10-601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Course Introduction

Matt Gormley
Lecture 1
Aug. 26, 2019

SYLLABUS HIGHLIGHTS

Syllabus Highlights

The syllabus is located on the course webpage:

http://www.cs.cmu.edu/~mgormley/courses/10601

The **course policies** are **required** reading.

Syllabus Highlights

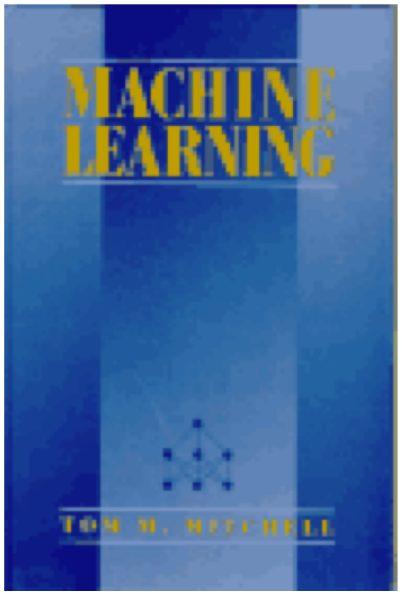
- **Grading:** 40% homework, 15% exam 1, 15% exam 2, 25% final exam, 5% participation
- **Midterm Exam 1:** evening exam, Thu, Oct. 03
- **Midterm Exam 2:** evening exam, Thu, Nov. 14
- **Final Exam:** final exam week, date TBD
- **Homework:** ~3 written and ~5 programming
  - 6 grace days for homework assignments
  - Late submissions: 80% day 1, 60% day 2, 40% day 3, 20% day 4
  - No submissions accepted after 4 days w/o extension
  - Extension requests: see syllabus
- **Recitations:** Fridays (optional, interactive sessions)
- **Readings:** required, online PDFs, recommended for after lecture
- **Technologies:** Piazza (discussion), Autolab (programming), Canvas (quiz-style), Gradescope (open-ended)
- **Academic Integrity:**
  - Collaboration encouraged, but must be documented
  - Solutions must always be written independently
  - No re-use of found code / past assignments
  - Severe penalties (i.e. failure)
- **Office Hours:** posted on Google Calendar on "People" page

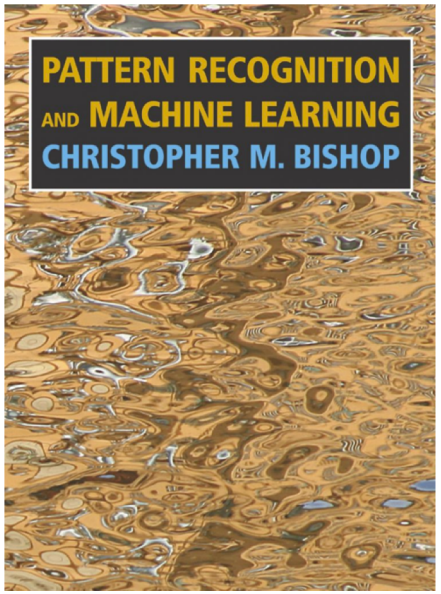
Lectures

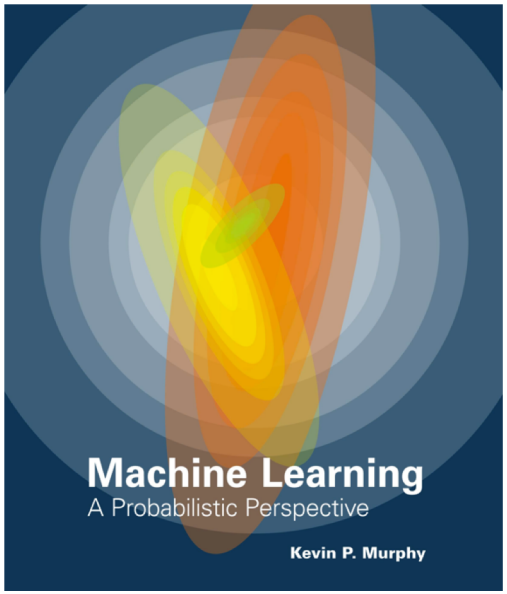
- You should ask lots of questions
  - Interrupting (by raising a hand) to ask your question is strongly encouraged
  - Asking questions later (or in real time) on Piazza is also great
- When I ask a question…
  - I want you to answer
  - Even if you don't answer, think it through as though I'm about to call on you
- Interaction improves learning (both in-class and at my office hours)

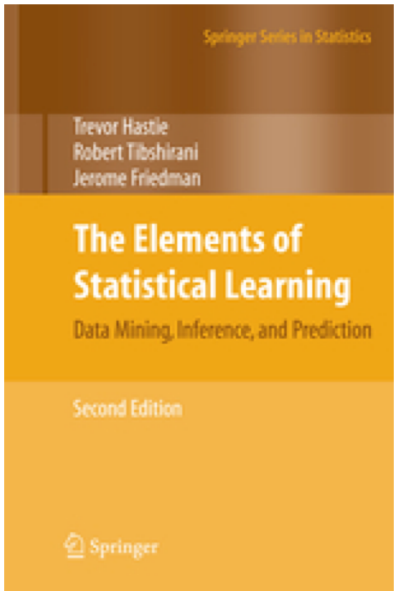
Textbooks

You are not *required* to read a textbook, but it will help immensely!







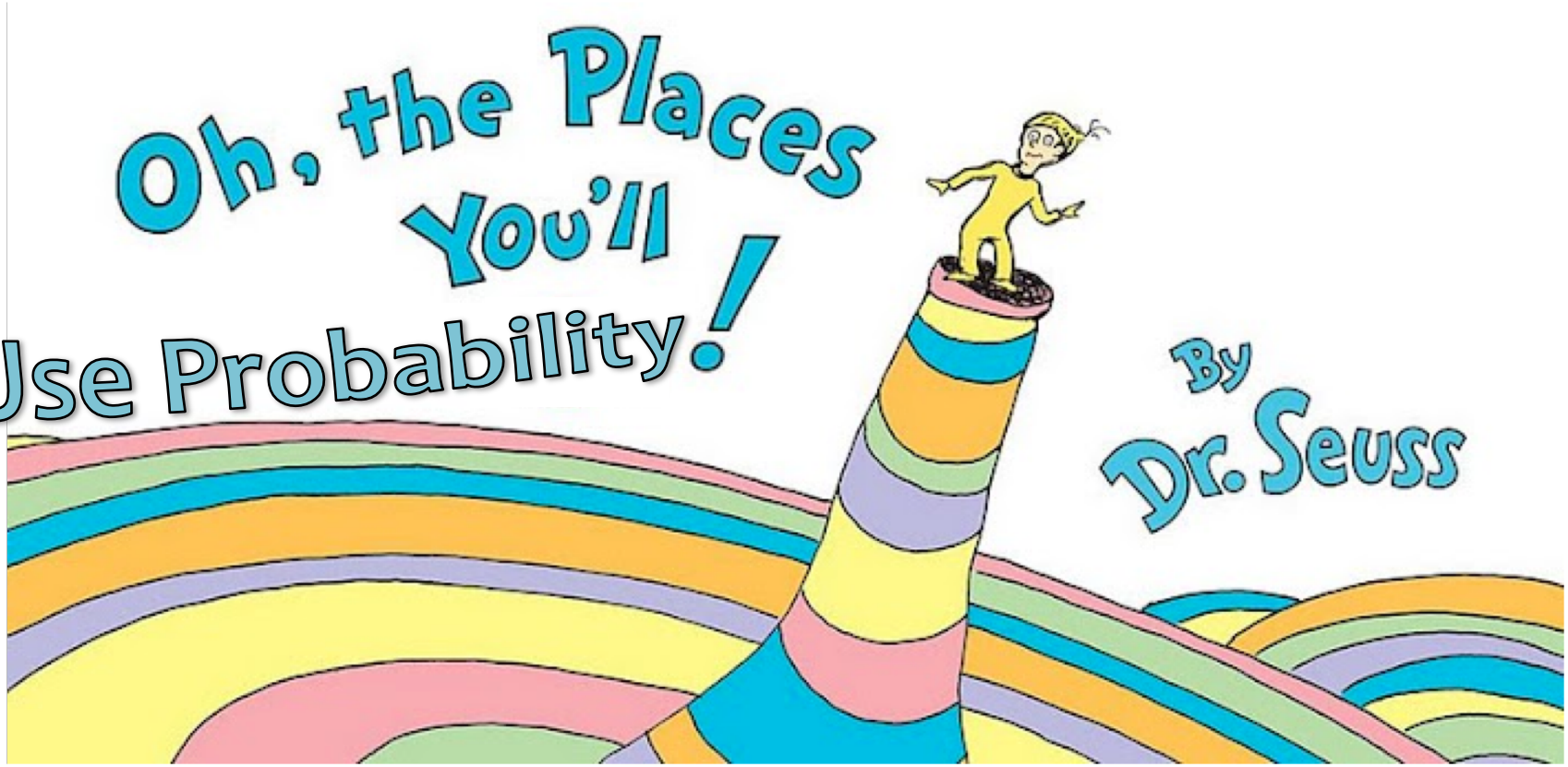


PREREQUISITES

Prerequisites

What they are:

- Significant programming experience (15-122)
  - Written programs of 100s of lines of code
  - Comfortable learning a new language
- Probability and statistics (36-217, 36-225, etc.)
- Mathematical maturity: discrete mathematics (21-127, 15-151), linear algebra, and calculus



Use Probability!

Oh, the Places You'll Use Probability!

Supervised Classification

- Naïve Bayes

$$p(y|x_1, x_2, \ldots, x_n) = \frac{1}{Z} p(y) \prod_{i=1}^{n} p(x_i|y)$$

- Logistic regression

$$P(Y = y|X = x; \boldsymbol{\theta}) = p(y|x; \boldsymbol{\theta})$$
$$= \frac{\exp(\boldsymbol{\theta}_y \cdot \mathbf{f}(x))}{\sum_{y'} \exp(\boldsymbol{\theta}_{y'} \cdot \mathbf{f}(x)}$$

Note: This is just motivation – we'll cover these topics later!

Oh, the Places You'll Use Probability!

ML Theory

(Example: Sample Complexity)

- Goal: h has small error over D.

  True error: $err_D(h) = \Pr_{x \sim D}(h(x) \neq c^*(x))$

  How often $h(x) \neq c^*(x)$ over future instances drawn at random from D

- But, can only measure:

  Training error: $err_S(h) = \frac{1}{m}\sum_i I\big(h(x_i) \neq c^*(x_i)\big)$

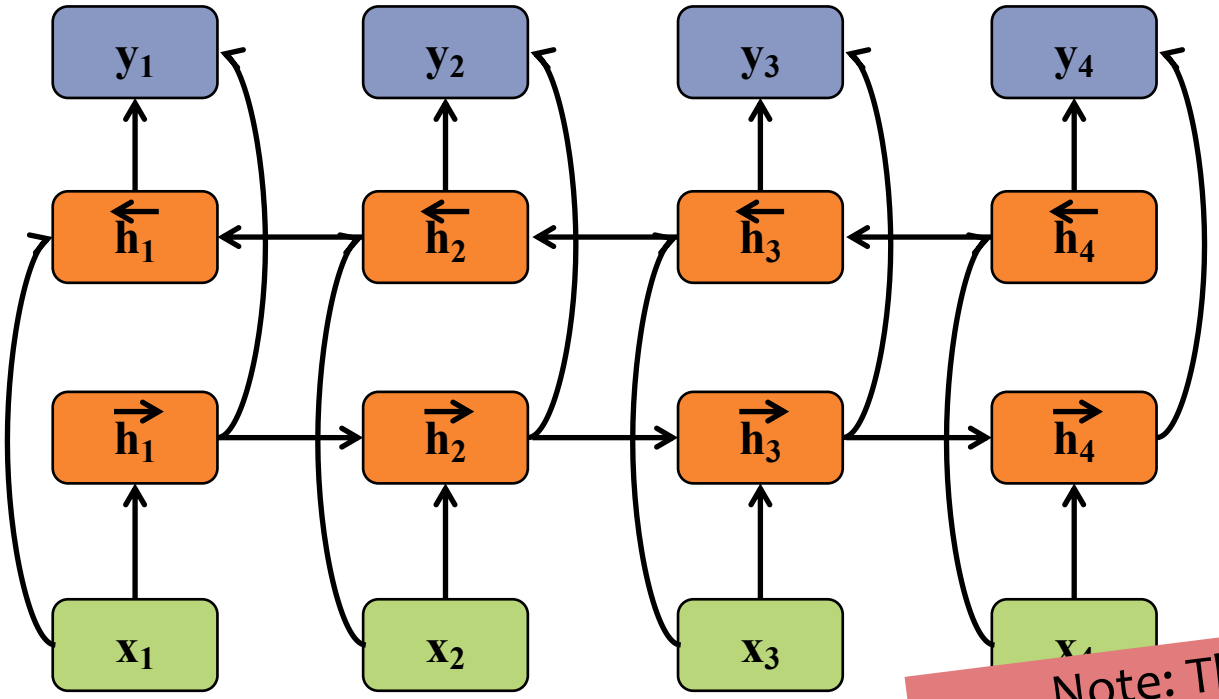
  How often $h(x) \neq c^*(x)$ over training instances

Sample complexity: bound $err_D(h)$ in terms of $err_S(h)$

Note: This is just motivation – we'll cover these topics later!

Oh, the Places You'll Use Probability!

Deep Learning
(Example: Deep Bi-directional RNN)

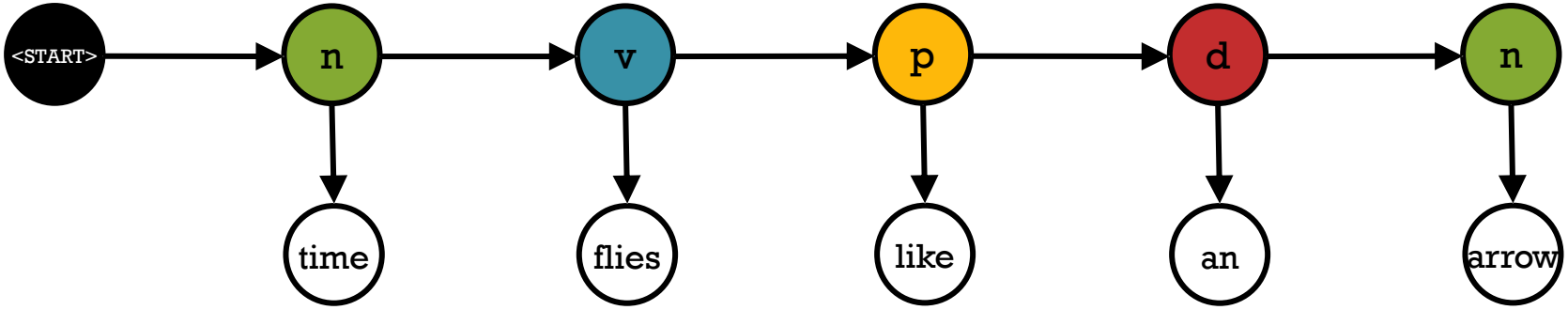


Note: This is just motivation – we'll cover these topics later!

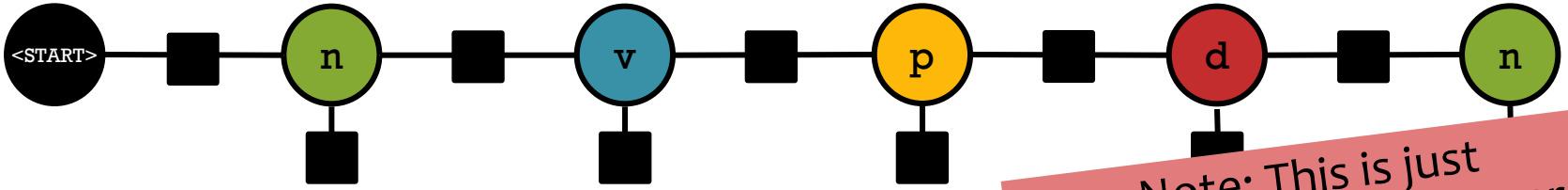
Oh, the Places You'll Use Probability!

Graphical Models

- Hidden Markov Model (HMM)



- Conditional Random Field (CRF)



Note: This is just motivation – we'll cover these topics later!

Prerequisites

What if I'm not sure whether I meet them?

- Don't worry: we're not sure either
- However, we've designed a way to assess your background knowledge so that you know what to study!

  (see instructions of Canvas portion of HW1)

Reminders

- **Homework 1: Background**
  - **Out: Wed, Aug. 28 (2nd lecture)**
  - **Due: Wed, Sep. 04 at 11:59pm**
  - Two parts:
    1. written part to Gradescope,
    2. programming part to Autolab
  - unique policy for this assignment:
    1. **two submissions** for written (see writeup for details)
    2. **unlimited submissions** for programming (i.e. keep submitting until you get 100%),

Q&A