

# STA 414/2104: Statistical Methods for Machine Learning and Data Mining (Jan-Apr 2014)

## Instructor:

Radford Neal, Office: SS6026A, Phone: (416) 978-4970, Email: radford@stat.utoronto.ca

Office Hours: Tuesdays 3:10-4:00pm, in SS6026A.

## Lectures:

Tuesdays 12:10-2:00pm and Thursdays 12:10-1:00pm, in SS 2127. The first lecture is January 7. The last lecture is April 3. There are no lectures February 17 to 21 (Reading Week).

Graduate students in STA 2104 will make presentations the week of April 7 to 11.

## Evaluation:

For undergraduates in STA 414:

52% Three or four assignments.  
48% Three 50-minute tests, each worth 16%, held in lecture time on February 6, March 13, and April 3.

For graduate students in STA 2104:

48% Three or four assignments.  
42% Three 50-minute tests, each worth 14%, held in lecture time on February 6, March 13, and April 3.  
10% A 12-minute individual presentation on a conference paper that you have read.

The assignments are to be done by each student individually. Any discussion of the assignments with other students should be about general issues only, and should **not** involve giving or receiving written, typed, or emailed notes.

Graduate students may discuss the conference paper that they will present with anyone, in order to help understand it, but they must

prepare their presentation themselves. (They may if they wish solicit feedback from others after a practice run of their presentation.)

### **Textbook:**

The book *Machine Learning: A Probabilistic Perspective*, by Kevin P. Murphy, is strongly recommended (but not required).

I will be posting lecture slides, and links to online references.

### **Computing:**

Assignments will be done in R. Statistics Graduate students will use the Statistics research computing system. Undergraduates and graduate students from other departments will use CQUEST. You can request an account on CQUEST if you're an undergraduate student in this course (you need to fill out a form if you're a grad student).

You can also use R on your home computer by downloading it for free from [www.R-project.org](http://www.R-project.org). From that site, here is the Introduction to R.

You might also be interested in trying out my faster implementation of R, called pqR, available from [pqR-project.org](http://pqR-project.org) (However, it currently is distributed only in source form, and is easily installed only on Linux/Unix systems.)

### **Some useful on-line references**

*Information Theory, Inference, and Learning Algorithms*, by David MacKay.

David MacKay's thesis.

*The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition)*, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

*Gaussian Processes for Machine Learning*, by Carl Edward Rasmussen and Christopher K. I. Williams.

Proceedings of the International Conference on Machine Learning (ICML)

Proceedings of the annual conference on Neural Information Processing Systems (NIPS).

## **Lecture slides:**

Note that slides may be updated as mistakes are corrected, or the amount of material covered in the week becomes apparent.

Week 1 (Introduction: Murphy's Ch. 1)

## **Web pages for past related courses:**

STA 414/2104 (Spring 2013)

STA 414/2104 (Spring 2012)

STA 414/2104 (Spring 2011)

STA 414/2104 (Spring 2007)

STA 414/2104 (Spring 2006)

CSC 411 (Fall 2006)

STA 410/2102 (Spring 2004) - has many examples of R programs