

Table of Contents

Preface	vii
How To Obtain the C4.5 Software	ix
1 Introduction	1
1.1 Example: Labor negotiation settlements	3
1.2 Other kinds of classification models	12
1.3 What lies ahead	16
2 Constructing Decision Trees	17
2.1 Divide and conquer	17
2.2 Evaluating tests	20
2.3 Possible tests considered	24
2.4 Tests on continuous attributes	25
3 Unknown Attribute Values	27
3.1 Adapting the previous algorithms	28
3.2 Play/Don't Play example again	30
3.3 Recapitulation	32
4 Pruning Decision Trees	35
4.1 When to simplify?	36
4.2 Error-based pruning	37
4.3 Example: Democrats and Republicans	41
4.4 Estimating error rates for trees	42
5 From Trees to Rules	45
5.1 Generalizing single rules	47
5.2 Class rulesets	50
5.3 Ranking classes and choosing a default	54
5.4 Summary	55
6 Windowing	57
6.1 Example: Hypothyroid conditions revisited	58
6.2 Why retain windowing?	58
6.3 Example: The multiplexor	60
7 Grouping Attribute Values	63
7.1 Finding value groups by merging	64

7.2	Example: Soybean diseases	65
7.3	When to form groups?	66
7.4	Example: The Monk's problems	67
7.5	Uneasy reflections	69
8	Interacting with Classification Models	71
8.1	Decision tree models	71
8.2	Production rule models	78
8.3	Caveat	80
9	Guide to Using the System	81
9.1	Files	81
9.2	Running the programs	84
9.3	Conducting experiments	89
9.4	Using options: A credit approval example	91
10	Limitations	95
10.1	Geometric interpretation	95
10.2	Nonrectangular regions	96
10.3	Poorly delineated regions	98
10.4	Fragmented regions	100
10.5	A more cheerful note	102
11	Desirable Additions	103
11.1	Continuous classes	103
11.2	Ordered discrete attributes	104
11.3	Structured attributes	104
11.4	Structured induction	105
11.5	Incremental induction	106
11.6	Prospectus	107
	Appendix: Program Listings	109
	Brief descriptions of the contents of files	110
	Notes on some important data structures	112
	Source code for the system	115
	Alphabetic index of routines	288
	References and Bibliography	291
	Author Index	297
	Subject Index	299