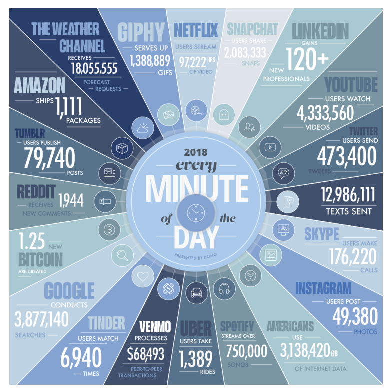
CS 5135/6035 Learning Probabilistic Models
Lecture 1: Course Overview

Gowtham Atluri

August 25, 2018

A Gentle Introduction

Data Deluge



2018 Internet Data estimates[1]

[1] https://www.domo.com/blog/data-never-sleeps-6/

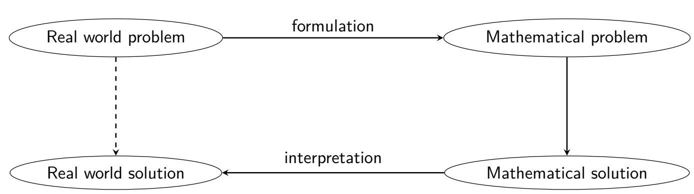
Data Deluge

- Data is being collected at an unprecedented rate
  - More data is being collected every year than ever collected
- **Science** - large scale experiments in astronomy, high energy physics, next generation genomics datasets, climate data, etc.
- **Business** - e-commerce, online advertising, electronic trading, self-driving cars, etc.
- **Society** - Government data, social media, mobile health, public health, crime data, etc.

We seek to harness this data and help discover actionable insights to accelerate science, advance businesses, and address societal problems.

Learning Probabilistic **Models**

- Mathematical modelling is a process of representing real world problems in mathematical terms in an attempt to find solutions to the problems.
  - Model is typically as a simplification or abstraction of a (complex) real world problem



Learning Probabilistic **Models**

Types of Models

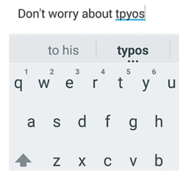
Deterministic Models

- Output is fully determined by parameter values
- Examples:
  - current through a conductor $I = \frac{V}{R}$
  - area of a circular lake: $A = \pi r^2$
  - predicting the price of a house based on relevant factors: $y = \beta x$

Probabilistic Models

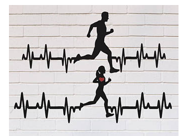
- Posses inherent randomness
- Output can be a prediction that is different even for same parameter values
- Output can be a probability of occurrence of an event
  - E.g., Weather forecasting models: what is the probability it will rain today?

Applications where deterministic models are ineffective



Spelling corrections

Speech-to-text conversion



Activity detection in smart watches

Movie recommendations

Learning **Probabilistic** Models

> "As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."
>
> — Albert Einstein

- Source of uncertainty:
  - **incomplete/noisy data**
    - not all data can be collected
  - **incomplete knowledge**
    - not all functions of a gene are known
  - **inherent randomness**
- Probability theory is a mathematical language for **representing and manupulating uncertainty**.

The inevitable reconciliation of **Fortuna** (goddess of chance) and **Sapientia** (wisdom incarnate). 16th century wood engraving.

Learning **Probabilistic** Models

- Probability theory is a mathematical language for **representing and manupulating uncertainty**.

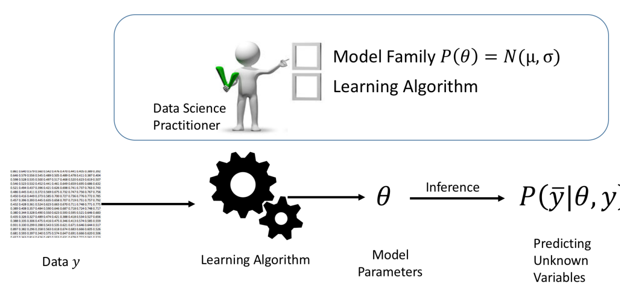
Advantages of probability models

- They are conceptually simple
  - Probability distributions are used to represent all uncertain unobserved quantities in a model and how they relate to the data.
- Support hierarchical construction
  - Simple probabilitic models of one or a few variables can be used to construct larger, more complex models.
- Easier to understand even complex models
  - The compositionality of probabilistic models makes it much easier to understand the models.

Learning Probabilistic Models

- Given data from a real-world phenomenon
  1. How do we choose a suitable family of models to learn?
  2. How do we learn a probabilistic model?
  3. How do we esure that the learned model fits the data well?
  4. How do we use it for predicting unknown variables?
- Keywords: Learning, Evaluation, and Inference
  - **Learning:** Given a set of samples that are known/assumed to be generated from a model, the goal is to determine the parameters of the model.
  - **Evaluation:** Given the samples and the learned model, the goal is to determine how well the model fits the data.
  - **Inference:** Given a set of model parameters and an observation of some variable(s), the goal is to predict states of other variables.
- Inference is also referred to as *probabilistic reasoning*.

Learning Probabilistic Models



What is in it for Computer Scientists?

- We (in CS) seek to develop systems that can automatically collect necessary data, make decisions and complete tasks.
- Examples:
  - Autonomous vehicles
  - Automated diagnosis of cancers
  - Translating sentences from one language to another
- Computational algorithms are an integral part of probabilistic models that offer effective solutions in these applications.
- We are interested in **learning algorithms** and **inference algorithms**
- To fully appreciate the utility and effectiveness of these algorithms, we need to understand the principles in probability and statistics.

Course Overview

Prerequisites

Courses

This course will build on some foundational concepts from:

- Basic probability and statistics
- Calculus
- Analysis of algorithms
- Discrete mathematics/ Graph Theory

We will do a quick review some of the basic concepts early-on.

Programming Language

- We will use Julia for programming exercises.
  - You are not expected to know Julia before hand.
  - Familiarity with Matlab will be useful.

Course contents

- Probability foundations
  - Discrete probability distributions
  - Continuous probability distributions
  - Probabilistic Inference
  - Parameter Estimation
- Maximum likelihood estimation
  - Univariate, Multivariate, Logistic Regression
  - Mixture Models, Factor Analysis
- Bayesian approach
  - Single parameter models
  - Multi-parameter models
  - Priors: Default, Conjugacy, Jeffreys, Mixture
  - Hierarchical models
  - Model Selection
- Bayesian computation
  - Random Sampling
  - Monte Carlo Integration
  - Markov chain Monte Carlo Methods

Module 1: Probability Foundations

- Random Variables, Domain, Distribution
- Axioms, Principles
  - Conditional Probability, Bayes' Rule
  - Independence, Marginalization, etc.
- Standard Probability Distributions
  - Discrete
  - Continuous
- Multivariate Probability Distributions
- Probabilistic Reasoning
- Parameter Estimation
  - Max. Likelihood Estimation
  - Bayesian Estimation
- Properties of Estimators

Module 2: Maximum Likelihood Estimation

Topics

- General approach to MLE
  - I.I.D
  - Likelihood $\mathcal{L}(\theta|x)$, Log-Likelihood ℓ, Maximizing ℓ
  - Optimization algos: Gradient Descent/Newton Method
- Univariate Parameter Est. using MLE
- Multivariate Parameter Est. using MLE
- Logistic Regression
  - Max. Conditional Likelihood
- Latent variables
  - Mixture Models: Discrete latent vars.
  - Factor Models: Continuous latent vars.
- Expectation-Maximization
  - General Approach
  - Proof of correctness

Module 3: Bayesian Parameter Estimation

- General approach to Bayesian estimation
  - Prior, Likelihood, Posterior
  - Why/Why not Bayesian estimation?
- Priors
  - Noninformative
  - Conjugate Priors
  - Natural Conjugacy
  - Mixture of Priors
  - Jeffrey's Prior
- Posterior
  - Univariate
  - Multivariate: Nuisance Parameter, Marginal Posterior
- Summarization of Posterior
  - Point Estimation (Bayes' Risk)
  - Interval Estimation

Module 4: Bayesian Computation

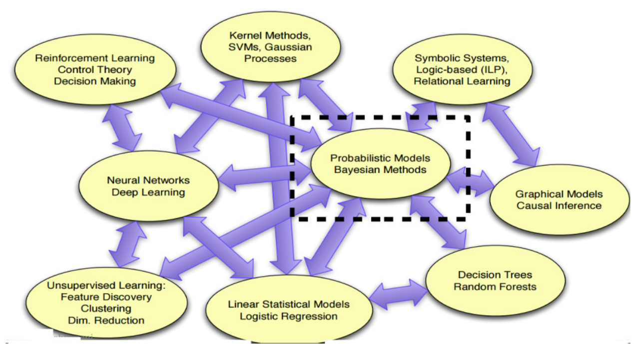
- Sampling from Posterior
  - Pseudo random number generator
  - Inverse-Transform Method
  - Accept-Reject Method
- Monte Carlo Integration
  - General Approach
  - Importance Sampling
- Markov Chain Monte Carlo Methods
  - Markov Chain: Stationarity and other properties
  - Metropolis-Hastings
    - General Approach
    - Random-walk Metropolis-Hastings
    - Independent Metropolis-Hastings
  - Gibbs Sampling
    - Application: Hierarchical Models

Learning Outcomes

Upon successful completion of this course, you will be able to...

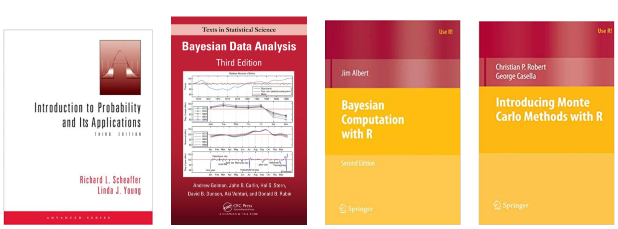
- determine the suitabile probabilistic models for different problem settings
- implement algorithms for learning models based on existing frameworks
- be able to use inference algorithms make inferences in real-world applications
- read previous and existing research literature in this area and critique about the strengths and weaknesses

This course in the larger context



Reading Material

- Mostly tutorials, papers, and lecture notes
- Relevant chapters from the textbooks



Grading

Score breakdown:

- Quizzes: 5%
- In-class exercises = 25%
- Midterm exams (2 x 20%): 40%
- Final: 30%

Grading scale:

- A: 91-100
- A-: 86-90
- B+: 81-85
- B: 76-80
- B-: 71-75
- C+: 66-70
- C: 61-65

Teaching Methodology

- This class will be taught in a flipped classroom style
  - The lectures videos will be made available prior to the class.
  - These videos will be typically 50 mins/class.
  - Students are expected to watch the lectures prior to participating in the class.
  - There will be a quiz at the beginning of each class.
  - Students will be working in groups on practice questions during the class.
    - Please bring your computer to every class.

Flipped Classroom

A flipped classroom is an instructional strategy and a type of blended learning that reverses the traditional learning environment by delivering instructional content, often online, outside of the classroom.[a]

[a] https://en.wikipedia.org/wiki/Flipped_classroom

Programming

- We will be using julia
- What is it?
  - a new programming language for scientific computing
    - developed by a group mostly from MIT
    - fully open source, i.e., free
    - convenient syntax for building math constructs like vectors, matrices, etc.
    - largely similar to Matlab
- Some pointers:
  - Tutorial on Julia Syntax [https://github.com/madeleineudell/intro-to-julia/blob/master/Julia%20Syntax%20Tutorial.ipynb]
  - Julia Documentation [https://docs.julialang.org/en/v0.6.0/]
- We will be using Julia 0.6.4 [https://julialang.org/downloads/oldreleases.html]
  - Even though Julia 1.0 was released, many packages we will need are yet to be updated.

Some guidelines for success in this course

- Assess early-on if the course content is aligned with your interest
- Assess if you have the necessary background
  - MATH 4008 - Introduction to Probability
  - Math 2063 - Multivariable Calculus
  - Math 2076 - Linear Algebra
  - Programming experience in Matlab/R/Python
- Active participation in the classroom is strongly encouraged.
- If you have difficulty following the course, speak to the instructor.
- Use discussion boards to ask questions. Participate in discussion boards for exam preparation.
- Follow student code of conduct [https://www.uc.edu/conduct/Code_of_Conduct.html].
  - No Plagiarism!

Previous students about this course

"Flipped education: **Best way of learning so far in my educational life**. This would have been **extremely difficult** to follow if this style was not adapted."

"I think the **material and videos were perfect**."

"The **exercises have been really great tools to better comprehend the material**. I like how things are structured with background and explanation of method, equations behind the method, Julia implementation and then hands-on implementation."

How important is this probabilistic models course for your overall training in the area of data science? **9.4/10**

Suggestions for YOU from previous students

"Be sure to watch the lectures, **make sure you take the time to understand the mathematics**, and **you will be better prepared for Machine Learning and Data Analytics than any other course at UC**. I know because I've taken them all."

"Watch all the lectures with **extra attention to details**. Make note of every word that the professor says and keep up with the pace of the subject. **LPM will be beautiful**."

"**Do the homework assignments diligently** and **take printed notes of the each lecture** and **make your own notes** in order to grasp the concepts better from the video lecture."

Previous students on what they wish they'd known

"I wish I would have known that **I needed to reserve a fairly unmovable two-hour block** on Monday and Wednesday evening **to watch the lectures**."

"... this should be heavily stressed in the first class... **if you are late you can't take the quiz**"

"Make sure your calculus and statistics foundations are good. **If you do not like math-centric courses do not take this course.**"

Probability and its history

What is probability?

Definition:

- the chance that a given event will occur [2]
- the ratio of the number of outcomes in an exhaustive set of **equally likely** outcomes that produce a given event to the total number of possible outcomes.

- Scenario: Tossing a coin
- A **trial** or **experiment** is one toss of a coin
- Possible outcomes: {Heads, Tails}
- Favourable outcome/event: Heads
- Probability: $\frac{1}{2}$

[2] http://www.m-w.com/dictionary/probability

History of probability - I

The Pascal - Fermat correspondence of 1654

- Cited by historians as the origin of probability theory.
- **The problem of points** - posed by a gambler, Chevalier De Mere.
- **Game:** Two players of equal skill play a game with an ultimate monetary prize. The first to win a fixed number of rounds wins the prize.
- **Scenario:** How should the stakes be divided if the game is interrupted after several rounds, but before either player has won the required number?
- There is uncertainty in how the game will unfold.

History of probability - II

- **Existing solutions of the time:** Split the prize in the ratio of points already scored. *Credit for being ahead in the game.*

- Assume player A needs to win a rounds and player B needs b rounds.The game can go at most $a + b - 1$ rounds.
- Number of all possible scenarios 2^{a+b-1}.

- **Pascal and Fermat's solution:** The fair division of the stake will be the proportion of the scenarios that lead to a win by A versus the proportion that lead to a win by B.
  - Takes into account the 'chance' of winning the game.

History of probability - different contexts - I

Other problems from the mid-17th century that required dealing with *uncertainty*

- Annuities
  - Two parties A and B agree that A pays B a lump sum, while B pays back A in annual installments for n years.
  - In the case of life annuity, B pays A a set sum every year until death.
  - To receive \$100 annually from B, how much should B pay A (assuming 5% interest)? Depends on B's life expectancy.

History of probability - different contexts - II

Other problems from the mid-17th century that required dealing with *uncertainty*

- Legal system
  - A judge is a *'trier of fact'* i.e., to determine fact based on presented evidence.
  - How much and what kind of evidence was required to produce a 'degree' of conviction in the mind of the judge?
  - Leibniz proposed to calculate the probability of statements of fact in order to determine whether they were true.
    - Belief proportioned to evidence.
  - The statements with the highest probability score would be judged to be true.

History of probability - different contexts - III

Other problems from the mid-17th century that required dealing with *uncertainty*

- Plague in London
  - John Graunt, in 1661, computed a *'life table'* that assigns probability of survival to each age using records of birth and death from Church of England.

Table 1. Graunt's Life Table.

Age Interval	Prop. Deaths in Interval	Prop. Surviving til start of Interval
0-6	0.36	1.00
7-16	0.24	0.64
17-26	0.15	0.40
27-36	0.09	0.25
37-46	0.06	0.16
47-56	0.04	0.10
57-66	0.03	0.06
67-76	0.02	0.03
77-86	0.01	0.01

History of probability - different contexts - III

Other problems from the mid-17th century that required dealing with *uncertainty*

- Plague in London
  - John Graunt, in 1661, computed a *'life table'* that assigns probability of survival to each age using records of birth and death from Church of England.
  - He made a great discovery concerning the prevalence of the plague. A 50 year old person had about as much chance of dying in the next year as did a 20 year old. Cause of death is not age related, very likely plague related.
  - The systematic study of quantitative facts about the state – births, deaths, incidence of diseases, emigration, etc – is referred to as political arithmetic.
  - This 17th century political arithmetic was first referred to as **'State-istics'** in late 18th century that is now popularly referred to as **'Statistics'**.

History of probability - books