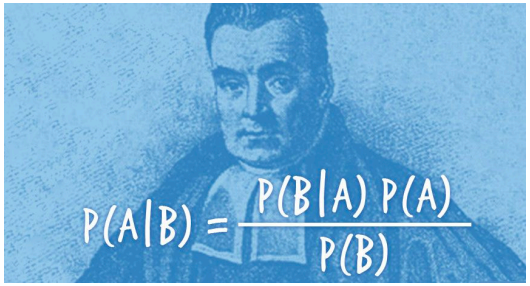
CS5350/6350 Machine Learning

Fall 2021

Instructor: Shandian Zhe



Shandian Zhe: Probabilistic Machine Learning

zhe@cs.utah.edu

Assistant Professor, School of Computing, University of Utah

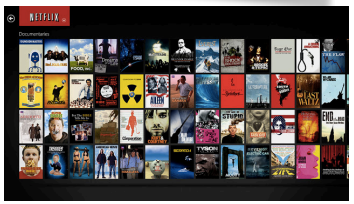
Research Topics:

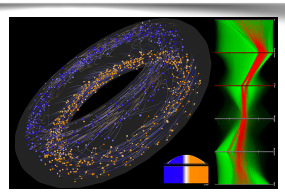
1. Bayesian Nonparametrics
2. Bayesian Deep Learning
3. Probabilistic Graphical Models
4. Large-Scale Learning System
5. Tensor/Matrix Factorization
6. Embedding Learning

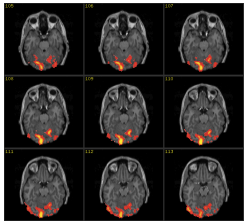
Applications:

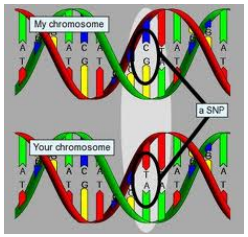
- Collaborative Filtering
- Online Advertising
- Physical Simulation
- Brain Imaging Data Analysis

....









Outline

- Machine learning definition, applications and course content
- Course requirements/policies (homework assignments, projects, final exams, etc.)
- Basic knowledge review (random variables, mean, variance, independency, etc.)

What is (machine) learning?

Let’s play a game

The badges game

Attendees of the 1994 conference on Computational Learning Theory received conference badges labeled + or —

Only one person (Haym Hirsh) knew the function that generated the labels

Depended *only* on the attendee's name

The task for the attendees: Look at as many examples as you want in the conference and find the unknown function

Let’s play

Name	Label
Claire Cardie	−
Peter Bartlett	+
Eric Baum	−
Haym Hirsh	−
Shai Ben-David	−
Michael I. Jordan	+

How were the labels generated?

What is the label for my name? Yours?

Playing the badge game → a typical learning procedure

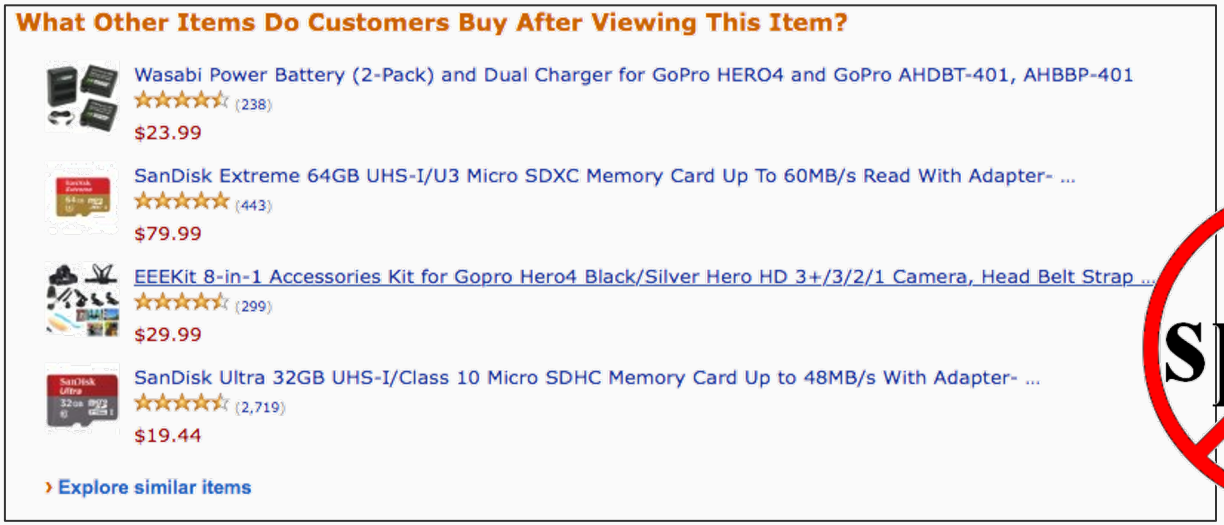
If the players are machines →it is a machine learning procedure!

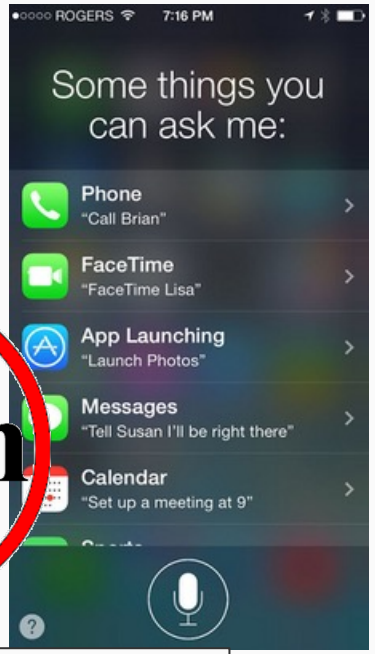
Alpha-Go!

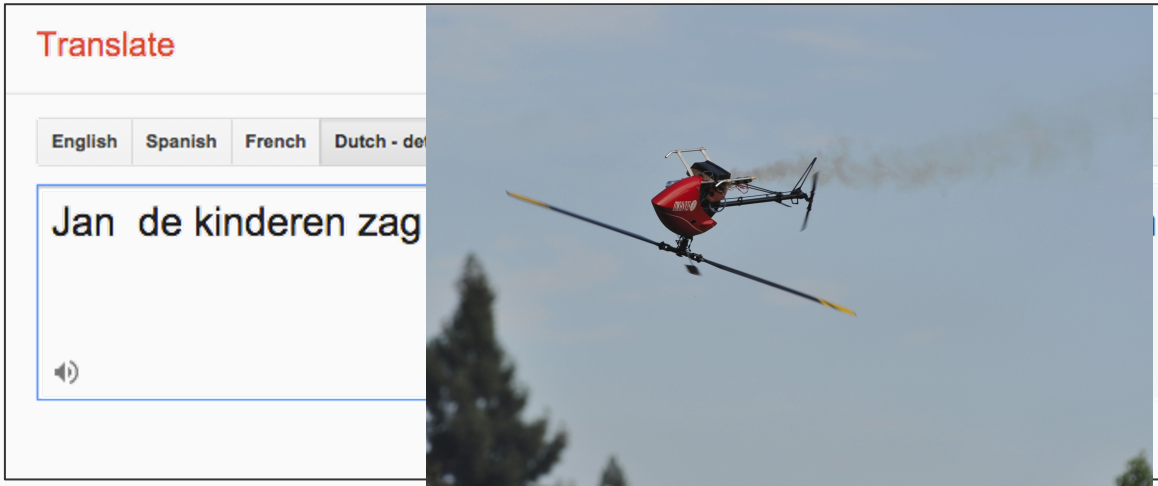
A ML algorithm rather AI

Machine learning is everywhere!

And you are probably already using it







Machine learning is everywhere!

And you are probably already using it

- Is an email spam?
- Find all the people in this photo
- If I like these three movies, what should I watch next?
- Based on your purchase history, you might be interested in…
- Will a stock price go up or down tomorrow? By how much?
- Handwriting recognition
- What are the best ads to place on this website?
- I would like to read that Dutch website in English
- Ok Google, Drive this car for me. And, fly this helicopter for me.
- Does this genetic marker correspond to Alzheimer's disease?

But what is learning?

Let’s try to define (machine) learning

What is machine learning?

"Field of study that gives computers the ability to learn without being explicitly programmed"

Arthur Samuel (1950s)

Some Studies in Machine Learning Using the Game of Checkers

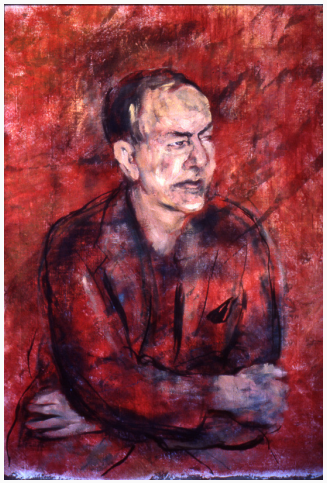
From 1959!

Arthur L. Samuel

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.

Learning as generalization

"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task (or tasks drawn from the same population) more effectively the next time."



Herbert Simon (1983)

Economist, psychologist, political scientist, computer scientist, sociologist, Nobel Prize (1978), Turing Award (1975)…

Learning as generalization

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**."

Tom Mitchell (1999)

Learning = generalization

Learning = generalization

Motivation: Why study machine learning?

- Build computer programs/systems with new capabilities
- Understand the nature of human learning
- Ultimate goal: develop robots that can learn as human beings!

Machine learning is the future

- Gives a system the ability to perform a task in a situation which has never been encountered before
- Big data: Learning allows programs to interact more robustly with messy data
- Starting to make inroads into end-user facing applications already

This course

Focuses on the underlying concepts and algorithmic ideas in the field of machine learning

This course is not about

- Using a specific machine learning tool
- Any single learning paradigm, e.g., deep learning

How will you learn?

- Take classes to learn the models and algorithms
- Finish the homework assignments to deepen your understanding
- Implement the learning models/algorithms by yourself!
- Doing course project for using machine learning techniques to solve problems!

Workload

- 6 homework assignments (most including both latex and programming problems)
- Project (report and a lot of programming)
- Final exam

Warning: This course is one of the most challenging course in CS department. The workload is heavy; you need to plan on ~20 hours per week (on average).

Be cautious when you make the decision☺

Overview of this course

Syllabus

https://www.cs.utah.edu/~zhe/teach/cs6350.html

This course

- The course website contains all the detailed information
- The course website is linked to my homepage

My home page http://www.cs.utah.edu/~zhe/

Course website https://www.cs.utah.edu/~zhe/teach/cs6350.html

This course

Focuses on the underlying concepts and algorithmic ideas in the field of machine learning

This course is not about

- Using a specific machine learning tool
- Any single learning paradigm, e.g., deep learning

How will you learn?

- Take classes to learn the models and algorithms
- Finish the homework assignments to deepen your understanding
- Implement the learning models/algorithms by yourself!
- Doing course project for using machine learning techniques to solve problems!

Canvas

- Feel free to post questions and discuss
- Our TM will respond as fast as they can

Workload

- 6 homework assignments (most including both latex and programming problems)
- Project (report and a lot of programming)
- Final exam

Warning: The workload is heavy; you need to plan on around 20 hours per week.

Be cautious when you make the decision☺
Be sure to plan on enough time on this course!

Basic Knowledge Review

Basic Knowledge Review

- Random events and probabilities
  - We use sets to represent random events, each element in the set is an atomic outcome
    - Example: tossing a coin for 5 times
    - Event A = {H,H,H,T, T}, B = {T,H,T,H,T}, …
  - We use probability to measure the chance an event happens: p(A), p(B)
  - Both A and B happen: $A \cap B$.
  - A or B happens: $A \cup B$.
  - $p(A \cup B) = p(A) + p(B) - p(\mathrm{A} \cap B)$.
  - What is the general version?

Basic Knowledge Review

- Random variables
  - For research convenience /rigor descriptions, we use numbers to represent the sample outcomes. Those numbers are called random variables. The events are represented by random variables falling in some region.
  - Example: tossing a coin, we introduce a R.V. X,
  - X = 1, H; X=0, T.
  - We toss a coin for 5 times, we have 5 R.V. X1, X2, X3, X4, X5
  - Event: we have less than 3 heads:
    - X1+X2+X3+X4+X5 <3
    - Probability: p(X1+X2+X3+X4+X5<3)

Basic Knowledge Review

- Joint probability and conditional probability

$$p(A,B) = p(A)p(B|A) = p(B)p(A|B)$$
$$p(X,Y) = p(X)p(Y|X) = p(Y)p(X|Y)$$

- Independency

$$p(A,B) = p(A)p(B)$$
$$p(X,Y) = p(X)p(Y)$$

What conclusion can you make?

- Conditional independency

$$p(A,B|C) = p(A|C)p(B|C)$$
$$p(X,Y|Z) = p(X|Z)p(Y|Z)$$

Basic Knowledge Review

- Expectation

$$E(X) = \int Xp(X)dX$$

$$E(g(X)) = \int g(X)p(X)dX$$

- Variance

$$Var(X) = E(X^2) - E(X)^2 \geq 0$$

when X is a vector

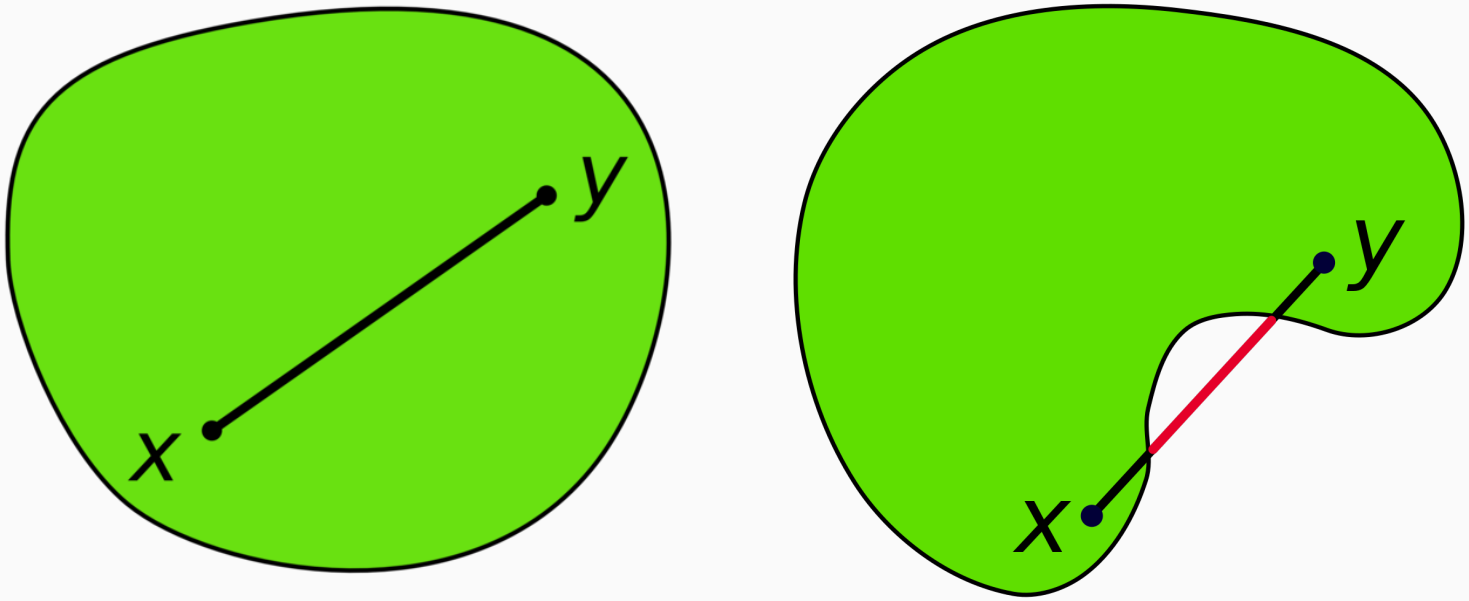
$$Cov(X) = E(XX^T) - E(X)E(X)^T \geq 0$$

- Conditional Expectation/Variance?

$$E(X|Y), Var(X|Y)$$

Basic Knowledge Review

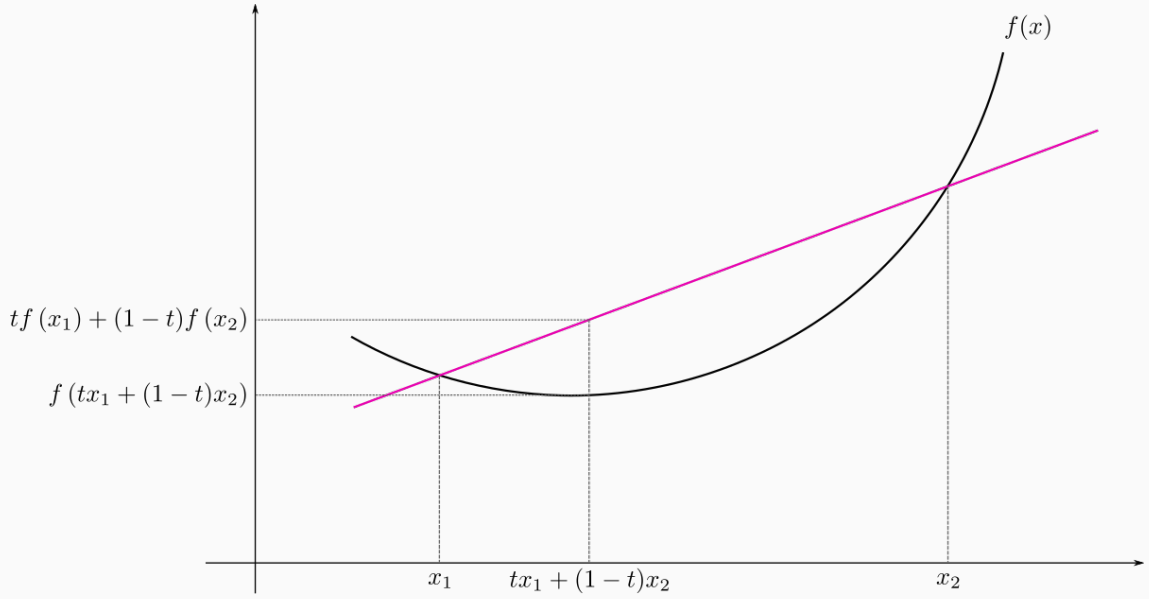
- Convex region/set



Basic Knowledge Review

- Convex function $f: X \longrightarrow R$
- The input domain X is a convex region/set

$$\forall x_1, x_2 \in X, \forall t \in [0,1]: \qquad f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$



Basic Knowledge Review

- Examples of convex functions

Single variable

$$f(x) = e^x$$

$$f(x) = -\log(x)$$

multivariable

$$f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$$

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{x}$$

- How to determine a convex function?

When differentiable $f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y})$

When twice differentiable $\nabla \nabla f(\mathbf{x}) \succeq 0$

Basic Knowledge Review

- Jensen's inequality (for convex function)

When X is random variable

$$f\big(E(X)\big) \leq E(f(X))$$

$$f\big(E(g(X))\big) \leq E(f(g(X)))$$

Basic Knowledge Review

- Matrix derivative

$$
\begin{aligned}
\partial \mathbf{A} &= 0 && (\mathbf{A} \text{ is a constant}) \\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} \\
\partial(\mathbf{X} + \mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} \\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial \mathbf{X}) \\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) \\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) \\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) \\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial \mathbf{X})\mathbf{X}^{-1} \\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) \\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) \\
\partial \mathbf{X}^T &= (\partial \mathbf{X})^T \\
\partial \mathbf{X}^H &= (\partial \mathbf{X})^H
\end{aligned}
$$

Hint: Use matrix cookbook as your reference!

https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf