Appendix A

Symbols and Notations

The notation and symbols used in this thesis assume a three layer neural network (NN) architechure with one input layer, one hidden layer, and one output layer. This appendix summarizes the symbols used throughout this thesis with reference to the three layer architecture. The symbols are listed alphabetically with their interpretation.

Symbols	*Interpretation*
α	momentum term
η	learning rate
κ	Zhang's notation for a specified performance level
ρ	Röbel's generalization factor
τ	Zhang's notation for allowable error tolerance per connection
Φ	Engelbrecht's notation for pattern informativeness
Ψ	used in this thesis as the next-time-change in output
Υ	used as the next-time-change in target for Kohara's algorithm

Symbols	*Interpretation*
C_{al}	cost of the subsetselection criteria for Zhang's accelerated learning (AL)
C_{dps}	cost of the subsetselection criteria for Röbel's dynamic pattern selection (DPS)
C_{fc}	cost of training a NN on a training set
C_{fsl}	cost of fixed set learning
C_s	cost of selecting patterns
C_{sai}	cost of the subsetselection criteria for SAILA
C_{sla}	cost of the subselection criteria for selective learning algorithm (SLA)
$C_{S_{al}}$	cost of selecting a pattern into D_T for Zhang's accelerated learning
$C_{S_{dps}}$	cost of selecting a pattern into D_T for Röbel's dynamic pattern selection
$C_{S_{sai}}$	cost of selecting a pattern into D_T for SAILA
C_{AL}	cost of training a network using Zhang's algorithm
C_{DPS}	cost of training a network using Röbel's algorithm
C_{SA}	cost of training a network using Engelbrecht's algorithm
C_{SL}	cost of training a network using selective learning algorithm
C_v	cost of updating a weight between the input and hidden layers
C_V	cost of updating all weights between the input and hidden layers
C_w	cost of updating a weight between the hidden and output layers
C_W	cost of updating all weights between the hidden and output layers
D_C	set of candidate training patterns
D_G	test set or the generalization set
D_T	actual training set
D_V	validation set
I	total number of input units
J	total number of hidden units
K	total number of output units
N_V	total number of weights between input and hidden layers of a network
N_W	total number of weights between hidden and output layers of a network
p	a single pattern
P_C	number of patterns in the candidate set D_C
P_G	number of patterns in the generalization set D_G
P_s	number of patterns in a subset D_s
P_T	number of patterns in the training set D_T
P_V	number of patterns in the validation set D_V
o_k	k-th output unit
$o_k^{(p)}$	activation of output unit o_k for pattern p
v_{ji}	weight between j-th hidden unit and i-th input unit
w_{kj}	weight between k-th output unit and j-th hidden unit
y_j	j-th hidden unit
$y_j^{(p)}$	activation of hidden unit y_j for pattern p
z_i	i-th input unit

Appendix B

Definitions

This appendix summarizes definitions of key terms used in this thesis. The terms are defined in alphabetical order.

Active learning: Active learning is any form of learning in which the learning algorithm has some deterministic control during training over what part of the input space it receives information (page 10).

Accelerated Learning: Accelerated learning (AL) is Zhang's algorithm for active learning. Patterns with the highest prediction error are selected for training. New patterns are selected as soon as the error on the training subset is reduced to a specified performance level. AL is an incremental approach to active learning (page 61).

Bias: A bias is a unit or neuron added to the input and hidden layers with a constant activation value of -1. The purpose of adding a bias unit is to offset the origin of the logistic activation function (page 36).

Dynamic Pattern Selection: Dynamic pattern selection (DPS) is Röbel's algorithm for active learning. The most informative patterns are the patterns with the maximum prediction error and are selected for training. New set of patterns are selected for training as soon as the network overfits the current training subset. DPS is an example of incremental learning (page 59).

Epoch: An epoch is one learning pass through the training set. One learning pass involves the presentation of training patterns, the calculation of the activation of each neuron and modification of the weights (page 36).

Gradient Descent Optimization: In Gradient descent optimization (GD) the minimum of the objective function is searched in the negative gradient of the objective function. In NNs, the objective function is the error function which is a function of the weights of the NN (page 27).

Incremental Learning: Incremental learning is a form of active learning, where a subset of the training patterns that satisfies a selection criterion is selected for training. Patterns are however selected and removed from the candidate set D_C into the actual training set D_T. The effect of incremental learning is that the training set D_T is grown while the candidate set D_C is pruned during training (page 46).

Momentum: Momentum is a term added to weight adjustments to help avoid oscillations in weight changes during training. This term is proportional to the magnitude of previous weight changes (page 37).

Mean Squared Error: In the context of neural networks, the mean squared error (MSE) is defined as the mean of the squared sum of the error between target values $t_k^{(p)}$ and the actual NN output values $o_k^{(p)}$:

$$MSE = \frac{\sum_{p=1}^{P} \sum_{k=1}^{K} (t_k^{(p)} - o_k^{(p)})^2}{2PK}$$

where P is the total number of patterns and K is the number of output units (page 11).

Pattern Presentations: A pattern presentation is a single pattern presented to the network for training. Pattern presentations are the total number of patterns presented so far to the network at a particular epoch (page 76).

Sensitivity Analysis for Incremental Learning: Sensitivity analysis for incremental learning (SAILA) is Engelbrecht's algorithm for active learning. Patterns are selected for training using the changes in output caused by perturbations in input parameters (page 56).

Selective Learning: Selective learning is an active learning algorithm, where a subset of the training patterns that satisfies a selection criterion is selected for training. Unlike incremental learning, the candidate set remain fixed while the size of the actual training set varies from time to time (page 46).

Selective Learning Algorithm: The selective learning algorithm (SLA) is a new active learning algorithm proposed in this thesis, which uses information on the next-time-changes to select patterns for training (page 54).

Subset selection Criterion: Subset selection criteria are criteria tested to determine whether a NN should select additional patterns into the current training subset D_T (page 52).

Sum Squared Error: SSE is the sum of squared errors, defined as

$$SSE = \frac{1}{2}\sum_{p=1}^{P}\sum_{k=1}^{K}(t_k^{(p)} - o_k^{(p)})^2$$

(page 31).