Machine Learning Assignment 1

Literature review:

# Introduction

Content Describe in your own words what the paper is about and what the main challenge or problem it attempted to resolve.

Without hidden layers in a neural network, it is unable to learn certain mappings between similar inputs and different outputs using similarity of patterns.

However, if add internal representation units to augment the original input pattern, the network can perform any mapping from input to output.

Since, network with hidden layers units cannot use simple rule, such as “delta rule”, for all problem, 3 response is proposed:

1. Competitive Learning: hidden units develop by simple unsupervised learning rules.
   1. The disadvantage of the response is that there is no guarantee that hidden units appropriate for the required mapping are developed.
2. assume an internal representation:
   1. appropriate for verb learning and word perception
3. develop a **learning procedure** which is adjustable for task variations.
   1. Boltzmann machines:
      1. Uses stochastic units
      2. Reach equilibrium in two different phases
      3. limited to symmetric networks
   2. stochastic units by Barto (1985)
   3. **generalized delta rule.(used by this paper)**
      1. deterministic units
      2. involves only local computations
      3. a clear generalization of the delta rule
   4. learning-logic (Parker (1985))
      1. a similar generalization with “generalized delta rule”
   5. Le Cun (1985) has also studied a roughly similar learning scheme.

# Innovation

How innovative is the research work in the paper? Here you should describe the “novelty” of the paper. What does the paper contribute: a new method? an algorithm? a methodology? a comparison between methods? or something else?

**THE GENERALIZED DELTA RULE**steps:

1. uses the input vector to produce its own output vector
2. compares this with the desired output, or target vector.
3. Difference is reduce by change weights

**The delta rule and gradient descent:**

For , j mean the jth perceptron in the output layer, i means the ith perceptron in the in layer. This formula means the change of the weights is

This formula shows how weights will change by measuring difference between target output and actual output. is the learning rate. This can be derived by taking the partial derivative of Error (defined by ) with respect to ( ).

This function means the “error signal” of an perceptron is calculated by its firing strength’s derivative multiplies the sum weighted “error signal” of its connected upper perceptron. If the perceptron is output perceptron, the last term is . Error signal means the derivative of error with respect to net input, defined by the formula

**The delta rule for semilinear activation functions in feedforward networks:**

By adding hidden units w may converging at local minima.

The activation function is defined as , its derivative is calculated: .

**Symmetry breaking:**

If weights are initiated equally, the error signal could be the same, because it is calculated by weight multiplies output error. Then the change for all neuron are the same, which again results in same weights.

To solve it local maximum risk, small random weights is initiated.

# Technical quality

How would you rate the technical quality of the work in the paper? Here you should consider the quality of the work done. For example, a paper comparing classifiers on the basis of their accuracy on a training set would be of poor quality because generally the accuracy measured on a training set is higher than the actual test error. Other indicators of poor quality might be that the results in the paper could not be replicated by someone reading the paper (because they were not described clearly enough); a comparison was made between two things but the two things couldn’t fairly be compared; or conclusions were drawn from too few experiments.

# Application and X-factor

Do you think the application domain is appropriate for the proposed technique? What other application domains could the research work be applied? Also in this section, give a couple of suggestions for further developments of the research work. Do you think the work described in the paper could spark a good discussion in class? What did you find interesting about the work?

Page 10 : Error for different pattern presentations can be cancelled? (one negative and one positive)

Page 11: cannot understand formula (7)

# Presentation

How would you rate the quality of the presentation of the paper? Things you could consider here include: how easy it was to follow the argument in the paper; presentation style; depth of the argument; and clarity of the presentation.

# References

To answer the questions above in your report you will need to read other books or papers. List any references to other papers and books you read here. Referencing should be in the Harvard format. See the information at the UTS Library http://www.lib.uts.edu.au/help/referencing.

The size of the report should suit the need to clearly express your ideas. The following word counts can be considered

as a typical volume for each section:

Content: 300

Innovation: 300

Technical Quality: 200

Application and X-Factor: 200

Presentation: 100