

D.P. Adorno, L.Bellomonte, M.Brai // – 2009. J. Stat. Mech. P03007.

12. Ткачук Р.А. Метод побудови оптимальної обробки електроретиносигналу та оцінювання вірогідності її результату / Р.А. Ткачук // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2008. – № 2. – С. 156 – 162.

13. Яворский Б.И. Взаимосвязь между представлением цифрового рекурсивного резонатора в частотной области и пространстве состояния / Б.И. Яворский // Автоматизация производственных процессов. – 1989. – № 28. – С. 151–155.

Надійшла 24.6.2012 р.

Рецензент: д.т.н. Лупенко С.А.

УДК 004.85

Ю.В. МОЖАРОВСЬКИЙ

Національний технічний університет України «Київський політехнічний інститут»

## ПОРІВНЯННЯ ШВИДКОДІ МЕТОДІВ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ ПРЯМОГО ПОШИРЕННЯ

*У статті проведено аналіз швидкості методів навчання нейронних мереж залежно від структури нейронної мережі та цільової задачі. Наведено рекомендації щодо застосування цих методів.*

*The analysis of neural network algorithms in dependence of its structure and problem is made. The usage recommendations of these methods are provided.*

Ключові слова: штучні нейронні мережі, навчання.

### Вступ

Нейронні мережі застосовуються для вирішення цілого ряду завдань, які є важко вирішуваними традиційними методами: розпізнавання образів, адаптивне управління, оптимізація, кластеризація, прогнозування. Вони ефективно впроваджуються у кібернетиці, космонавтиці, механіці, металургії і т.д.

Одним з найважливіших етапів застосування нейронної мережі є її навчання. Цей процес потребує набору пар прикладів «вхід – очікуваний вихід». Під час навчання ваги та зміщення мережі коригуються на кожній ітерації, щоб мінімізувати цільову функцію, найчастіше — середньоквадратичну помилку між виходом нейронної мережі та очікуваним виходом для навчальної множини.

Нехай  $e_i(n)$  – помилка між  $i$ -м виходом нейронної мережі та очікуваним виходом для  $n$ -го навчального елемента. Сумарна помилка для цього елемента виражається як

$$E(n) = \frac{1}{2} \sum_{i=1}^M e_i^2(n), \quad (1)$$

де  $M$  – розмір виходу нейронної мережі.

Середньоквадратична помилка визначається як

$$MSE = \frac{1}{N} \sum_{n=1}^N E(n), \quad (2)$$

де  $N$  – величина навчальної вибірки.

Підставивши (1) у (2), отримаємо

$$MSE = \frac{1}{2N} \sum_{n=1}^N \sum_{i=1}^M e_i^2(n). \quad (3)$$

Існує багато алгоритмів навчання, і дуже важко спрогнозувати, який з них працюватиме найшвидше. Це залежить від багатьох факторів, включаючи величину навчальної вибірки, кількість ваг і зміщень мережі, тип задачі та складність.

Проблема визначення найшвидшого методу не є новою [1, 2], але характерною для попередніх робіт є зосередженість на конкретній цільовій задачі. У даній статті порівнюються шість алгоритмів навчання мереж прямого поширення на чотирьох задачах, як простих вигаданих, так і реальних.

### Постановка задачі

Для проведення аналізу використовується прикладний пакет Matlab “Neural Network Toolbox”, комп'ютер з ОС Microsoft Windows XP SP3, AMD Athlon 64 X2 Dual Core Processor 4200+ (2.21 Ghz), 2 GB RAM.

В табл. 1 наведено список алгоритмів, їх позначення та складність алгоритму, де  $n$  — розмір входу нейронної мережі.

## Алгоритми навчання нейронних мереж

Позначення	Алгоритм	Обчислювальна складність
GDA	Адаптивний метод градієнтного спуску [3]	$O(n)$
RPROP	Resilient backpropagation [4]	$O(n)$
BFGS	BFGS алгоритм [5]	$O(n^2)$
LM	Алгоритм Левенберга-Марквардта [6]	$O(n^2)$
CGF	Fletcher-Reeves Update [7]	$O(n)$
CGP	Polak-Ribière Update [8]	$O(n)$

В табл. 2 наведено список задач з відповідною структурою мережі та цільовою помилкою.

Таблиця 2

## Задачі навчання

Задача	Структура мережі	Тип задачі	Цільова помилка
Sin	1-10-1	Апроксимація функції	$1e-5$
Parity	5-10-10-1	Класифікація	0.0001
Price	9-150-1	Апроксимація функції	0.005
AlphaDigit	320-150-36	Класифікація	0,05

## Навчальна множина Sin

Перша задача для аналізу роботи методів — проста апроксимація функції. Навчальна множина складається з 50 пар  $(x, \sin(x))$ , де  $x \in (-\pi, \pi)$ . Мережа має структуру 1-10-1 з функцією активації  $\tanh(x)$  у прихованому шарі та лінійною у вихідному.

В табл. 3 наведено дані з 50 експериментів для кожного методу. Початкові ваги і зміщення ініціалізувалися випадковим чином. Навчання припинялося при досягненні  $MSE = 0.001$ . Найшвидшим виявився алгоритм Левенберга-Марквардта, його середній час збіжності більш ніж вдесятеро менший наступного за швидкістю методу.

Таблиця 3

## Результати навчання Sin

Алгоритм	Мін. час, с	Макс. час, с	Середній час, с
GDA	0,79700	5,85900	2,95782
RPROP	0,23500	29,07800	3,36370
BFGS	0,18700	0,54700	0,34344
LM	0,01500	0,06200	0,02532
CGF	0,28100	1,31300	0,57382
CGP	0,37500	1,59400	0,75188

Те, який метод швидший, істотно залежить від цільового значення MSE. Саме це можна побачити на рис. 1, де зображено залежність середньоквадратичної помилки від часу сходження (середнього для 50 спроб) для декількох алгоритмів. Можна побачити, що алгоритм Левенберга-Марквардта збігається значно швидше з часом, ніж інші алгоритми.

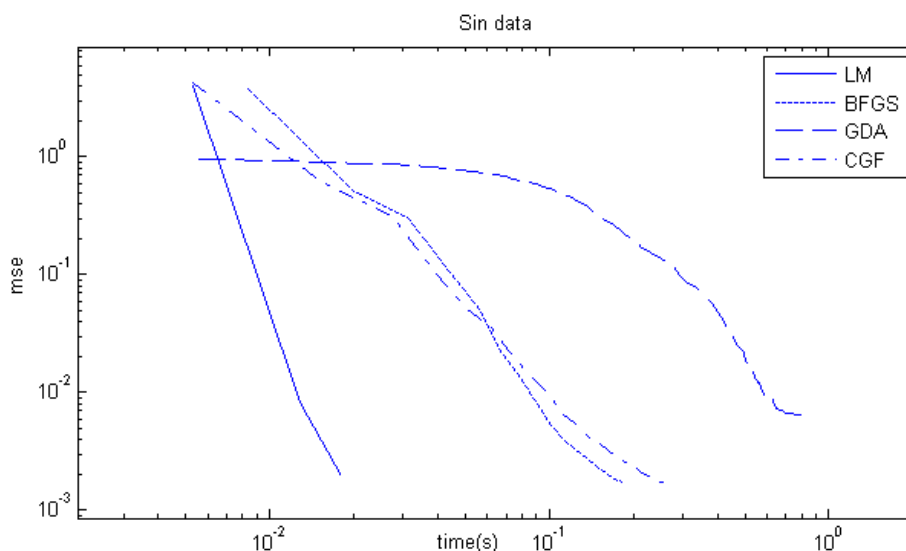


Рис. 1. Sin-залежність MSE від часу

Далі на рис. 2 зображено залежність середнього часу збіжності алгоритмів від цільового значення MSE. Можна побачити значне зменшення швидкості деяких методів при зменшенні MSE, особливо — RPROP та GDA. В той самий час LM та BFGS виявляються у лідерах.

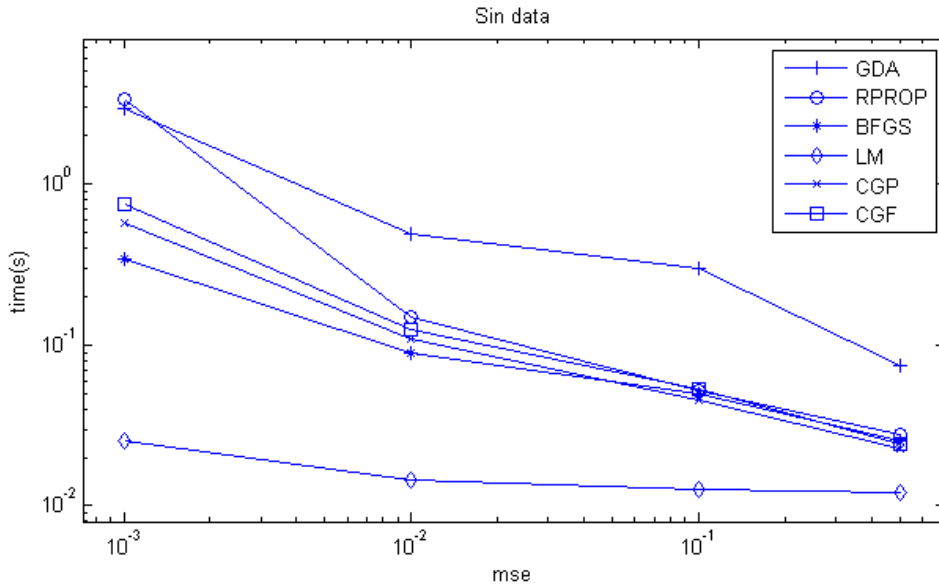


Рис. 2. Sin-залежність часу від MSE

**Навчальна множина Parity**

Наступна задача – класифікація бінарних векторів з п'яти елементів. Якщо вектор має парну кількість одиниць, тоді він парний і очікуваний вихід 1, у іншому випадку очікуваний вихід -1. Навчальна множина складається з 50 елементів. Мережа має структуру 5-10-10-1 з функцією активації  $\tanh(x)$  у всіх шарах.

В табл. 4 наведено дані з 50 експериментів для кожного методу. Початкові ваги і зміщення ініціалізувалися випадковим чином. Навчання припинялося при досягненні  $MSE = 0.0001$ . Як і минулого разу, найшвидшим виявився алгоритм Левенберга-Марквардта. Значно відстав метод BFGS.

Таблиця 4

**Результати навчання Parity**

Алгоритм	Мін. час, с	Макс. час, с	Середній час, с
GDA	1,18700	2,03100	1,40684
RPROP	0,31300	1,31300	0,57224
BFGS	0,67200	1,45300	0,93748
LM	0,07800	0,43800	0,13940
CGF	0,40600	3,21900	1,10472
CGP	0,42200	1,26600	0,70370

На рис. 3 зображено залежність середньоквадратичної помилки від часу збіжності (середнього для 50 спроб) для декількох алгоритмів.

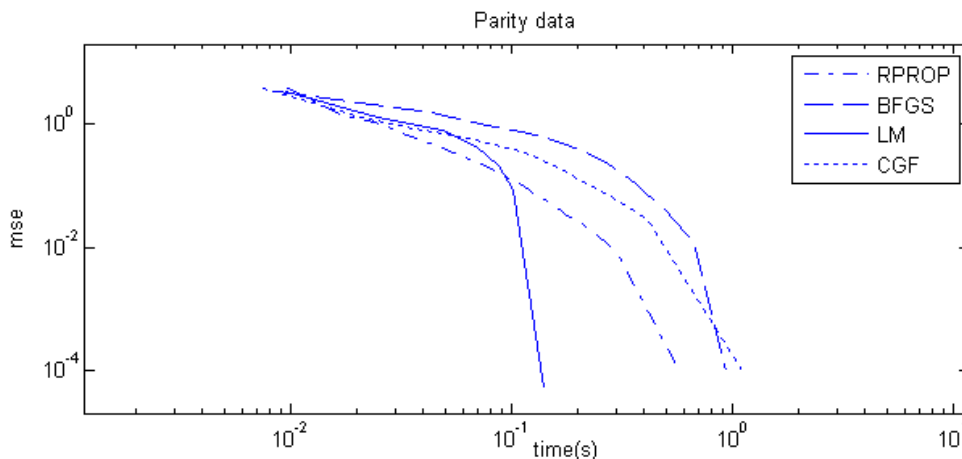


Рис. 3. Parity-залежність MSE від часу

На рис. 4 зображено залежність середнього часу збіжності алгоритмів від цільового значення MSE. Алгоритм Левенберга-Марквардта виявився найшвидшим.

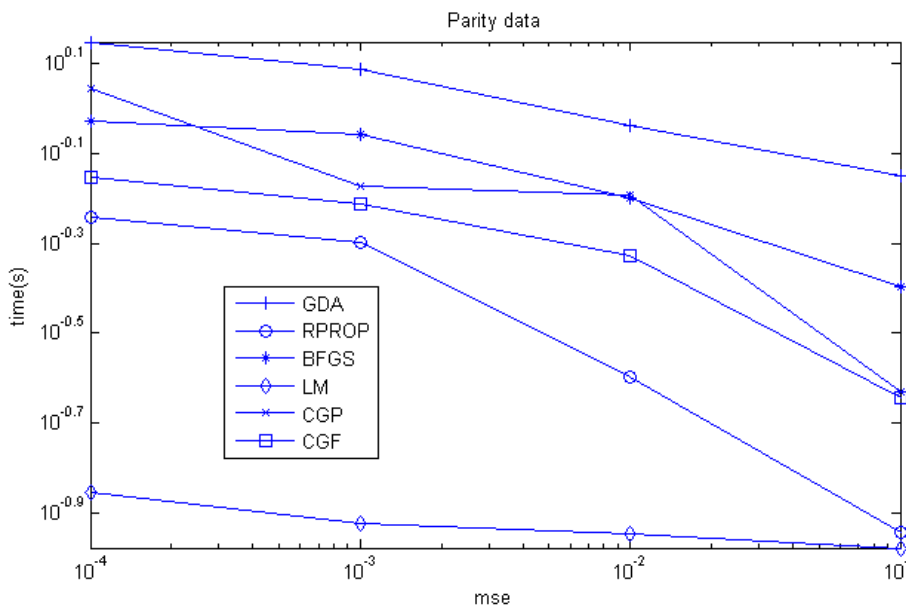


Рис. 4. Parity-залежність часу від MSE

**Навчальна множина Price**

Третя задача – апроксимація функції залежності курсу акцій компанії від курсу дев'яти інших компаній [9]. Навчальна множина складається з 950 елементів. Мережа має структуру 9-150-1 з функцією активації  $\tanh(x)$  у прихованому шарі та лінійною у вихідному.

В табл. 5 наведено дані з 30 експериментів для кожного методу. Початкові ваги і зміщення ініціалізувалися випадковим чином. Навчання припинялося при досягненні  $MSE = 0.005$ .

Таблиця 5

**Результати навчання Price**

Алгоритм	Мін. час, с	Макс. час, с	Середній час, с
GDA	169,1560	318,6570	248,7000
RPROP	26,7350	44,7030	35,7354
BFGS	386,4220	478,8600	428,0984
LM	12,2340	26,9530	16,6312
CGF	23,4850	34,7340	27,5661
CGP	24,3750	35,6090	28,6146

На рис. 5 зображено залежність середньоквадратичної помилки від часу збіжності для декількох алгоритмів. Алгоритм LM працює повільніше на початку навчання, CGF та CGP значно його випереджають. Але з часом ефективність LM зростає і він випереджає всі інші.

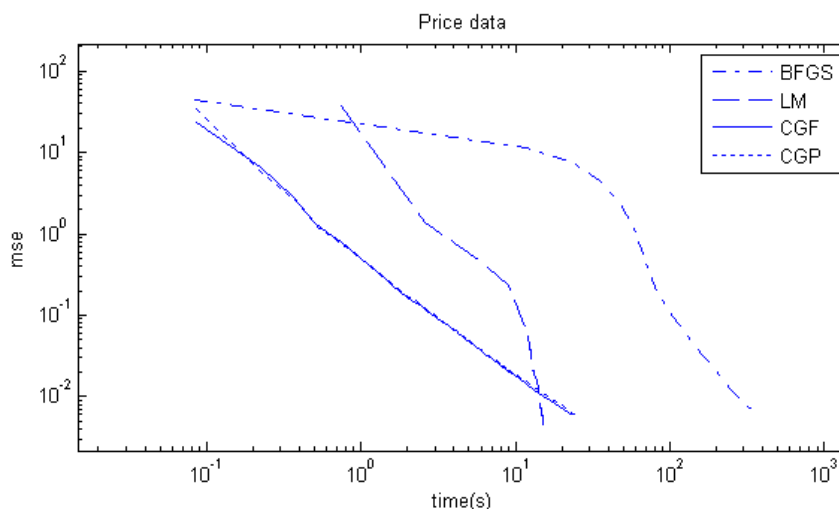


Рис. 5. Price-залежність MSE від часу

На рис. 6 зображено залежність середнього часу збіжності алгоритмів від цільового значення MSE. Можна відмітити загальну неефективність BFGS та сильне зменшення швидкості GDA при зменшенні цільового значення MSE.

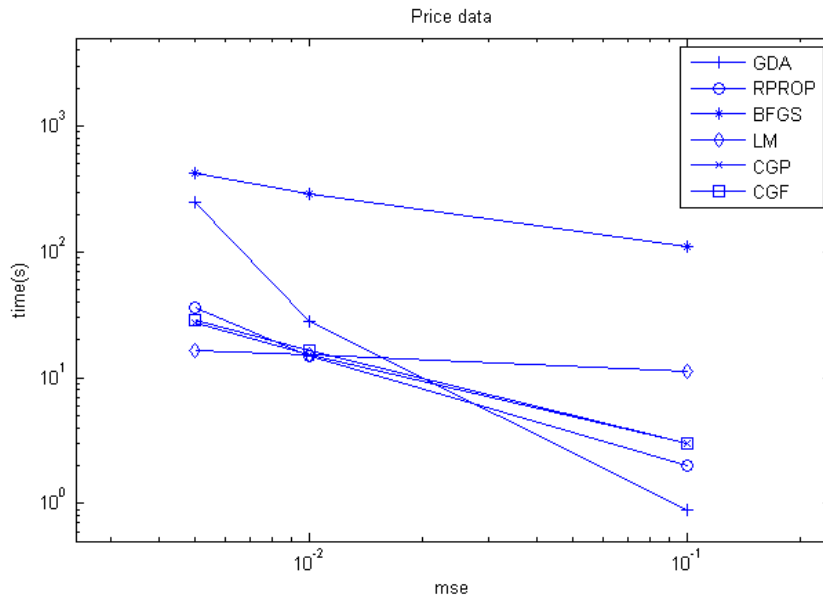


Рис. 6. Price-залежність часу від MSE

**Навчальна множина AlphaDigit**

Остання задача – навчити нейронну мережу розпізнавати рукописні літери та цифри. Навчальна множина [10] містить зображення рукописних цифр від 0 до 9 та великих латинських букв від "A" до "Z" розміром 20x16 пікселів. Для кожного символу представлено 39 зображень. Всього множина AlphaDigit містить 1404 зображення. Мережа має структуру 320-150-36 з функцією активації  $\tanh(x)$  у всіх шарах.

Зображення символів представлено матрицею зі значеннями від 0 (чорний колір) до 255 (білий колір). Для обробки мережею значення нормуються в інтервалі [-1;1] та формується вхідний вектор  $\vec{x}$  довжиною  $N_i \cdot N_j$ , де  $N_i$  – висота матриці,  $N_j$  – ширина. Вихідний вектор  $\vec{y}$  складається з K елементів, де K – кількість класів зображень. Приймаючи зображення символу на вході, будемо активувати відповідний елемент значенням 1, а всі інші елементи – значенням -1. Схематично процес зображено на рис. 7.

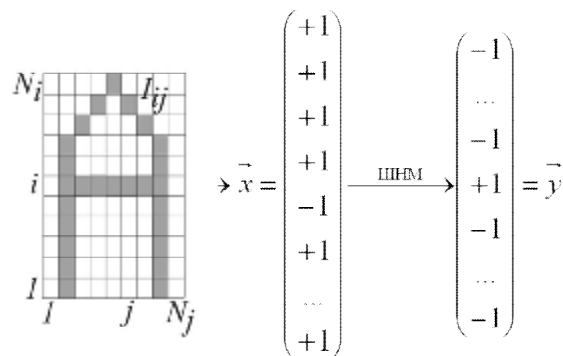


Рис. 7. Передача зображення символу нейронній мережі

В табл. 6 та рис. 8 наведено дані з 30 експериментів для кожного методу. Початкові ваги і зміщення ініціалізувалися випадковим чином. Навчання припинялося при досягненні  $MSE = 0.05$ . Алгоритми BFGS та LM відносяться до ньютонівських алгоритмів та потребують значних ресурсів ОЗП комп'ютера. Мережа і навчальна вибірка є досить великими, тому під час експерименту виникла проблема нестачі пам'яті. Робота BFGS та LM не є можливою для даної задачі.

В цьому випадку найефективнішим за часом збіжності є алгоритм RPROP.

Таблиця 6

**Результати навчання AlphaDigit**

Алгоритм	Мін. час, с	Макс. час, с	Середній час, с
GDA	248.718000	470.625000	320.807933
RPROP	31.734000	34.172000	32.871833
BFGS	-	-	-
LM	-	-	-
CGF	53.954000	150.188000	73.189100
CGP	58.890000	136.953000	78.233800

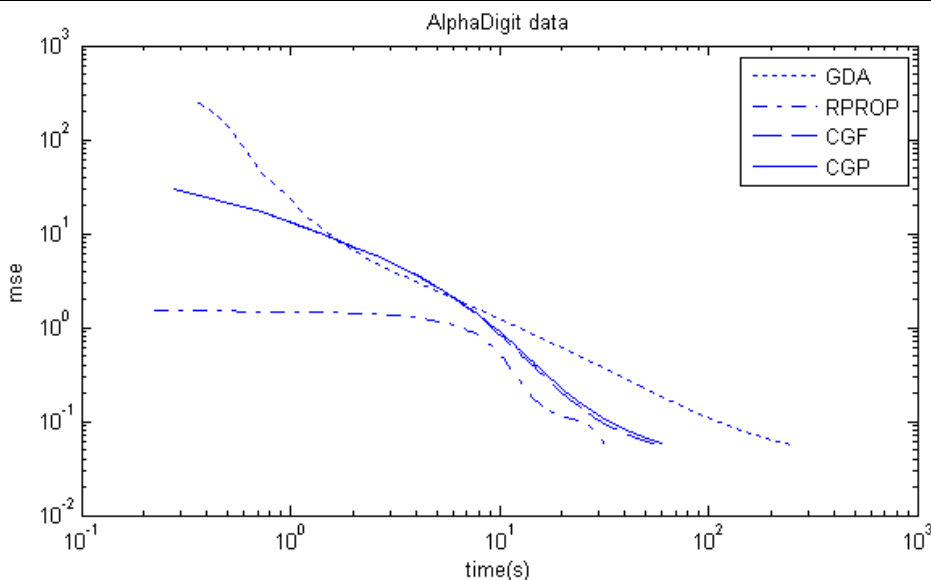


Рис. 8. AlphaDigit-залежність MSE від часу

### Висновки

За результатами тестів можна зробити висновок, що для мереж, які мають незначну кількість ваг, найефективнішим за часом є алгоритм Левенберга-Марквардта, який виявився значно швидшим за інші алгоритми. Особливо це помітно при зменшенні цільової середньоквадратичної помилки. При збільшенні структури мережі алгоритм програє за швидкістю, але при малих цільових помилках у підсумку все ж таки збігається швидше інших.

Алгоритми Левенберга-Марквардта та BFGS мають значні потреби у ресурсах ОЗП комп'ютера, тому використання цих методів виявилось не можливим в задачі 4 через величину мережі та вибірки.

Також з графіків видно, що алгоритм RPROP є найефективнішим у навчанні мереж великої структури. Він не потребує значних ресурсів ОЗП, є легким у реалізації.

### Література

1. Comparison of Neural Network Learning Algorithms for prediction Enhancement of a Planning Tool [Електронний ресурс]. – Режим доступу : <http://projets.comelec.telecom-paristech.fr/EW2007/papers/1569014736.pdf>
2. A Comparison of Artificial Neural Network Learning Algorithms for Vibration-Based Damage Detection [Електронний ресурс] — Режим доступу: <http://www.scientific.net/AMR.163-167.2756>
3. Variable Learning Rate [Електронний ресурс]. – Режим доступу : [http://www.kxcad.net/cae\\_MATLAB/toolbox/nnet/backpr12.html](http://www.kxcad.net/cae_MATLAB/toolbox/nnet/backpr12.html)
4. A direct adaptive method for faster backpropagation learning: The RPROP algorithm [Електронний ресурс]. □ Режим доступу : <http://paginas.fe.up.pt/~ee02162/dissertacao/RPROP%20paper.pdf>
5. Станислав Осовский. Нейронные сети для обработки информации / Станислав Осовский. □ М. : Финансы и статистика, 2002. – С. 63–64. – ISBN 5-729-02567-4.
6. Станислав Осовский. Нейронные сети для обработки информации / Станислав Осовский. □ М. : Финансы и статистика, 2002. – С. 65–67. – ISBN 5-729-02567-4.
7. Fletcher-Reeves Update [Електронний ресурс]. – Режим доступу : [http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher\\_reeves.html](http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher_reeves.html)
8. Polak-Ribière Update [Електронний ресурс]. – Режим доступу : [http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher\\_reeves\\_polak\\_ribiere.html](http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher_reeves_polak_ribiere.html)
9. Stock Prices [Електронний ресурс]. – Режим доступу : <http://funapp.cs.bilkent.edu.tr/DataSets/Data/SP.dat>
10. Binary 20x16 digits of "0" through "9" and capital "A" through "Z" [Електронний ресурс]. – Режим доступу: <http://cs.nyu.edu/~roweis/data/binaryalphadigs.mat>

Надійшла 5.6.2012 р.  
Рецензент: д.т.н. Данилов В.Я.