Winter 2017   COSC 4P80: Assignment 1

Instructor: B. Ombuki-Berman
Assigned: Monday, January 30th, 2017

Due date: Tuesday, February 28th, 12:00 pm (no lates)

Task: **Handwritten Digit Recognition using Back-propagation and Variants**

The objective of this assignment is to learn about the differences/similarities in supervised learning rules and their effect on the learning process in feed forward neural architectures. The learning methods are tested on a popular digit recognition database from the US Postal Service. In order to accomplish this you will be investigating three popular learning rules as they are applied to a practical application. You are to explain the observed results concerning learning parameters (e.g., learning rate) and generalization ability as well as the role of the neural architecture (such as activation function and hidden layer(s) size).

Implementation

You are required to construct feed forward neural networks that are able to classify handwritten digits. The data has been collected from the United States Postal Office and consists of 8 x 8 pixel images organized into a file per digit from 0-9. In total there are 700 training cases and 400 test cases per digit. Note: before using this data it is advisable to test your programs on other simpler data (Don't have to hand in).

Part A
Build a feed forward neural network based on backpropagation learning with momentum. Vary the learning rate and momentum, recording network performance. For each of these also vary the number of hidden nodes. Alter the activation function between the logistic and tanh functions. What conclusions can you make based on these findings? Split the training data into training and validation sets. Use the holdout and cross validation techniques and examine how they improve generalization capability. What conclusions can be made? How is the final output affected (when using the testing set)? Ensure that you are evaluating the results over enough runs so that they are statistically significant. Use statistical methods to back up your explanation (See hints below).

Part B
Repeat (A) using any two variants of backpropagation such as rProp, quickProp, cascade correlation, delta-bar-delta or conjugate gradient and Levenberg-Marquardt (some of these will be discussed in class)

Assignment Requirements and Grading:

The results are to be handed in via a technical paper using a IEEE style file format with the following headings: (including an abstract)

1) Abstract, Introduction & problem definition
   - What is the goal of this work? Outline of rest of paper, Etc
   - Clear definition of the problem and data being used, applicability of neural networks for this problem

2) Neural Networks (with subsection for each of the learning rules)

   - Paragraph describing feed forward networks, and another describing learning rules in general.
   - Section for each of the chosen learning rules. Each section should explain weight update, as well as how learning is accomplished.
   - What is similar/different between the networks? A small pseudocode algorithm may be useful. After reading this section the reader should be clear as to how the learning rules work. Explain generalization techniques, problems, etc here as well.

3) Methods of Analysis
   - What are you going to analyze and how?
     This part will probably have a separate subsection for each of the analytical methods used. Each subsection should be about 2-3 paragraphs (in most cases), explaining exactly how the method is used (try not to say the words neural network in this section – ie. Just explain the method, worry about the neural net in the next section). These subsections are therefore going to be explaining statistical stuff (no need to explain average, std. deviation, and other straight forwards). Choose whatever you like to make your point clear, possible things include:
       i. Comparison of means
       ii. t-tests/tests for normality/Mann-Whitney U tests
       iii. Confidence intervals
       iv. ANOVA/Kruskall-Wallis
       v. Contingency table analysis
       vi. ???
     There is no specific statistical test (s); you are free to choose those tests you feel are suitable to analyze the networks.
4) Results and Discussion
   - Here you will outline the parameters used and show plots, etc of your results.
   - Also, explain (using the methods of analysis) what the results lead you to conclude.
   - This should be the biggest section (followed by with (3) and (2), respectively)
   - Ensure that the effect of the learning rule, number of hidden nodes, etc… are all tested
5) Conclusions
   - In 2-4 paragraphs clearly outline your conclusions and why they are valid.
   - So what is the effect of the learning rule? Architecture, i.e., number of hidden nodes etc? Etc…

6) References – include the references you used for work, including book (s) and research articles

You must use the IEEE style found at
http://www.ieee.org/conferences_events/conferences/publishing/templates.html
The report must be written in a clear and concise manner.

* It is expected that the report will be approximately 10-12 pages (in my estimation, a report within this range is solid and sufficient), this is just a guideline, no penalization for extra pages.
** Failure to submit working code will result in a very low mark if not 0 (even if other sections of the report are completed correctly).
*** Academic misconduct according to the departmental regulation (plagiarism, cheating, false results, etc) is NOT tolerated, and will result in a 0 for this class.

NOTE:

MINIMUM MARKS:
- Simple testing using only averages and maybe standard deviation. Not mathematically clear that one network outperforms others (if one does), insufficient number of runs, etc.
- redundant methods of analysis.
- Write-up is not in technical format, has spelling errors, incorrect explanations, etc.
- No plots are shown in write-up
- No references, or not properly referenced
- Conclusion not well supported by results, or methods of analysis are unconvincing.

MAXIMUM MARKS
- Uses reasonable statistics, ANOVA, hypothesis testing, confidence intervals, contingency table analysis to determine quality of results, etc…
- Report is very clear, concise, well-written, correct explanations, etc…
- Shows plots (but not too many - enough to get the point across), other graphs, charts, etc
- Conclusions are well supported from methods of analysis.

Note: Do not overcomplicate the algorithms! They are straightforward and not hundreds of lines in length…if you have very large programs you are probably overcomplicating it! You should do statistical analysis in Excel, R, Maple (or even Minitab, MATLAB if you are familiar with them etc…). Most of this assignment is in the analysis of results, and understanding of learning rules, so concentrate on them, and start early!!