Backpropagation Can't Do Parity Generalisation

Chris Thornton

Cognitive and Computing Sciences

University of Sussex

Brighton

BN1 9QN

UK

Email: Chris.Thornton@cogs.susx.ac.uk

WWW: http://www.cogs.susx.ac.uk

Tel: (44)1273 678856

November 25, 1996

Abstract

It is accepted that the early connectionist learning methods such as the perceptron algorithm cannot solve parity learning problems. But since the early 1980s, there have been many demonstrations purporting to show that the backpropagation method *can* do so. However these demonstrations are misleading. Backpropagation in fact reliably *fails* to solve parity

Keyworks: Learning, Theoretical limitations

1 Backpropagation performance on parity generalisation

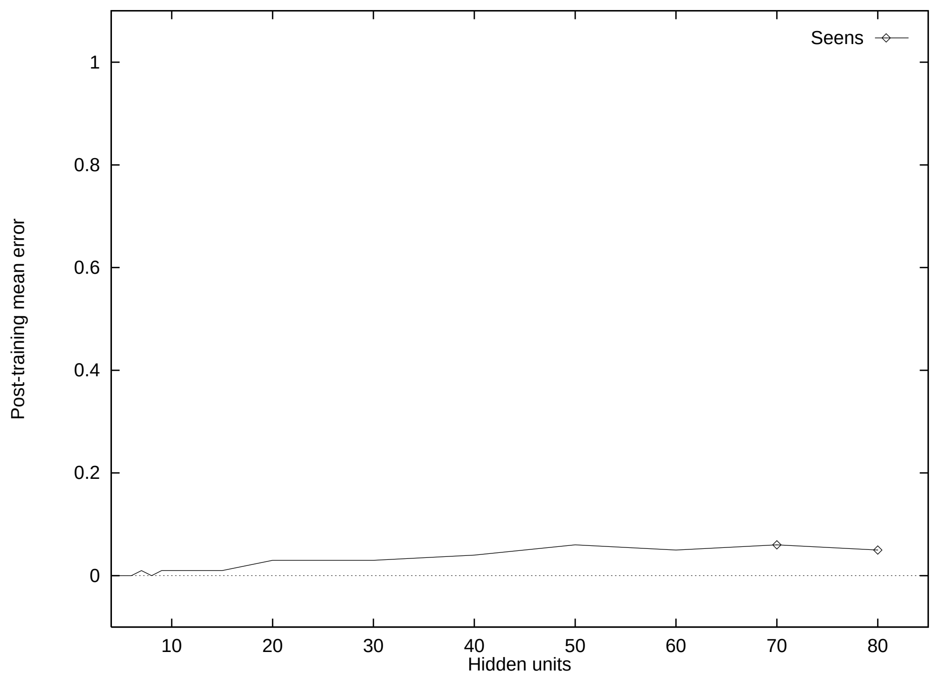
ing sets which contain all but one of the possible cases) can be constructed. C4.5 actually generalises incorrectly in *all 16 cases*; i.e., it always 'gets the answer wrong'.

Backpropagation performs no better. In an extensive empirical analysis, backpropagation was tested for its ability to generalise to one, randomly selected unseen case in the 4-bit parity mapping. In this analysis a standard, two-layer, (strictly) feed-forward network was used with the number of hidden units being varied between 3 and 80. Data were collected for 20 successful runs (i.e., achievement of negligible error on the training data) with each architecture. The learning rate was 0.2 and the momentum value was 0.9.

The results are summarised in Figure 1. This shows the post-training mean error for seens and unseens averaged over the 20 successful training runs which were performed in each architecture. The basic error value used here is simply the average difference between the target output and actual output produced. The graph shows negligible mean error for seen cases due to the fact that data were only recorded for successful runs. More interestingly, it shows that the mean error on the unseen case is very poor for all architectures used, i.e., no generalisation is achieved. (The reason why the generalisation error is so much *worse* than chance is explained below.)

2 Performance on related problems

Generalisation failures on parity mappings are sometimes dismissed on the grounds that the parity problem is an artificial construct upon which learning methods cannot be expected to perform properly. To show that this is not the case we need to demonstrate that backpropagation fails on other problems as well as parity. The key property of the parity mapping is that, in its full form, it is statistically neutral. That is to say, the conditional probability of


Seens
1
0.8
0.6
0.4
0.2
0
10
20
30
40
50
60
70
80
Post-training mean error
Hidden units

3 Dealing with neutrality in non-parity problems

Consider the 'likelihood problem' whose target mapping is shown below. This is a straightforward learning problem with a relatively obvious input/output rule. However, we can translate it into a modulus-addition problem with the following substitutions: person/0, computer/1, consumes/0, dislikes/1, heat/0, electricity/1, moisture/2, silicon/3, yes/0, no/1. Under this translation, the requirement that the input-set cardinalities are equal to, or a multiple of the output-set cardinality is met so we know that the problem is *necessarily* statistically neutral. It is in fact easy to confirm that every single conditional output-probability has the chance value, which is 0.5 here because there are just two output values.

x_1	x_2	x_3		y_1
person	consumes	heat	$\Longrightarrow$	yes
person	consumes	electricity	$\Longrightarrow$	no
person	consumes	moisture	$\Longrightarrow$	yes
person	consumes	silicon	$\Longrightarrow$	no
person	dislikes	heat	$\Longrightarrow$	no
person	dislikes	electricity	$\Longrightarrow$	yes
person	dislikes	moisture	$\Longrightarrow$	no
person	dislikes	silicon	$\Longrightarrow$	yes
computer	consumes	heat	$\Longrightarrow$	no
computer	consumes	electricity	$\Longrightarrow$	yes
computer	consumes	moisture	$\Longrightarrow$	no
computer	consumes	silicon	$\Longrightarrow$	yes
computer	dislikes	heat	$\Longrightarrow$	yes
computer	dislikes	electricity	$\Longrightarrow$	no
computer	dislikes	moisture	$\Longrightarrow$	yes
computer	dislikes	silicon	$\Longrightarrow$	no

The neutrality of the likelihood problem implies that we should expect generalisation performance on this problem by backpropagation (and C4.5) to be just as poor as it was in the case of parity. And in fact this is exactly what we *do* find. Backpropagation's generalisation performance on the likelihood problem using one, randomly selected unseen case is summarised in the Figure 2. Again, the generalisation performance is very poor in all architectures tested. As expected, C4.5 generalises incorrectly on all 16, minimally incomplete training sets for this problem.

It is interesting to note that the generalisation performance on this new problem hovers around the chance level. This is of course where we would expect

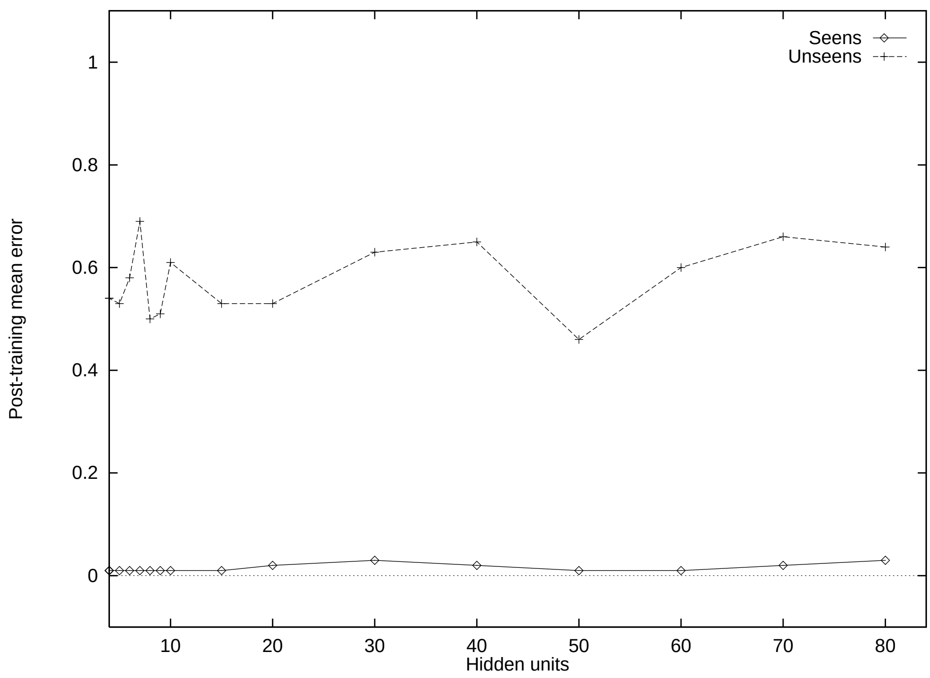


Figure 2: Post-training mean-error curves for likelihood generalisation.

it to be on the hypothesis that backpropagation cannot deal with statistically netural mappings. In the case of the true parity problem, it will be recalled that the generalisation was considerably *worse* than chance. The explanation for this appears to be that in deleting a single case from a parity mapping, a strong but misleading association is created between input cases one Hamming unit away from the deleted case and the complement of the output for those cases (i.e., the 'wrong' output). Backpropagation detects and exploits this phoney association and is thus led to always produce the complement of the correct generalisation.

4 Concluding comment

The paper has shown that backpropagation reliable fails to solve parity learning problems when they are posed as genuine supervised learning problems. The algorithm is thus subject to at least one of the limitations that Minsky and Papert attributed to the perceptron method in the late 1960s. The firm confidence which researchers sometimes place in backpropagation may therefore be less than fully justified.

References

[1] Minsky, M. and Papert, S. (1988). *Perceptrons: An Introduction to Computational Geometry* (expanded edn). Cambridge, Mass.: MIT Press.

[2] Quinlan, J. (1993). *C4.5: Programs for Machine Learning.* San Mateo, California: Morgan Kaufmann.

[3] Rumelhart, D., Hinton, G. and Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, *323* (pp. 533-6).