[Weiss *et al.*, 1987] Sholom M. Weiss, Robert S. Galen, and Prasad V. Tadepalli. Optimizing the Predictive Value of Diagnostic Decision Rules. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 521–526, 95 First Street, Los Altos, California 94022, 1987. The American Association for Artificial Intelligence, Morgan Kaufmann Publishers, Inc.

[Weiss *et al.*, 1990] S. Weiss, R. Galen, and P. Tadepalli. Maximizing the Predictive Value of Production Rules. *Artificial Intelligence*, page in press, 1990.

[Quinlan, 1988] J. R. Quinlan. An empirical comparison of genetic and decision-tree classifiers. In *Proceedings of the fifth international workshop on Machine Learning*, University of Michigan, June 1988. Machine Learning.

[Rumelhart *et al.*, 1987] David E. Rumelhart, James L. McClelland, and The PDP Research Group, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. MIT Press: Bradford Books, 1987.

[Sanger, 1990a] Terence D. Sanger. A tree-structured adaptive network for function approximation in high dimensional spaces, 1990. submitted to *IEEE Trans. Neural Networks.*

[Sanger, 1990b] Terence D. Sanger. A tree-structured algorithm for reducing computation in networks with separable basis functions, 1990. *Neural Computation*, in press.

[Schlimmer and Fisher, 1988] J. Schlimmer and D. Fisher. A case study in incremental concept learning. In *Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI86)*, pages 564–569, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, 1988. The American Association for Artificial Intelligence, Morgan Kaufmann Publishers, Inc.

[Schwartz *et al.*, 1990] S. Schwartz, J. Wiles, and S. Phillips. Connectionist, Rule-based and Bayesian decision aids: an empirical comparison, November 1990. Paper presented to the annual meeting of the Cognitive Science Society, Sydney.

[Schwartz *et al.*, 1991] S. Schwartz, J. Wiles, and S. Phillips. Connectionist, Rule-based and Bayesian decision aids, January 1991. Paper presented to the Third International Workshop on Artificial Intelligence and Statistics, Ft. Lauderdale.

[Shavlik *et al.*, 1989] J. W. Shavlik, R. J. Mooney, and G. G. Towell. Symbolic and Neural Net Learning Algorithms: An Experimental Comparison. Technical Report 857, Department of Computer Science, University of Wisconsin, Madison, Wisconsin, June 1989.

[Sun *et al.*, 1988] G. Z. Sun, Y. C. Lee, and H. H. Chen. A novel net that learns sequential decision process. In Dana Z. Anderson, editor, *Neural Information Processing Systems*, pages 760–766. American Institute of Physics, New York, 1988.

[Tenorio and Lee, 1989] Manoel Fernando Tenorio and Wei-Tsih Lee. Self organizing neural network for optimum supervised learning. Technical Report TR-EE 89-30, Purdue Univ. School of Elec. Eng., June 1989. References hanson-88.

[Utgoff, 1988] Paul E. Utgoff. Id5: An incremental id3. In *Proceedings of the fifth international workshop on Machine Learning*, University of Michigan, June 1988. Machine Learning.

[Utgoff, 1990a] Paul. E. Utgoff. An Incremental Method for Finding Multivariate Splits for Decision Trees, 1990.

[Utgoff, 1990b] Paul. E. Utgoff. Perceptron Trees: A Case Study in Hybrid Concept Representations, 1990.

[Weiss and Kapouleas, 1989] Sholom M. Weiss and Ioannis Kapouleas. An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods. In *Proceedings of the International Joint Conference on Artificial Intelligence.* IJCAI, Morgan Kaufmann, 1989.

[Fisher and McKusick, 1989] Douglas H. Fisher and Kathleen B. McKusick. An Empirical Comparison of ID3 and Back-propagation. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 788–793, August 1989.

[Friedman, 1988] Jerome H. Friedman. Multivariate adaptive regression splines. Technical Report 102, Stanford Univ. Lab for Computational Statistics, 1988.

[Gallant, 1988] Stephen I. Gallant. Connectionist Expert Systems. *Communications of the ACM*, 31(2):152–169, 1988.

[Golea, 1990] Golea. A Growth Algorithm for Neural Network Decision Trees. *Europhysics Letters*, 12:205–210, July 1990.

[Ikeda *et al.*, 1976] Saburo Ikeda, Mikiko Ochiai, and Yoshikazu Sawaragi. Sequential GMDH algorithm and its application to river flow prediction. *IEEE Trans. Systems, Man, and Cybernetics*, SMC-6(7):473–479, July 1976.

[Ivakhnenko, 1971] A. G. Ivakhnenko. Polynomial theory of complex systems. *IEEE Trans. Systems, Man, and Cybernetics*, SMC-1(4):364–378, October 1971.

[Jacobs *et al.*, 1990] Robert A. Jacobs, Michael I. Jordan, and Andrew G. Barto. Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. Technical Report COINS TR 90-27, U. Mass., Amherst, March 1990.

[Knerr *et al.*, 1990] S. Knerr, L. Personnaz, and G. Dreyfus. Single-layer learning revisited: A stepwise procedure for building and training a neural network. manuscript, 1990.

[McClelland and Rumelhart, 1988] James L. McClelland and David E. Rumelhart. *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises.* The MIT Press, Cambridge, MA, 1988.

[Mezard and Nadal, 1989] Marc Mezard and Jean-Pierre Nadal. Learning in feedforward layered networks: The tiling algorithm. *J. Physics A*, 22:2191–2203, 1989.

[Omohundro, 1987] Stephen M. Omohundro. Efficient Algorithms with Neural Network Behavior. *Complex Systems*, 1:273–347, 1987.

[Omohundro, 1990a] Stephen M. Omohundro. Bumptrees for Efficient Function, Constraint, and Classification Learning: Summary, 1990.

[Omohundro, 1990b] Stephen M. Omohundro. Geometric Learning Algorithms. *Physica D*, 42:307–321, 1990.

[Quinlan, 1983] J. R. Quinlan. Learning efficient classification procedures and their application to Chess End Games. In *Machine Learning*, pages 463–482. Palo Alto, CA: Tioga Publishing Company, 1983.

[Quinlan, 1987a] J. R. Quinlan. Generating Production Rules From Decision Trees. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, 1987.

[Quinlan, 1987b] J. R. Quinlan. Simplifying Decision Trees. *International Journal of Man Machine Studies*, 27, 1987.

References

[Atlas *et al.*, 1990a] L. E. Atlas, J. Connor, D. Park, A. Lippman, R. Cole, M. El-Sharkawi, R. J. Marks, Y. Muthusamy, and M. Rudnick. A Performance Comparison of Trained Multi-Layer Perceptrons and Trained Classification Trees. In *IEEE International Conference on Systems, Man, and Cybernetics, Cambridge, Massachusetts, November 14-17*, 1990.

[Atlas *et al.*, 1990b] Les Atlas, Ronald Cole, Jerome Connor, Mohamed El-Sharkawi, Robert J. Marks II, Yeshwant Muthusamy, and Etienne Barnard. Performance Comparisons between Backpropagation Networks and Classification Trees on Three Real-world applications, 1990.

[Atlas *et al.*, 1990c] Les Atlas, Ronald Cole, Yeshwant Muthasamy, Alan Lippman, Jerome Connor, Dong Park, Mohamed El-Sharkawi, and Robert J. Marks II. A Performance Comparison of Trained Multi-Layer Perceptrons and Trained Classification Trees. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 622–629, San Mateo, CA, 1990. Morgan Kaufmann.

[Atlas *et al.*, 1990d] Les Atlas, Ronald Cole, Yeshwant Muthusamy, A. Lippman, Jerome Connor, D. Park, M. El-Sharkawi, and Robert J. Marks II. Performance Comparisons between Backpropagation Networks and Classification Trees on Three Real-world applications. *Proceedings of the IEEE*, October 1990.

[Barron *et al.*, 1984] R. L. Barron, A. N. Mucciardi, F. J. Cook, J. N. Craig, and A. R. Barron. Adaptive learning networks: Development and application in the United States of algorithms related to GMDH. In S. J. Farlow, editor, *Self-Organizing Methods in Modeling*. Marcel Dekker, New York, 1984.

[Bentley, 1975] J. H. Bentley. Multidimensional binary search trees used for associated searching. *Communications ACM*, 18(9):509–517, 1975.

[Breiman *et al.*, 1984] L. Breiman, J. H. Friedman, R. A. Olshen, and J. Stone. *Classification and Regression Trees.* Wadsworth and Brooks, Monterey, CA, 1984.

[Chen and Wohl, 1989] Francine Chen and Peter Wohl. An extended version of ID3. In *DARPA Speech and Natural Language Workshop, Philadelpha, PA*, pages 284–289, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, February 1989.

[Cole *et al.*, 1990] R. A. Cole, Y. K. Muthusamy, L. Atlas, T. Leen, and M. Rudnick. Speaker-Independent Vowel Recognition: Comparison of Backpropagation and Trained Classification Trees. In *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences, Kailua-Kona, Hawaii, January 2-5*, 1990.

[Cybenko, 1989] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 1989.

[Dattatreya and Kanal, 1985] G. R. Dattatreya and L. N. Kanal. Decision Trees in Pattern Recognition. In L. Kanal and A. Rosenfeld, editors, *Progress in Pattern Recognition 2*. North-Holland, 1985. An extensive survey of the use of decision trees in the field of pattern recognition.

[Fahlman, 1990] Scott Fahlman. The Cascade-Correlation Learning Architecture. Technical Report CMU-CS-90-100, Carnegie Mellon Computer Science Department, Pittsburgh, PA, February 1990.

1.3 Performance

ID3/CART	Back-propagation
Serial, one node at a time	Parallel by level
Missing data by surrogate splits	
Can work with a subset of data (i.e. you can avoid performing costly tests)	Requires all data as input (although see [Gallant, 1988] hybrid)

2 Pointers to research

Here are some pointers to relevant research in the fields discussed here.

ID3 is described by [Quinlan, 1983]

An advanced version of ID3 is described in [Chen and Wohl, 1989].

ID4 is described in [Schlimmer and Fisher, 1988]

ID5 is described in [Utgoff, 1988]

C4 is described in [Quinlan, 1987a] and [Quinlan, 1987b] (extended version)

C4 is empirically measured in [Quinlan, 1988]

[Dattatreya and Kanal, 1985] gives an extensive survey of decision trees in pattern recognition.

CART is described in [Breiman *et al.*, 1984], [Sun *et al.*, 1988].

The work described in [Fahlman, 1990] is also tree-like.

Back-Propagation is described in Chapter 8 of [Rumelhart *et al.*, 1987], which introduces the algorithm, and Chapter 5 of [McClelland and Rumelhart, 1988], which includes simulator software.

Empirical comparison work is in [Fisher and McKusick, 1989], [Shavlik *et al.*, 1989], [Weiss and Kapouleas, 1989], [Atlas *et al.*, 1990b], [Atlas *et al.*, 1990c], [Atlas *et al.*, 1990d], [Cole *et al.*, 1990], [Atlas *et al.*, 1990a].

Sanger's work includes [Sanger, 1990b], [Sanger, 1990a]

Omohundro's papers include: [Omohundro, 1987], [Omohundro, 1990b], [Omohundro, 1990a].

Papers related to decision tree learning include: [Mezard and Nadal, 1989] (tiling) [Friedman, 1988] (MARS), [Ikeda *et al.*, 1976], [Barron *et al.*, 1984], [Ivakhnenko, 1971], [Tenorio and Lee, 1989] (all on GMDH), [Knerr *et al.*, 1990] (upstart), [Bentley, 1975] (k-d tree), and [Jacobs *et al.*, 1990] is also related, because it explores how one network can control another.

[Golea, 1990] describes a hybrid network/decision tree.

Also, see [Schwartz *et al.*, 1990] and [Schwartz *et al.*, 1991].

1.2 Learning

ID3/CART	**Back-propagation**
Learning consists of placing one hyperplane at a time	Learning consists of starting with random hyperplanes and shifting their position until they are in a locally maximal configuration
Learns quickly	Learns slowly (?)
Primarily Batch	Incremental/Batch
Determines its own architecture	Requires architecture as input
Not good at determining proper thresholds. CART: inefficient, ID3: doesn't do it at all ([Weiss *et al.*, 1987], [Utgoff, 1990a], and [Weiss *et al.*, 1990] address this issue, however)	Determining thresholds is part of learning algorithm

1 Neural network comparison framework.

1.1 Representation:

ID3/CART	**Back-propagation**
Each hyperplane divides an existing polygon into two	Each hyperplane completely divides n-space (n = #of input units/attributes)
	More compact?
Understandable (?)	Not so easy to understand (will people use them?)
Boolean fn. universal (DNF equiv.)	Boolean fn. universal([Cybenko, 1989])
ID3: monothetic (axis-parallel decision regions) (although see [Utgoff, 1990b])	polythetic (from [Fisher and McKusick, 1989])
CART: polythetic decisions over attributes possible	polythetic over attributes, and over results of those decisions
Subset	Superset (**every decision tree can be represented using a back-propagation network**)
Classification certainty is determined as a by-product of learning.	
A reliable data structure for memorizing data.	No guarantees to correctly store data (complex methodology)
Pruning algorithms for generalization	Generalization via node growing, brain damage, constrained resources.

Neural Networks and Decision Tree Induction: Exploring the relationship between two research areas

A NIPS '90 workshop, 11/30/1990, Keystone, Colorado

Workshop Handout/bibliography

Workshop co-chairs: L. Y. Pratt* S. W. Norton†

The fields of Neural Networks and Machine Learning have evolved separately in many ways. However, close examination of multilayer perceptron learning algorithms (such as Back-Propagation) and decision tree induction methods (such as ID3 and CART) reveals that there is considerable convergence between these subfields. They address similar problem classes (inductive classifier learning) and can be characterized by a common representational formalism of hyperplane decision regions. Furthermore, topical subjects within both fields are related, from minimal trees and network reduction schemes to incremental learning.

In this workshop, invited speakers from the Neural Network and Machine Learning communities will discuss their empirical and theoretical comparisons of the two areas, and then present work at the interface between these two fields which takes advantage of the potential for technology transfer between them. In a summary period, we'll discuss our conclusions, comparing the methods along the dimensions of representation, learning, and performance. We'll debate the "strong convergence hypothesis" that these two research areas are really studying the same problem.

Schedule of talks:

AM:		
7:30-7:50	Lori Pratt	Introduction
7:50-8:10	Tom Dietterich	Evidence For and Against Convergence: Experiments Comparing ID3 and BP
8:15-8:35	Les Atlas	Is backpropagation really better than classification and regression trees?
8:40-9:00	Ah Chung Tsoi	Comparison of the performance of some popular machine learning algorithms: CART, C4.5, and multi-layer perceptrons
9:05-9:25	Ananth Sankar	Neural Trees: A Hybrid Approach to Pattern Recognition
PM:		
4:30-4:55	Stephen Omohundro	A Bayesian View of Learning with Tree Structures and Neural Networks
5:00-5:20	Paul Utgoff	Linear Machine Decision Trees
5:25-5:45	Terry Sanger	Basis Function Trees as a Generalization of CART, MARS, and Other Local Variable Selection Techniques
5:50-6:30	Discussion, wrap-up.	

*pratt@paul.rutgers.edu, Rutgers University Computer Science Department, New Brunswick, NJ 08903. (201) 932-4634 (Rutgers)

†norton@learning.siemens.com, Siemens Corporate Research, Inc., 755 College Road East, Princeton, NJ 08540