

## 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
$\alpha$	momentum term
$\eta$	learning rate
$\kappa$	Zhang's notation for a specified performance level
ho	Röbel's generalization factor
au	Zhang's notation for allowable error tolerance per connection
$\Phi$	Engelbrecht's notation for pattern informativeness
$\Psi$	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
	$\overline{C_{al}}$	cost of the subsetselection criteria for Zhang's accelerated learning (AL)
$\begin{array}{lll} C_{fs} & \text{cost of training a NN on a training set} \\ C_{fsl} & \text{cost of fixed set learning} \\ C_s & \text{cost of selecting patterns} \\ C_{sai} & \text{cost of the subsetselection criteria for SAILA} \\ cost of the subsetselection criteria for selective learning algorithm (SLA) \\ cost of selecting a pattern into D_T for Zhang's accelerated learning C_{S_{als}} & \text{cost of selecting a pattern into } D_T for Röbel's dynamic pattern selection cost of selecting a pattern into D_T for Röbel's dynamic pattern selection C_{S_{als}} & \text{cost of training a network using Zhang's algorithm} \\ C_{SL} & \text{cost of training a network using Engelbrech's algorithm} \\ C_{SL} & \text{cost of training a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using selective learning algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating all weights between the input and hidden layers} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & \text{cost of updating a network using Engelbrech's algorithm} \\ C_v & cost of updating a network using Engelbrech's alg$	$C_{dps}$	cost of the subsetselection criteria for Röbel's dynamic pattern selection (DPS)
$\begin{array}{lll} C_{fsl} & \operatorname{cost} & \operatorname{of} & \operatorname{fixed} & \operatorname{set} & \operatorname{learning} \\ C_{s} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{patterns} \\ C_{sai} & \operatorname{cost} & \operatorname{of} & \operatorname{the} & \operatorname{subsetselection} & \operatorname{criteria} & \operatorname{for} & \operatorname{SAILA} \\ C_{sla} & \operatorname{cost} & \operatorname{of} & \operatorname{the} & \operatorname{subsetselection} & \operatorname{criteria} & \operatorname{for} & \operatorname{Salta} & \operatorname{cost} & \operatorname{of} & \operatorname{the} & \operatorname{subsetselection} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Röbel's} & \operatorname{dynamic} & \operatorname{pattern} & \operatorname{selection} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Röbel's} & \operatorname{dynamic} & \operatorname{pattern} & \operatorname{selection} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{apattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Röbel's} & \operatorname{dynamic} & \operatorname{pattern} & \operatorname{selection} & \operatorname{cost} & \operatorname{ost} & \operatorname{ost} & \operatorname{selecting} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Röbel's} & \operatorname{dynamic} & \operatorname{pattern} & \operatorname{selection} & \operatorname{cost} & \operatorname{ost} & \operatorname{ost} & \operatorname{ost} & \operatorname{selecting} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{ost} & \operatorname{selecting} & \operatorname{pattern} & patt$		cost of training a NN on a training set
$\begin{array}{lll} C_s & \text{cost of selecting patterns} \\ C_{sai} & \text{cost of the subsetselection criteria for SAILA} \\ C_{Sla} & \text{cost of the subsetselection criteria for selective learning algorithm (SLA)} \\ C_{Sal} & \text{cost of selecting a pattern into } D_T \text{ for Zhang's accelerated learning} \\ C_{Sal} & \text{cost of selecting a pattern into } D_T \text{ for Röbel's dynamic pattern selection} \\ C_{Sa} & \text{cost of selecting a pattern into } D_T \text{ for SAILA} \\ C_{DPS} & \text{cost of training a network using Zhang's algorithm} \\ C_{DPS} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SA} & \text{cost of training a network using selective learning algorithm} \\ C_{S} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{cost of updating all weights between the hidden and output layers} \\ c_{W} & \text{validation set} \\ I & \text{total number of input units} \\ I & \text{total number of input units} \\ I & \text{total number of weights between input and hidden layers of a network} \\ I & \text{total number of weights between hidden and output layers of a network} \\ I & \text{total number of weights between hidden and output layers of a network} \\ I & \text{total number of patterns in the candidate set } D_{C} \\ I & \text{number of patterns in a subset } D_{S} \\ I & \text{number of patterns in a subset } D_{S} \\ I & \text{number of patterns in the training set} \\ D_{T} & \text{number of patterns in the training set} \\ D_{T} & \text{number of patterns in the training set} \\ D_{T} & \text{number of patterns in the training set} \\ D_{T} & \text{number of patterns in the training set} \\ D_{$		cost of fixed set learning
$\begin{array}{lll} C_{sai} & \operatorname{cost} & \operatorname{of} & \operatorname{the subsetselection} & \operatorname{criteria} & \operatorname{for} & \operatorname{SAILA} \\ C_{sla} & \operatorname{cost} & \operatorname{of} & \operatorname{the subselection} & \operatorname{criteria} & \operatorname{for} & \operatorname{selective} & \operatorname{learning} & \operatorname{algorithm} & \operatorname{(SLA)} \\ C_{S_{ait}} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{a} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Zhang's} & \operatorname{accelerated} & \operatorname{learning} \\ C_{S_{min}} & \operatorname{cost} & \operatorname{of} & \operatorname{selecting} & \operatorname{a} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{Zhala} \\ C_{AL} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Zhang's} & \operatorname{algorithm} \\ C_{DPS} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Zhang's} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ C_{SL} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{selectiv} & \operatorname{learning} & \operatorname{algorithm} \\ \operatorname{cost} & \operatorname{oth} & \operatorname{unith} & \operatorname$		cost of selecting patterns
$\begin{array}{lll} C_{sla} & \operatorname{cost} \text{ of the subselection criteria for selective learning algorithm (SLA)} \\ C_{S_{alos}} & \operatorname{cost} \text{ of selecting a pattern into } D_T \text{ for Zhang's accelerated learning} \\ cost of selecting a pattern into } D_T \text{ for Röbel's dynamic pattern selection} \\ C_{S_{ani}} & \operatorname{cost} \text{ of selecting a pattern into } D_T \text{ for Röhel's dynamic pattern selection} \\ C_{S_{ani}} & \operatorname{cost} \text{ of training a network using Zhang's algorithm} \\ C_{DPS} & \operatorname{cost} \text{ of training a network using Engelbrecht's algorithm} \\ C_{SA} & \operatorname{cost} \text{ of training a network using selective learning algorithm} \\ C_{V} & \operatorname{cost} \text{ of updating a weight between the input and hidden layers} \\ C_{V} & \operatorname{cost} \text{ of updating all weights between the input and hidden layers} \\ C_{W} & \operatorname{cost} \text{ of updating all weights between the hidden and output layers} \\ C_{W} & \operatorname{cost} \text{ of updating all weights between the hidden and output layers} \\ C_{W} & \operatorname{cost} \text{ of updating all weights between the hidden and output layers} \\ D_{C} & \operatorname{est} \text{ of candidate training patterns} \\ D_{G} & \operatorname{test} \text{ set} \text{ or the generalization set} \\ D_{T} & \operatorname{actual training set} \\ D_{V} & \operatorname{validation set} \\ I & \operatorname{total number of input units} \\ K & \operatorname{total number of output units} \\ K & \operatorname{total number of weights between input and hidden layers of a network} \\ N_{W} & \operatorname{total number of weights between input and hidden layers of a network} \\ number of patterns in the candidate set D_{C} \\ number of patterns in the subset D_{S} \\ number of patterns in a subset D_{S} \\ number of patterns in the training set D_{T} \\ number of patterns in the validation set D_{V} \\ O_{E} \\ k-th output unit O_{S} \\ number of patterns in the validation set D_{V} \\ number of patterns in the validation set D_{V} \\ number of patterns in the validation set D_{V} \\ number of patterns in the validation set D_{V} \\ number of patterns in the validation set D_{V} \\ number of patterns in the validation set D_{V} \\ number of patterns \\ number of number of number of n$		cost of the subsetselection criteria for SAILA
$\begin{array}{lll} C_{S_{als}} & \text{cost of selecting a pattern into } D_T \text{ for Zhang's accelerated learning} \\ C_{S_{dys}} & \text{cost of selecting a pattern into } D_T \text{ for Röbel's dynamic pattern selection} \\ C_{S_{als}} & \text{cost of training a network using Zhang's algorithm} \\ C_{DPS} & \text{cost of training a network using Röbel's algorithm} \\ C_{SA} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SL} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{V} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the input and hidden layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ D_{C} & \text{set of candidate training patterns} \\ D_{G} & \text{test set or the generalization set} \\ D_{T} & \text{actual training set} \\ D_{V} & \text{validation set} \\ I & \text{total number of input units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ V_{W} & \text{total number of weights between hidden and output layers of a network} \\ P_{G} & \text{number of patterns in the candidate set } D_{C} \\ P_{G} & \text{number of patterns in the generalization set } D_{G} \\ P_{S} & \text{number of patterns in the training set } D_{T} \\ \text{number of patterns in the validation set } D_{V} \\ V_{C} & \text{k-th output unit} \\ O_{E} & \text{k-th output unit on the pattern } P_{C} \\ \text{number of patterns in the validation set } D_{V} \\ \text{k-th output unit} \\ O_{E} & \text{k-th output unit on the pattern } P_{C} \\ \text{number of patterns in the validation set } D_{V} \\ \text{k-th output unit} \\ O_{E} & \text{total output unit and } P_{C} \\ \text{total output unit and } P_{C}$		cost of the subselection criteria for selective learning algorithm (SLA)
$\begin{array}{lll} C_{S_{dys}} & \text{cost of selecting a pattern into } D_T \text{ for Röbel's dynamic pattern selection} \\ C_{S_{dai}} & \text{cost of selecting a pattern into } D_T \text{ for SAILA} \\ cost of training a network using Zhang's algorithm} \\ C_{DPS} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SA} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SL} & \text{cost of training a network using Selective learning algorithm} \\ C_{V} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the input and hidden layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weights between the hidden and output layers} \\ S_{CW} & \text{cost of updating all weight between the hidden layers} \\ S_{CW} & \text{cost of updating all weight between input and hidden layers} \\ S_{CW} & \text{cost of updating all weight between hidden and output layers of a network} \\ S_{CW} & \text{total number of output units} \\ S_{CW} & \text{total number of patterns in the candidate set } D_{C} \\ S_{CW} & \text{total number of patterns in the candidate set } D_{C} \\ S_{CW} & \text{total number of patterns in the training set } D_{T} \\ S_{CW} & \text{total number of patterns in the training set } D_{T} \\ S_{CW} & \text{total number of patterns in the validation set } D_{V} \\ S_{CW} & \text{total number of patterns in the validation set } D_{V} \\ S_{CW} & \text{total number of patterns in the validation set } D_{V} \\ S_{CW} & \text{total number of output unit } S_{CW} & total num$		cost of selecting a pattern into $D_T$ for Zhang's accelerated learning
$\begin{array}{lll} C_{S_{sai}} & \operatorname{cost} & \operatorname{selecting} & \operatorname{a} & \operatorname{pattern} & \operatorname{into} & D_T & \operatorname{for} & \operatorname{SAILA} \\ C_{AL} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Zhang's} & \operatorname{algorithm} \\ C_{DPS} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Engelbrecht's} & \operatorname{algorithm} \\ C_{SA} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Engelbrecht's} & \operatorname{algorithm} \\ C_{SL} & \operatorname{cost} & \operatorname{of} & \operatorname{training} & \operatorname{a} & \operatorname{network} & \operatorname{using} & \operatorname{Engelbrecht's} & \operatorname{algorithm} \\ C_{V} & \operatorname{cost} & \operatorname{of} & \operatorname{updating} & \operatorname{a} & \operatorname{weight} & \operatorname{between} & \operatorname{the} & \operatorname{input} & \operatorname{and} & \operatorname{hidden} & \operatorname{lavers} \\ C_{V} & \operatorname{cost} & \operatorname{of} & \operatorname{updating} & \operatorname{a} & \operatorname{weight} & \operatorname{between} & \operatorname{the} & \operatorname{hidden} & \operatorname{and} & \operatorname{output} & \operatorname{lavers} \\ C_{W} & \operatorname{cost} & \operatorname{of} & \operatorname{updating} & \operatorname{all} & \operatorname{weight} & \operatorname{between} & \operatorname{the} & \operatorname{hidden} & \operatorname{and} & \operatorname{output} & \operatorname{lavers} \\ C_{W} & \operatorname{cost} & \operatorname{of} & \operatorname{updating} & \operatorname{all} & \operatorname{weight} & \operatorname{between} & \operatorname{the} & \operatorname{hidden} & \operatorname{and} & \operatorname{output} & \operatorname{lavers} \\ C_{W} & \operatorname{cost} & \operatorname{of} & \operatorname{updating} & \operatorname{all} & \operatorname{weight} & \operatorname{between} & \operatorname{the} & \operatorname{hidden} & \operatorname{and} & \operatorname{output} & \operatorname{lavers} \\ D_{C} & \operatorname{set} & \operatorname{of} & \operatorname{candidate} & \operatorname{training} & \operatorname{patterns} \\ D_{C} & \operatorname{total} & \operatorname{number} & \operatorname{of} & \operatorname{input} & \operatorname{units} \\ D_{V} & \operatorname{total} & \operatorname{number} & \operatorname{of} & \operatorname{updating} & \operatorname{patterns} & \operatorname{number} & \operatorname{of} & \operatorname{a} & \operatorname{network} \\ D_{V} & \operatorname{total} & \operatorname{number} & \operatorname{of} & \operatorname{patterns} & \operatorname{patterns}$	$C_{S_{dns}}$	cost of selecting a pattern into $D_T$ for Röbel's dynamic pattern selection
$\begin{array}{lll} C_{AL} & \text{cost of training a network using Zhang's algorithm} \\ C_{DPS} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SA} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SL} & \text{cost of training a network using Selective learning algorithm} \\ C_{v} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the hidden and output layers} \\ C_{w} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ D_{C} & \text{set of candidate training patterns} \\ D_{G} & \text{test set or the generalization set} \\ D_{T} & \text{actual training set} \\ D_{V} & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_{V} & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ P_{C} & \text{number of patterns in the candidate set } D_{C} \\ number & \text{of patterns in the generalization set } D_{G} \\ number & \text{of patterns in the training set } D_{T} \\ number & \text{of patterns in the training set } D_{T} \\ number & \text{of patterns in the validation set } D_{V} \\ o_{k} & k\text{-th output unit} \\ o_{k}^{(p)} & \text{activation of output unit } o_{k} & \text{for pattern } p \\ \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_{j} & \text{thidden unit} \\ y_{j} & \text{thidden unit} \\ \end{array}$	$C_{S_{sai}}$	cost of selecting a pattern into $D_T$ for SAILA
$\begin{array}{lll} C_{SA} & \text{cost of training a network using Engelbrecht's algorithm} \\ C_{SL} & \text{cost of training a network using selective learning algorithm} \\ C_{v} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the hidden and output layers} \\ C_{w} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ D_{C} & \text{set of candidate training patterns} \\ D_{G} & \text{test set or the generalization set} \\ D_{T} & \text{actual training set} \\ D_{V} & \text{validation set} \\ I & \text{total number of input units} \\ K & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_{V} & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ p_{C} & \text{number of patterns in the candidate set } D_{C} \\ P_{G} & \text{number of patterns in the generalization set } D_{G} \\ P_{s} & \text{number of patterns in the training set } D_{T} \\ P_{V} & \text{number of patterns in the validation set } D_{V} \\ o_{k} & \text{k-th output unit} \\ o_{k}^{(p)} & \text{activation of output unit } o_{k} \text{ for pattern } p \\ \text{weight between } p\text{-th hidden unit and } p\text{-th hidden unit} \\ y_{j} & \text{yieight between } k\text{-th output unit and } j\text{-th hidden unit} \\ y_{j} & \text{yieight between } k\text{-th output unit and } j\text{-th hidden unit} \\ \end{array}$		cost of training a network using Zhang's algorithm
$\begin{array}{lll} C_{SL} & \text{cost of training a network using selective learning algorithm} \\ C_{V} & \text{cost of updating a weight between the input and hidden layers} \\ C_{V} & \text{cost of updating all weights between the input and hidden layers} \\ C_{w} & \text{cost of updating all weights between the hidden and output layers} \\ C_{W} & \text{cost of updating all weights between the hidden and output layers} \\ D_{C} & \text{set of candidate training patterns} \\ D_{G} & \text{test set or the generalization set} \\ D_{T} & \text{actual training set} \\ D_{V} & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_{W} & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ P_{C} & \text{number of patterns in the candidate set } D_{C} \\ P_{G} & \text{number of patterns in a subset } D_{s} \\ P_{T} & \text{number of patterns in a subset } D_{s} \\ P_{T} & \text{number of patterns in the training set } D_{T} \\ N_{W} & \text{k-th output unit} \\ O_{k}^{(p)} & \text{activation of output unit } o_{k} \text{ for pattern } p \\ \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_{j} & \text{weight between } k\text{-th output unit} \\ y_{j} & \text{th idden unit} \\ y_{j} & \text{th idden unit} \\ \end{array}$	$C_{DPS}$	cost of training a network using Röbel's algorithm
$\begin{array}{lll} C_v & \text{cost of updating a weight between the input and hidden layers} \\ C_V & \text{cost of updating all weights between the input and hidden layers} \\ C_w & \text{cost of updating a weight between the hidden and output layers} \\ C_W & \text{cost of updating all weights between the hidden and output layers} \\ D_C & \text{set of candidate training patterns} \\ D_G & \text{test set or the generalization set} \\ D_T & \text{actual training set} \\ D_V & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_V & \text{total number of weights between hidden and output layers of a network} \\ P_G & \text{number of patterns in the candidate set } D_C \\ P_G & \text{number of patterns in the generalization set } D_G \\ P_s & \text{number of patterns in the training set } D_T \\ number & \text{of patterns in the training set } D_V \\ k-\text{th output unit} \\ o_k^{(p)} & \text{activation of output unit } o_k \text{ for pattern } p \\ weight & \text{between } j\text{-th hidden unit} \\ weight & \text{between } k\text{-th output unit and } j\text{-th hidden unit} \\ y_j & \text{j-th hidden unit} \\ \end{array}$	$C_{SA}$	cost of training a network using Engelbrecht's algorithm
$\begin{array}{lll} C_V & \text{cost of updating all weights between the input and hidden layers} \\ C_W & \text{cost of updating a weight between the hidden and output layers} \\ C_W & \text{cost of updating all weights between the hidden and output layers} \\ D_C & \text{set of candidate training patterns} \\ D_G & \text{test set or the generalization set} \\ D_T & \text{actual training set} \\ D_V & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of output units} \\ N_V & \text{total number of weights between input and hidden layers of a network} \\ N_W & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ P_C & \text{number of patterns in the candidate set } D_C \\ P_S & \text{number of patterns in the generalization set } D_G \\ P_S & \text{number of patterns in a subset } D_S \\ P_T & \text{number of patterns in the training set } D_T \\ P_V & \text{number of patterns in the validation set } D_V \\ o_k & k\text{-th output unit} \\ o_k^{(p)} & \text{activation of output unit } o_k \text{ for pattern } p \\ v_{ji} & \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_j & j\text{-th hidden unit} \\ \end{array}$	$C_{SL}$	cost of training a network using selective learning algorithm
$\begin{array}{lll} C_w & \text{cost of updating a weight between the hidden and output layers} \\ C_W & \text{cost of updating all weights between the hidden and output layers} \\ D_C & \text{set of candidate training patterns} \\ D_G & \text{test set or the generalization set} \\ D_T & \text{actual training set} \\ D_V & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_V & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ p_C & \text{number of patterns in the candidate set } D_C \\ P_G & \text{number of patterns in the generalization set } D_G \\ P_s & \text{number of patterns in a subset } D_s \\ P_T & \text{number of patterns in the training set } D_T \\ P_V & \text{number of patterns in the validation set } D_V \\ o_k & k\text{-th output unit} \\ o_k^{(p)} & \text{activation of output unit } o_k \text{ for pattern } p \\ v_{ji} & \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_j & j\text{-th hidden unit} \\ \end{array}$	$C_v$	cost of updating a weight between the input and hidden layers
$\begin{array}{lll} C_W & \text{cost of updating all weights between the hidden and output layers} \\ D_C & \text{set of candidate training patterns} \\ D_G & \text{test set or the generalization set} \\ D_T & \text{actual training set} \\ D_V & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of output units} \\ K & \text{total number of weights between input and hidden layers of a network} \\ N_W & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ P_C & \text{number of patterns in the candidate set } D_C \\ P_G & \text{number of patterns in the generalization set } D_G \\ P_s & \text{number of patterns in a subset } D_s \\ P_T & \text{number of patterns in the training set } D_T \\ P_V & \text{number of patterns in the validation set } D_V \\ o_k & k\text{-th output unit} \\ o_k^{(p)} & \text{activation of output unit } o_k \text{ for pattern } p \\ v_{ji} & \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_j & j\text{-th hidden unit} \\ \end{array}$		cost of updating all weights between the input and hidden layers
$\begin{array}{llllllllllllllllllllllllllllllllllll$	$C_{w}$	cost of updating a weight between the hidden and output layers
$\begin{array}{lll} D_G & \text{test set or the generalization set} \\ D_T & \text{actual training set} \\ D_V & \text{validation set} \\ I & \text{total number of input units} \\ J & \text{total number of output units} \\ K & \text{total number of output units} \\ N_V & \text{total number of weights between input and hidden layers of a network} \\ N_W & \text{total number of weights between hidden and output layers of a network} \\ p & \text{a single pattern} \\ p_C & \text{number of patterns in the candidate set } D_C \\ P_G & \text{number of patterns in the generalization set } D_G \\ P_s & \text{number of patterns in a subset } D_s \\ P_T & \text{number of patterns in the training set } D_T \\ P_V & \text{number of patterns in the validation set } D_V \\ o_k & k\text{-th output unit} \\ o_k^{(p)} & \text{activation of output unit } o_k \text{ for pattern } p \\ v_{ji} & \text{weight between } j\text{-th hidden unit and } j\text{-th hidden unit} \\ y_j & j\text{-th hidden unit} \\ \end{array}$	$C_W$	cost of updating all weights between the hidden and output layers
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$D_C$	set of candidate training patterns
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$D_G$	test set or the generalization set
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$D_T$	actual training set
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$D_V$	validation set
$K$ 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 $p_C$ number of patterns in the generalization set $p_C$ number of patterns in a subset $p_C$ number of patterns in a subset $p_C$ number of patterns in the training set $p_C$ number of patterns in the validation set $p_C$ number of patterns in the validation set $p_C$ $p_C$ number of patterns in the validation set $p_C$	I	total number of input units
$N_V$ total number of weights between input and hidden layers of a network total number of weights between hidden and output layers of a network a single pattern number of patterns in the candidate set $D_C$ number of patterns in the generalization set $D_G$ number of patterns in a subset $D_s$ number of patterns in the training set $D_T$ number of patterns in the validation set $D_V$ $o_k$ $k$ -th output unit $o_k$ for pattern $p$ weight between $j$ -th hidden unit and $j$ -th hidden unit $y_j$ $j$ -th hidden unit		total number of hidden units
$N_W$ 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 $D_C$ number of patterns in the generalization set $D_G$ number of patterns in a subset $D_s$ number of patterns in the training set $D_T$ number of patterns in the validation set $D_V$ $O_k$ $ext{k-th output unit}$ $O_k^{(p)}$ activation of output unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit and $O_k$ in $O_k$ weight between $O_k$ $O_k$ weight between $O_k$ $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit and $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ the hidden unit $O_k$ for pattern $O_k$ weight between $O_k$ for pattern $O_k$ weight between $O_k$ for pattern $O_k$ weight between $O_k$ for pattern $O_k$ for pattern $O_k$ weight between $O_k$ for pattern $O_k$		total number of output units
$egin{array}{lll} p &  ext{a single pattern} \\ P_C &  ext{number of patterns in the candidate set $D_C$} \\ P_G &  ext{number of patterns in the generalization set $D_G$} \\ P_s &  ext{number of patterns in a subset $D_s$} \\ P_T &  ext{number of patterns in the training set $D_T$} \\ P_V &  ext{number of patterns in the validation set $D_V$} \\ o_k &  ext{$k$-th output unit} \\ o_k^{(p)} &  ext{activation of output unit $o_k$ for pattern $p$} \\ v_{ji} &  ext{weight between $j$-th hidden unit and $j$-th hidden unit} \\ y_j &  ext{$j$-th hidden unit} \\ \end{array}$	$N_V$	- · · · · · · · · · · · · · · · · · · ·
$P_C$ 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	$N_W$	total number of weights between hidden and output layers of a network
$P_G$ 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 $j$ -th hidden unit	p	3 1
$P_s$ 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	$P_C$	
$P_T$ 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		
$P_V$ 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		
$egin{array}{lll} o_k & k & k & k & k & k & k & k & k & k &$		-
$o_k^{(p)}$ 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 $j$ -th hidden unit	$P_V$	
$egin{array}{lll} v_{ji} &  ext{weight between $j$-th hidden unit and $i$-th input unit} \ weight between $k$-th output unit and $j$-th hidden unit} \ y_j &  ext{$j$-th hidden unit} \end{array}$		k-th output unit
$egin{array}{lll} v_{ji} &  ext{weight between $j$-th hidden unit and $i$-th input unit} \ weight between $k$-th output unit and $j$-th hidden unit} \ y_j &  ext{$j$-th hidden unit} \end{array}$	$o_k^{(p)}$	activation of output unit $o_k$ for pattern $p$
$y_j$	$v_{ji}$	weight between $j$ -th hidden unit and $i$ -th input unit
	$w_{kj}$	weight between $k$ -th output unit and $j$ -th hidden unit
(n)	$y_j$	j-th hidden unit
$y_i^{r}$ activation of hidden unit $y_j$ for pattern $p$	$y_j^{(p)}$	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).