Data Mining: Concepts and Techniques

3rd Edition

Solution Manual

Jiawei Han, Micheline Kamber, Jian Pei

The University of Illinois at Urbana-Champaign

Simon Fraser University

Version January 2, 2012

©Morgan Kaufmann, 2011

For Instructors' references only.

Do not copy! Do not distribute!

Preface

For a rapidly evolving field like data mining, it is difficult to compose "typical" exercises and even more difficult to work out "standard" answers. Some of the exercises in *Data Mining: Concepts and Techniques* are themselves good research topics that may lead to future Master or Ph.D. theses. Therefore, our solution manual is intended to be used as a guide in answering the exercises of the textbook. You are welcome to enrich this manual by suggesting additional interesting exercises and/or providing more thorough, or better alternative solutions.

While we have done our best to ensure the correctness of the solutions, it is possible that some typos or errors may exist. If you should notice any, please feel free to point them out by sending your suggestions to *hanj@cs.uiuc.edu*. We appreciate your suggestions.

To assist the teachers of this book to work out additional homework or exam questions, we have added one additional section "**Supplementary Exercises**" to each chapter of this manual. This section includes additional exercise questions and their suggested answers and thus may substantially enrich the value of this solution manual. Additional questions and answers will be incrementally added to this section, extracted from the assignments and exam questions of our own teaching. To this extent, our solution manual will be incrementally enriched and subsequently released in the future months and years.

Notes to the current release of the solution manual.

Due to the limited time, this release of the solution manual is a preliminary version. Many of the newly added exercises in the third edition have not provided the solutions yet. We apologize for the inconvenience. We will incrementally add answers to those questions in the next several months and release the new versions of updated solution manual in the subsequent months.

Acknowledgements

For each edition of this book, the solutions to the exercises were worked out by different groups of teach assistants and students. We sincerely express our thanks to all the teaching assistants and participating students who have worked with us to make and improve the solutions to the questions. In particular, for the first edition of the book, we would like to thanks Denis M. C. Chai, Meloney H.-Y. Chang, James W. Herdy, Jason W. Ma, Jiuhong Xu, Chunyan Yu, and Ying Zhou who took the class of *CMPT-459: Data Mining and Data Warehousing* at Simon Fraser University in the Fall semester of 2000 and contributed substantially to the solution manual of the first edition of this book. For those questions that also appear in the first edition, the answers in this current solution manual are largely based on those worked out in the preparation of the first edition.

For the solution manual of the second edition of the book, we would like to thank Ph.D. students and teaching assistants, Deng Cai and Hector Gonzalez, for the course *CS412: Introduction to Data Mining and Data Warehousing*, offered in the Fall semester of 2005 in the Department of Computer Science at the University of Illinois at Urbana-Champaign. They have helped prepare and compile the answers for the new exercises of the first seven chapters in our second edition. Moreover, our thanks go to several students from the *CS412* class in the Fall semester of 2005 and the *CS512: Data Mining: Principles and Algorithms* classes

in the Spring semester of 2006. Their answers to the class assignments have contributed to the advancement of this solution manual.

For the solution manual of the third edition of the book, we would like to thank Ph.D. students, Jialu Liu, Brandon Norick and Jingjing Wang, in the course *CS412: Introduction to Data Mining and Data Warehousing*, offered in the Fall semester of 2011 in the Department of Computer Science at the University of Illinois at Urbana-Champaign. They have helped checked the answers of the previous editions and did many modifications, and also prepared and compiled the answers for the new exercises in this edition. Moreover, our thanks go to teaching assistants, Xiao Yu, Lu An Tang, Xin Jin and Peixiang Zhao, from the *CS412* class and the *CS512: Data Mining: Principles and Algorithms* classes in the years of 2008–2011. Their answers to the class assignments have contributed to the advancement of this solution manual.