

Course Outline

Statistics 9850/1 B

Data Mining (9851) & Advanced Data Analysis (9850)

Professor A. I. McLeod, WSC Rm 235

aimcleod@uwo.ca

January 3, 2010

Location: WSC Rm 240, MWF, 12:30-1:30

Lectures run from Monday January 3, 2011 to Wednesday April 13, 2011.

Course Name: There are two course titles. This arose because I suggested that we “modernize” the title by changing the title to “Data Mining” but because OCGS approval is needed to change the title in our joint Biostatistics option, we need to keep the older title. As a result, students in Statistics will take Data Mining and students in Biostatistics will take Advanced Data Analysis. The courses will be identical apart from their titles.

Topics

The lectures will cover advanced statistical methods and other selected topics from the text. In addition, some other topics in data visualization and data analysis will be presented. Implementation of these methods using R software will be discussed in the labs and lectures as well.

<http://cran.r-project.org/>

Course Textbook

Hastie, Tibshirani and Friedman (2009). The Elements of Statistical Learning, 2nd Edition. This book is downloadable in pdf format:

<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

Computing

We will use R software. This is high-quality software and it is freely available. It is used by many large companies as well as being widely used at many universities and research institutes world wide. Articles about R have appeared in the business section of the New York Times as well as in the Wall Street Journal.

CRAN (Comprehensive R Archive): <http://cran.r-project.org>

For this course, minimal programming in R is all that is required. Reading the online manual “Introduction to R” is sufficient for the purpose of this course. This manual is available in PDF and HTML formats here:

<http://cran.r-project.org/manuals.html>