Chapter 4

CONCLUSION

The first objective of this thesis was to propose a new active learning algorithm using changes in output as selection criterion. The second objective of this thesis was to compare the performance of selected active learning algorithms on both clean and noisy data. The algorithms were compared in relation to their accuracy, generalization, convergence and computational cost. Four active learning algorithms were selected for this comparison. Two of these algorithms namely, DPS and AL use the error in prediction as selection criterion. That is, patterns are selected based on the error of the patterns. Two algorithms, SLA and SAILA, which use perturbations in output as selection criterion were also selected for comparison. Patterns which influence the change in output values most are selected for training, using the output selection criterion.

Röbel's algorithm (DPS) performed well with clean data (TS1, TS2 and F1), having a faster convergence and better generalization than the other algorithms. This performance can be attributed to the selection of patterns that contribute most to the error of the network. Training on such patterns took into account the current state of the network and thus brought the output closer to the target function. The performance of DPS degraded in the presence of outliers and noise in the training data, consequently the generalization ability deteriorated.

AL performed badly in all the functions except F1 and TS2. For the sine functions (TS1, TS2, TS3), AL selected very few patterns for training which resulted in very large errors.

SAILA achieved a considerable good accuracy for the function with outliers and the complex function TS3. This is because SAILA used perturbations in output values, i.e. changes caused to the output by the input as its selection criterion, thus avoiding the selection of outliers patterns. SAILA was however, slow in learning most of the functions, even for those functions for which a low generalization error has been obtained. SAILA's slow learning *can* be attributed to the fact that SAILA only chose patterns at the highest peak of the derivative and then tries to fit the network from this point. A suggestion to improve training using the SAILA algorithm is to select patterns at the lowest peak also, i.e at the turning point where derivative is zero in addition to the patterns selected at the highest peak. The network will then fit the problems being solved at the two extreme points of the derivatives simultaneously. A faster convergence and a lower training time maybe achieved compared to the current SAILA algorithm.

SLA achieved a good accuracy for both clean data and data with outliers and noise. SLA used much less patterns (i.e. a low computational cost) than all the algorithms for all the problems. Thus, SLA showed to be more robust in the occurrence of outliers and noise. SLA has demonstrated good and comparable results both in the training and generalization ability of the network.

Active learning algorithms using perturbations in output performed better with functions with noise and outliers while, algorithms using change in error as selection criteria performed better with clean data. A good subset selection criterion is very important in any active learning algorithm. AL had a poor subset selection criterion, selecting too few patterns for training. Even though, DPS and AL used the same selection criterion, DPS outperformed AL in all the problems. This better performance of DPS is a result of a better subset selection criterion used by Röbel. A network trained with too few information will generalize badly, as in the case of AL.

For problems with clean data, DPS is preferred, though DPS used more patterns for training than SLA. However SLA is preferred with problems with outliers and noise. SLA is also preferred for clean data because of low computational costs.

4.1 Future of Active Learning in Neural Networks

Active learning has been shown to demonstrate a better performance than the conventional backpropagation algorithm. Various research have compared these two learning paradigms and have published their results [Zhang 1994, Röbel 1994c, Engelbrecht *et al* 1998]. Because of the demonstrated performance of active learning, research to improve on active learning must be continuously carried out.

A suggestion to further improve on active learning is to first cluster input patterns. A clustering algorithm can be used to group similar patterns into clusters, where similarity is measured as the Euclidean distance between input vectors. At each subsetselection interval, the most informative pattern is selected from each of the clusters. The clustering active learning approach can potentially reduce computational complexity and improve accuracy.

-