Stat 406
Algorithms for classification and prediction
Lecture 1
Introduction

Administrivia

- Class web page
  www.cs.ubc.ca/~murphyk/Teaching/Stat406-Spring10
- Join http://groups.google.ca/group/stat406-spring10
- Office hours: Fri 2-3pm, LSK 308d, or by appointment
- Tutorial/ discussion section: Wed 4-5 LSK 302
- TA: Pavel Krupski p.krupskii@stat.ubc.ca
- Midterm: Fri Mar 5th
- Last class: Wed Apr 14th
- Final exam: TBD

Grading

Grading
- Midterm (open-book):35%
- Final exam (open-book): 40%
- Weekly Assignments: 25%

Homeworks

Weekly homeworks, out on Mon, due back on Mon

- Collaboration policy:
  - You can collaborate on homeworks if you write the name of your collaborators on what you hand in; however, you must understand everything you write, and be able to do it on your own (eg. in the exam!)

- Sickness policy:
  - If you cannot do an assignment or an exam, you must come see me in person; a doctor's note (or equivalent) will be required.

Workload

- This class will be quite time consuming.
- Attending lectures: 3h.
- Attending discussion: 1h
- Weekly homeworks: about 4h.
- Weekly reading: about 5h.
- Total: 13h/week.

Pre-requisites

- You should know
  - Basic multivariate calculus e.g.,

  $$\nabla_{\mathbf{x}}\mathbf{x}^T\mathbf{a} = \mathbf{a}$$

  - Basic linear algebra e.g.,

  $$A\vec{u}_i = \lambda_i \vec{u}_i$$

  - Basic probability/ statistics e.g.

  $$Cov(X,Y) = E\left[(X - EX)(Y - EY)\right] = E[XY] - E[X]E[Y]$$

  - Some programming concepts (eg from R)

Textbook

- "Machine learning: a probabilistic approach"
- Draft copies available from Copiesmart in the UBC Village (next to Macdonald's) for about $56.50
- Extra credit (up to 5% of your grade) for finding errors (5 points) or typos (1 point) – more details later
- Please bring your book to every class (and the exams).
- Ch 1 is online for free --- read it, so you can see if you still want to take this class or drop it!

Other good books

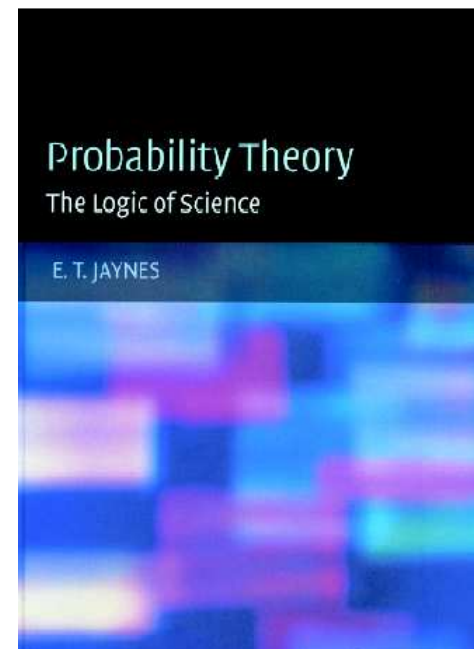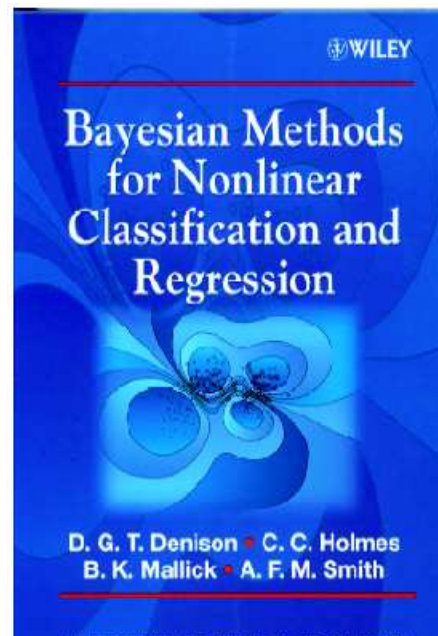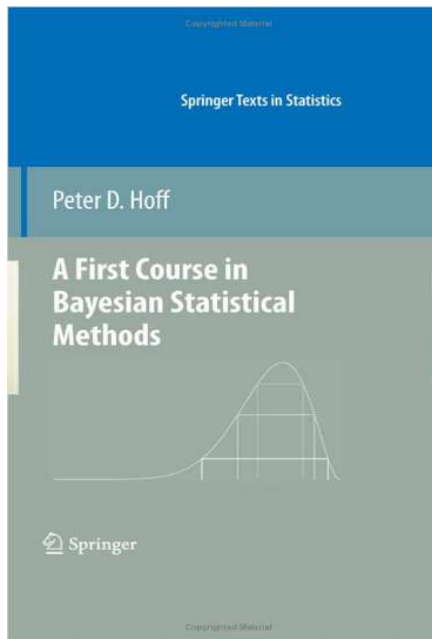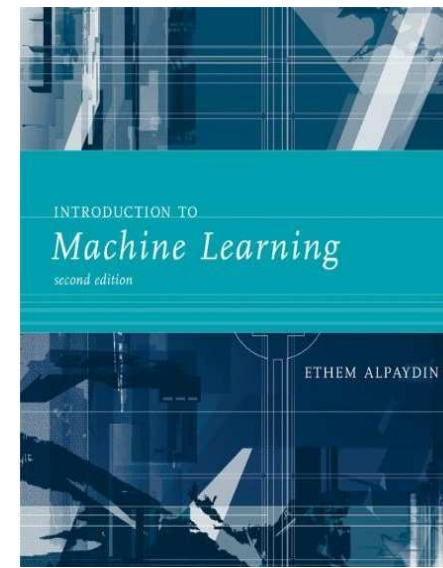If you want a book that is already "debugged", see one of these
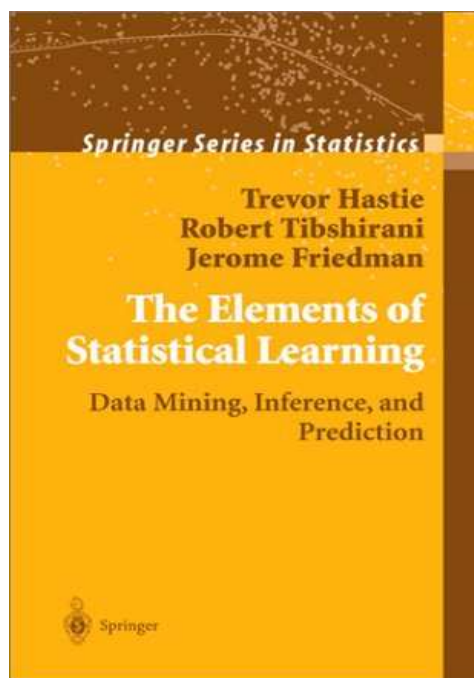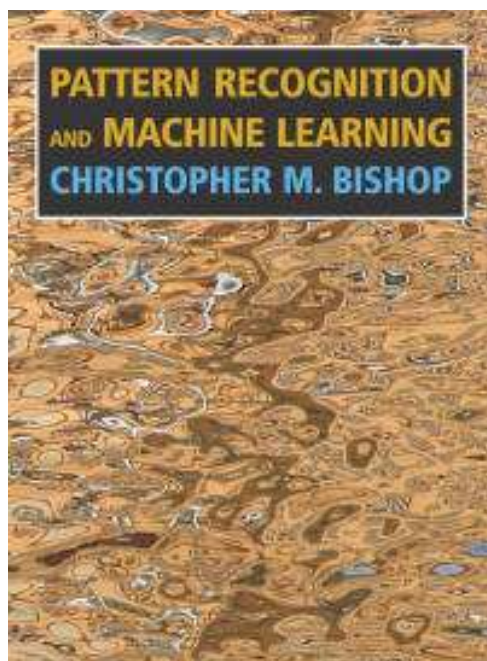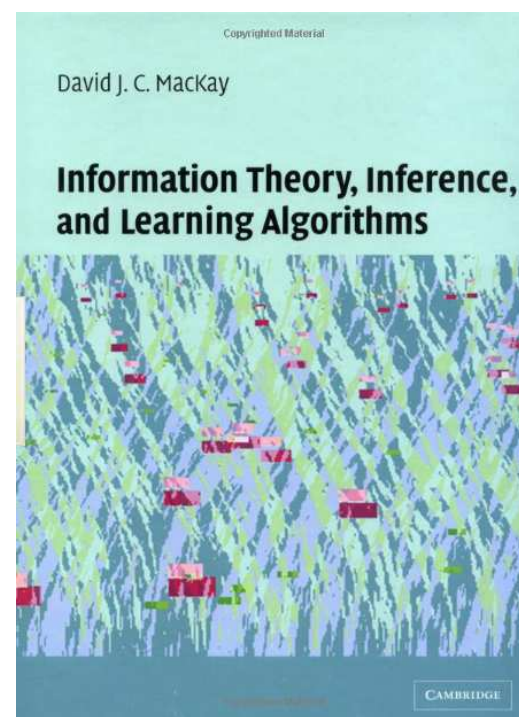
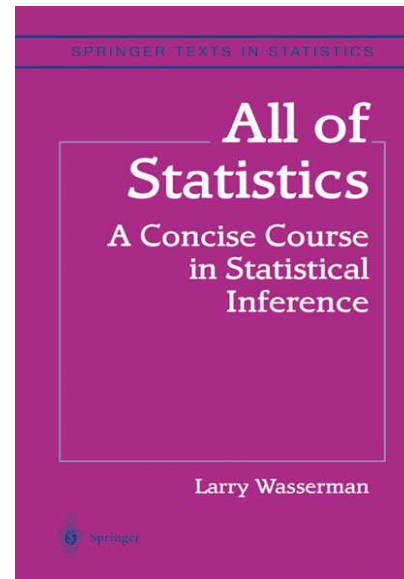

Free online   Free online

If you plan to "major" in machine learning, you should buy and read all of these!

Other good books

Stuart Russell, Peter Norvig — Artificial Intelligence: A Modern Approach, Third Edition

SPRINGER TEXTS IN STATISTICS — All of Statistics: A Concise Course in Statistical Inference — Larry Wasserman — Springer

INTRODUCTION TO Machine Learning — second edition — ETHEM ALPAYDIN

Springer Texts in Statistics — Peter D. Hoff — A First Course in Bayesian Statistical Methods — Springer

WILEY — Bayesian Methods for Nonlinear Classification and Regression — D. G. T. Denison · C. C. Holmes · B. K. Mallick · A. F. M. Smith

Probability Theory — The Logic of Science — E. T. JAYNES

Matlab

- Matlab is a mathematical scripting language widely used for machine learning (and engineering and numerical computation in general).
- Everyone should have access to Matlab. If not, ask for a Stats guest account.
- You can buy a student version for $170 from the UBC bookstore. Please make sure it has the Stats toolbox.
- Matt Dunham has written an excellent Matlab tutorial which is on the class web site – please study it carefully!
- On Wed 6th, 4-5pm, LSK 302, I will offer a brief Matlab tutorial.

Software

- My textbook (MLAPA) is accompanied by Matlab code, which will be made available via the course web page

- Currently there is not 100% consistency between the book and the code; this will be fixed during the semester

- I also have 2 older code projects associated with the book (PMTK and PMTK2) which you can ignore, since they are deprecated

- We will also use several other Matlab packages such as Netlab, GPML, etc.

Learning objectives

- By the end of this class, you should be able to
- 1. For any given data analysis problem, be able to identify a suitable probabilistic model from those that are currently available, or if none exists, to devise one of your own.
- 2. If the model is novel, be able to derive a suitable algorithm. This includes deriving the update equations for fitting the parameters (eg gradient expressions for the MLE/MAP
- 3. If the model is novel, be able to implement the above equations in clear and efficient code (preferably Matlab or R).
- 4. Be able to evaluate the performance of your model/ algorithm in an objective way, and compare to other models/ algorithms.

Lecture style

- This year, rather than spend time making lecture slides, I will just project my book onto the screen, and use the blackboard.

- Each "lecture" then becomes just an executive summary of relevant parts of the book.

- I will try to include some in-class demos and/or interactive activities (in the spirit of the Carl Weiman initiative)

- You are expected to carefully read the book either before and/or after each lecture to learn all the details – the relevant sections will be listed on the class web page.

- Homeworks are an essential aid to learning the material; assessing performance is a relatively minor concern