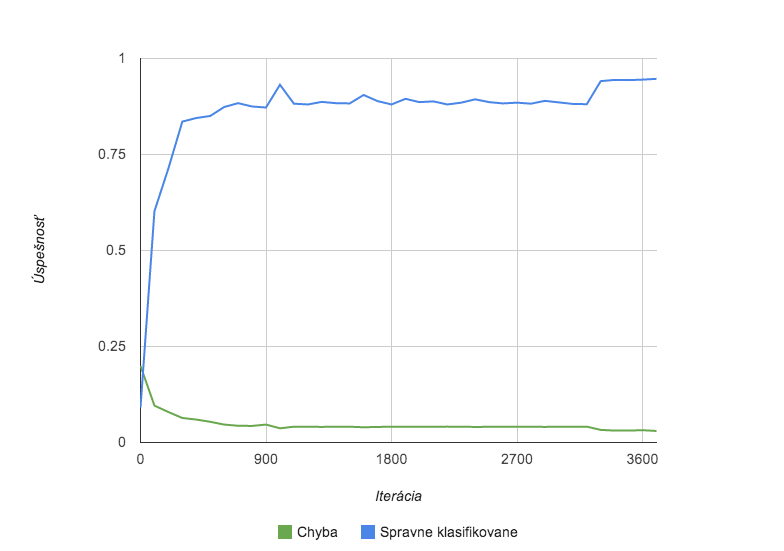
*Experiment s parametrami:*

learning rate 0,6; momentum 0,8 ; 12 neurónov v skrytej vrstve



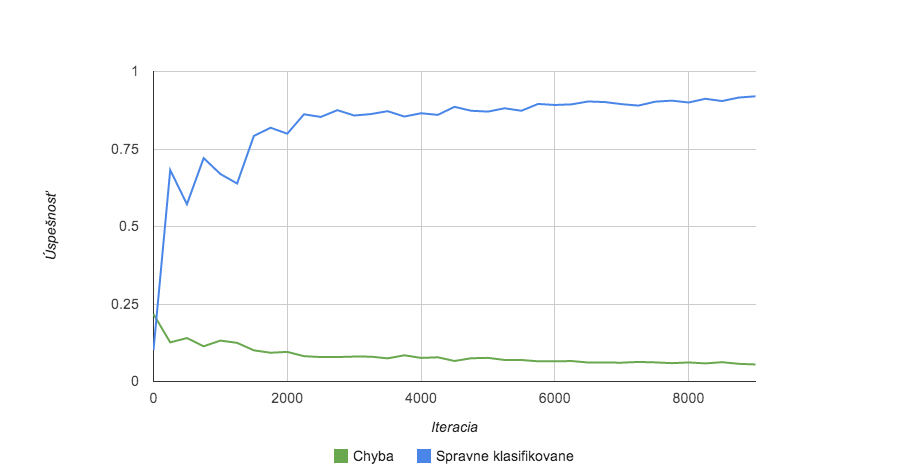
*Experiment s parametrami:*

learning rate 0,2; momentum 0,8 ; 12 neurónov v skrytej vrstve



*Experiment s parametrami:*

learning rate 0,3; momentum 0,8; 14 neurónov v skrytej vrstve



*Experiment s parametrami:*

learning rate 0,3; momentum 0,8; 16 neurónov v skrytej vrstve

# 

# Zhodnotenie

Implementácia neurónovej siete predstavovala netriviálny problém. Z pohľadu implementácie bolo najnáročnejšou fázou odladenie samotnej implementácie a zabezpečenie jej funkčnosti a správnosti výsledkov.

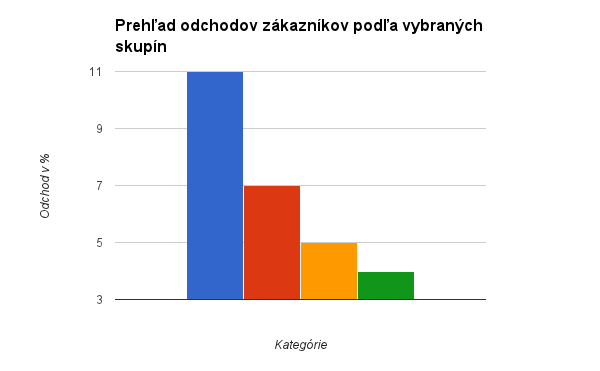
Na tieto účely slúžilo najmä implementovanie a otestovanie siete pre množinu XOR, kde sa dala manuálne overiť správnosť, rovnako bola sieť testovaná aj pre množinu dát – paritu, ktorá svojo komplexnosťou preverila funkčnosť a výpočtovú zložitosť – čas behu siete.

Po tejto časti práce prišlo na rad otestovanie vlastných dát. úloha, kt. bola postavená a ktorá mala z dát overiť správnosť hypotézy je podrobnejšie zdokumentovaná v časti testovanie.

Hypotéza, ktorá predpokladala na základe skúseností, že klienti, ktorí majú najväčšiu tendenciu končiť zmluvy (a prechádzať ku konkurencií) sú muži, vo veku 18-25 rokov, bývajúci v meste s dĺžkou kontraktu 2-5 rokov sa ukázala ako opodstatnená, čo dokumentuje nárast ukončenia zmlúv oproti iným skupinám.

Úspešnosť správnej identifikácie „odchodu“ na základe 12 atribútov sa pohybovala v rozmedzí 90-95 % čo možno hodnotiť ako spoľahlivé zatrieďovanie.

Nárast sa pohybuje medzi 3-5 percentami oproti najbližšej uvažovanej skupine:



|  |  |  |  |
| --- | --- | --- | --- |
| **M,M, V, 2-5** | **M,D, V, 2-5** | **Z,M, V, 2-5** | **Z,D, V, 2-5** |

Precision, recall