

Richard Berk

Criminal Justice Forecasts of Risk

A Machine Learning Approach

 Springer

Contents

| | | |
|----------|--|----|
| 1 | Getting Started | 1 |
| 1.1 | Setting the Stage | 1 |
| 1.1.1 | A Brief Motivating Example | 4 |
| 2 | Some Important Background Material | 7 |
| 2.1 | Policy Considerations | 7 |
| 2.1.1 | Criminal Justice Forecasting Goals | 7 |
| 2.1.2 | Decisions to be Informed by the Forecasts | 9 |
| 2.1.3 | Outcomes To be Forecast | 10 |
| 2.2 | Data Considerations | 11 |
| 2.2.1 | Stability Over Time | 11 |
| 2.2.2 | Training Data and Test Data | 12 |
| 2.2.3 | Representative Data | 13 |
| 2.2.4 | Large Samples | 14 |
| 2.2.5 | Data Continuity | 15 |
| 2.2.6 | Missing Data | 16 |
| 2.3 | Statistical Considerations | 17 |
| 2.3.1 | Actuarial Methods | 18 |
| 2.3.2 | Building in the Relative Costs of Forecasting Errors | 19 |
| 2.3.3 | Effective Forecasting Algorithms | 22 |
| 3 | A Conceptual Introduction to Classification and Forecasting | 27 |
| 3.1 | Populations and Samples | 27 |
| 3.2 | Classification and Forecasting Using Decision Boundaries | 27 |
| 3.3 | Classification by Data Partitions | 30 |
| 3.4 | Forecasting by Data Partitions | 33 |
| 3.5 | Finding Good Data Partitions | 34 |
| 3.6 | Enter Asymmetric Costs | 35 |
| 3.7 | Recursive Partitioning Classification Trees | 37 |
| 3.7.1 | How Many Terminal Nodes? | 39 |
| 3.7.2 | Classification Tree Instability | 40 |

| | | |
|----------|--|-----|
| 4 | A More Formal Treatment of Classification and Forecasting | 43 |
| 4.1 | Introduction | 43 |
| 4.2 | Data Generation Models | 44 |
| 4.3 | Notation | 45 |
| 4.4 | Classification | 45 |
| 4.5 | Estimation in the Real World | 47 |
| 4.5.1 | Estimation Bias | 48 |
| 4.5.2 | Estimation Variance | 50 |
| 4.5.3 | The Bias-Variance Tradeoff | 51 |
| 4.5.4 | Uncertainty | 53 |
| 4.6 | A Bit More on the Joint Probability Model | 54 |
| 4.7 | Summary | 56 |
| 5 | Tree-Based Forecasting Methods | 59 |
| 5.1 | Introduction | 59 |
| 5.2 | Splitting the Data | 60 |
| 5.3 | Building the Costs of Classification Errors | 62 |
| 5.4 | Classification Tables | 63 |
| 5.5 | Ensembles of Trees: Random Forests | 64 |
| 5.5.1 | Variable Importance | 66 |
| 5.5.2 | Response Functions | 68 |
| 5.5.3 | Forecasting | 69 |
| 5.5.4 | Procedure Uncertainty | 70 |
| 5.5.5 | Forecast Uncertainty | 71 |
| 5.6 | Tree-Based Alternatives to Random Forests | 73 |
| 5.6.1 | Stochastic Gradient Boosting | 73 |
| 5.6.2 | Bayesian Trees | 75 |
| 5.7 | Why Do Ensembles of Classification Trees Work So Well? | 77 |
| 6 | Examples | 81 |
| 6.1 | A Simplified Example | 81 |
| 6.2 | A More Complex Example | 88 |
| 7 | Implementation | 101 |
| 7.1 | Making it Real | 101 |
| 7.2 | Demonstration Phase | 101 |
| 7.3 | Hardware and Software | 102 |
| 7.4 | Data Preparation for Forecasting | 103 |
| 7.5 | Personnel | 104 |
| 7.6 | Public Relations | 105 |
| 8 | Some Concluding Observations About Actuarial Justice and More | 107 |
| 8.1 | Current Trends in Actuarial Justice | 107 |
| 8.2 | Larger Issues | 109 |
| 8.2.1 | Forecasts Despite Individual Uniqueness | 109 |
| 8.2.2 | Risk and Forecasting | 110 |

| | | |
|-------|--------------------------------------|------------|
| 8.2.3 | Black-Box Methods | 110 |
| 8.3 | Dark Structure and New Science | 111 |
| 8.4 | A Closing Sermonette | 112 |
| | References | 113 |