
CONTENTS

1	Introduction	11
1.1	Aims and an outline	15
1.2	Datasets	16
1.2.1	Dust Data	17
1.2.2	Restenosis Data	18
1.2.3	Breast Cancer Data	22
1.2.4	Gastric Cancer Data	22
2	Predicting Outcomes	25
2.1	Predicting by learning	25
2.2	Predicting specific outcome types	26
2.2.1	Classification	26
2.2.2	Event–History	29
3	Recursive partitioning	33
3.1	Growing trees	33
3.1.1	The splitting process	34
3.1.2	Stopping rules	37
3.1.3	Finding predictors for terminal nodes	39
3.2	Using the tree as a predictor	40
3.3	Assessing tree performance	40
3.4	Finding trees of right size	45
3.5	Dealing with missing values	48
3.6	Adapting to other response types	49
3.6.1	Regression setting	50
3.6.2	Event History setting	52

4	Predicting with CART	60
4.1	Dust Data	60
4.2	Restenosis Data	67
4.3	Breast Cancer Data	71
4.4	Gastric Cancer Data	75
5	Improving prediction through refinement	79
5.1	Tuning minimum node size	79
5.1.1	Motivation	80
5.1.2	Implementation	82
5.1.3	Results	83
5.2	Allowing for nonbinary splits	87
5.2.1	Motivation	87
5.2.2	Implementation	91
5.2.3	Results	99
6	Improving prediction via stabilization	106
6.1	Assessing tree stability	106
6.1.1	Graphing the split criterion	107
6.1.2	Bootstrap confidence intervals for the cutpoints	108
6.1.3	Assessing factor importance rankings	109
6.2	Accounting for the multiple testing situation	112
6.2.1	Implementation	113
6.2.2	Results	114
6.3	Stabilizing the predictor	121
6.3.1	Tree-based predictor aggregation via the bootstrap	123
6.3.2	Node-level stabilization via the bootstrap	128
	Summary and outlook	132
A	Algorithmic Implementation	135
B	Table of Symbols used	137
	Bibliography	139