

COMP3002 Artificial Intelligence
Semester 1, 2003

Assignment 2. Backpropagation Neural Network
Due in week 11 (at the beginning of your tutorial class); can be done in pairs.

The goal of this assignment is to implement backpropagation neural network, evaluate its performance using 10-fold cross validation, and also to investigate the effect of the number of hidden neurons.

1. Download a dataset from the UCI Machine Learning Repository

<http://www.ics.uci.edu/~mllearn/MLRepository.html>.

It should contain more than 200 examples described with only numerical features. If there are missing values, choose a method to handle them. The following datasets are suitable: breast-cancer-w, glass, liver-disorders, pima-diabetes, new-thyroid, vowel.

2. Normalize data between 0 and 1.

3. Implement backpropagation neural network with one hidden layer. It should be fully connected (i.e. each neuron is connected with all neurons from the previous layer). Use one output neuron for each class and binary encoding of the target outputs. Implement the learning rule without momentum. When calculating the accuracy, an example is considered correctly classified if the respective output neuron (corresponding to the target class) has highest activation.

4. To investigate the effect of the number of hidden neurons on the performance, select 5 different numbers of hidden neurons. For each of these experiments evaluate the performance (classification accuracy and speed) using 10-fold cross validation. (You are not required to use stratification).

Write a brief report:

1. Description of data:

- data set used; number of attributes and classes
- preprocessing applied - e.g. were there missing values and how they were handled; how was the normalization done, anything else specific to the chosen data set

2. Experimental setting

- NN architecture – how many input, hidden and output neurons were used
- other parameters – e.g. learning rate, range of the initial weights
- stopping criteria – what was the stopping criteria; parameters used

3. Results – for each number of hidden neurons, fill in the following table

# fold	accuracy on training set [%]	accuracy on test set [%]	# training epochs
1			
2			
...			
10			
average	mean \pm st.dev.	mean \pm st.dev.	mean \pm st.dev.

where mean and st. dev. are the mean value and standard deviation (respectively) when the results are averaged over the 10 folds

4. Discussion – briefly discuss the choice of the parameters and results (e.g. training vs. testing set accuracy, variations of the results across the folds, effect of the number of hidden neurons on the accuracy and training time, and anything else you consider important).

Enclose a print out of your code.

Submission: both hard copy and electronic version (using netfile or as an e-mail attachment) to your tutor.

Deadline: week 11, at your tutorial class

Marking scheme (guidelines only):

1. Description of the data – 2 marks
2. Experimental setup – 2 marks
3. Implementation – 5 marks
4. Results – 2 marks
5. Discussion – 3 marks