

Data Mining Algorithms in C++

Data Patterns and Algorithms for Modern Applications

Discover hidden relationships among the variables in your data, and learn how to exploit these relationships. This book presents a collection of data-mining algorithms that are effective in a wide variety of prediction and classification applications. All algorithms include an intuitive explanation of operation, essential equations, references to more rigorous theory, and commented C++ source code.

Many of these techniques are recent developments, still not in widespread use. Others are standard algorithms given a fresh look. In every case, the focus is on practical applicability, with all code written in such a way that it can easily be included into any program. The Windows-based DATAMINE program lets you experiment with the techniques before incorporating them into your own work.

What You'll Learn

- Use Monte-Carlo permutation tests to provide statistically sound assessments of relationships present in your data
- Discover how combinatorially symmetric cross validation reveals whether your model has true power or has just learned noise by overfitting the data
- Work with feature weighting as regularized energy-based learning to rank variables according to their predictive power when there is too little data for traditional methods
- See how the eigenstructure of a dataset enables clustering of variables into groups that exist only within meaningful subspaces of the data
- Plot regions of the variable space where there is disagreement between marginal and actual densities, or where contribution to mutual information is high

Who This Book Is For

Anyone interested in discovering and exploiting relationships among variables. Although all code examples are written in C++, the algorithms are described in sufficient detail that they can easily be programmed in any language.

Discover hidden relationships among the variables in your data, and learn how to exploit these relationships. This book presents a collection of data-mining algorithms that are effective in a wide variety of prediction and classification applications. All algorithms include an intuitive explanation of operation, essential equations, references to more rigorous theory, and commented C++ source code.

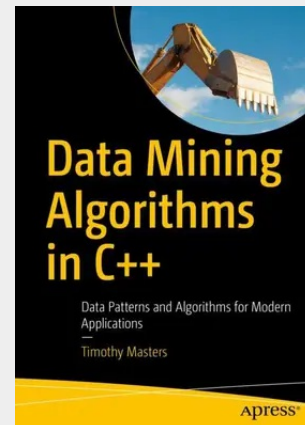
Many of these techniques are recent developments, still not in widespread use. Others are standard algorithms given a fresh look. In every case, the focus is on practical applicability, with all code written in such a way that it can easily be included into any program. The Windows-based DATAMINE program lets you experiment with the techniques before incorporating them into your own work.

What You'll Learn

Use Monte-Carlo permutation tests to provide statistically sound assessments of relationships present in your dataDiscover how combinatorially symmetric cross validation reveals whether your model has true power or has just learned noise by overfitting the dataWork with feature weighting as regularized energy-based learning to rank variables according to their predictive power when there is too little data for traditional methodsSee how the eigenstructure of a dataset enables clustering of variables into groups that exist only within meaningful subspaces of the dataPlot regions of the variable space where there is disagreement between marginal and actual densities, or where contribution to mutual information is high

Who This Book Is For

Anyone interested in discovering and exploiting relationships among variables. Although all code examples are written in C++, the algorithms are described in sufficient detail that



80,24 €

74,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9781484233146

Medium: Buch

ISBN: 978-1-4842-3314-6

Verlag: Apress

Erscheinungstermin: 19.12.2017

Sprache(n): Englisch

Auflage: 1. Auflage 2017

Produktform: Kartoniert

Gewicht: 576 g

Seiten: 286

Format (B x H): 178 x 254 mm

