

CS5350/6350 Machine Learning

Spring 2018

Instructor: Shandian Zhe



Outline

- Machine learning definition, applications and course content
- Course requirements/policies (homework assignments, projects, mid-term/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!

Machine learning is everywhere!

And you are probably already using it

What Other Items Do Customers Buy After Viewing This Item?



Wasabi Power Battery (2-Pack) and Dual Charger for GoPro HERO4 and GoPro AHDBT-401, AHBBP-401
★★★★★ (238)
\$23.99



SanDisk Extreme 64GB UHS-I/U3 Micro SDXC Memory Card Up To 60MB/s Read With Adapter- ...
★★★★★ (443)
\$79.99



EEEKIT 8-in-1 Accessories Kit for Gopro Hero4 Black/Silver Hero HD 3+/3/2/1 Camera, Head Belt Strap ...
★★★★★ (299)
\$29.99



SanDisk Ultra 32GB UHS-I/Class 10 Micro SDHC Memory Card Up to 48MB/s With Adapter- ...
★★★★★ (2,719)
\$19.44

› Explore similar items

spam



Translate

English Spanish French Dutch - de

Jan de kinderen zag



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

Arthur L. Samuel

From 1959!

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 biological learning
- Understanding teaching

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

Related fields

All very active research areas!

- The artificial intelligence dream: Computers that are as intelligent as humans
 - Machine learning closely tied to AI
- Theoretical CS and mathematics
 - Formalizing and understanding learning mathematically
 - Uses ideas from probability and statistics, linear algebra, theory of computation
- Philosophy, cognitive psychology, neuroscience, linguistics, robotics,...
- Many, many application areas
 - medicine, engineering, other areas of CS like compilers, psychology, marketing...

Overview of this course

The main question through the semester

What is learning?

Different formal answers to this problem will give us:

- Various families of learning algorithms
- Techniques for developing new learning algorithms

We will see...

1. Different kinds of models
2. Different learning protocols
3. Learning algorithms
4. Computational learning theory
5. Representing data

We will see different “models”

Or: functions that a learner learns

- Decision trees
- Linear classifiers
- Non-linear classifiers, kernels, neural networks
- Ensembles of classifiers

Different learning protocols

- **Supervised learning**
 - A *teacher* supplies a collection of examples with labels
 - The *learner* has to learn to label new examples using this data
- **Unsupervised learning**
 - No *teacher*, *learner* has only unlabeled examples
 - Data mining
- **Semi-supervised learning**
 - *Learner* has access to both labeled and unlabeled examples
- **Active learning**
 - *Learner* and *teacher* interact with each other
 - *Learner* can ask questions
- **Reinforcement learning**
 - Learner learns by interacting with the environment

Different learning protocols

- **Supervised learning**
 - A *teacher* supplies a collection of examples with labels
 - The *learner* has to learn to label new examples using this data
- **Unsupervised learning**
 - No *teacher, learner* has only unlabeled examples
 - Data mining
- **Semi-supervised learning**
 - *Learner* has access to both labeled and unlabeled examples
- **Active learning**
 - *Learner* and *teacher* interact with each other
 - *Learner* can ask questions
- **Reinforcement learning**
 - Learner learns by interacting with the environment

Who has seen supervised learning before?

Learning algorithms

- **Online algorithms:** Learner can access only one labeled at a time
 - Perceptron, Winnow
- **Batch algorithms:** Learner can access to the entire dataset
 - Naïve Bayes
 - Support vector machines, logistic regression
 - Decision trees and nearest neighbors
 - Boosting
- **Unsupervised/semi-supervised algorithms**
 - Expectation maximization
 - K-Means

Learning algorithms

- **Online algorithms:** Learner can access only one labeled at a time
 - Perceptron, Winnow
- **Batch algorithms:** Learner can access to the entire dataset
 - Naïve Bayes
 - Support vector machines, logistic regression
 - Decision trees and nearest neighbors
 - Boosting
- **Unsupervised/semi-supervised algorithms**
 - Expectation maximization
 - K-Means

Who has used any of
these algorithms before

Representing data

What is the best way to represent data for a particular task?

- The importance of the right features
- Dimensionality reduction

The theory of machine learning

What does it mean to learn?

- Online learning
 - Learner sees examples in a stream and stop making mistakes as we go along (or minimize regret in our decisions).
- Probably Approximately Correct (PAC) Learning
 - After seeing a collection of examples, the learner will (with high probability) produce a function that makes small error.
- Bayesian learning
 - Based on our observations, what is the probability distribution over possible functions that produced the data?

How will you learn?

- Take classes to learn the theories 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!

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

What will you learn?

1. A broad theoretical and practical understanding of machine learning paradigms and algorithms
2. Ability to implement learning algorithms
3. Identify where machine learning can be applied and make the most appropriate decisions (about algorithms, models, supervision, etc)

Brief View of Course Policy

Course Policy

- Workloads
 - 5-6 homework assignments (30%)
 - Written assignments
 - Programming part
 - One course project (30%)
 - Each project consists of at most 2 people
 - mid-term course project report (one week before the mid-term)
 - Final course project report (one week before final)
 - Mid-term exam (20%)
 - Final exam (20%)

Course Policy

- Course project
 - Use machine learning to solve any problems you want
 - Suggest: to form a group as soon as possible
 - You can come to office to brainstorm project ideas

Course Policy

- Written assignments: we only accept latex version, no hand-written versions or their scans; submitted via Canvas.
- Program: Matlab, Python and R are strong suggested; ensure running on CADE machines

Course Policy

- 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 <http://www.cs.utah.edu/~zhe/cs6350.html>

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: $A \cup B$.

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. X_1, X_2, X_3, X_4, X_5
 - Event: we have less than 3 heads:
 - $X_1+X_2+X_3+X_4+X_5 <= 3$
 - Probability: $p(X_1+X_2+X_3+X_4+X_5 <= 3)$

Basic Knowledge Review

- Joint probability and conditional probability

$$\begin{aligned} p(A, B) &= p(A)p(B|A) = p(B)p(A|B) \\ p(X, Y) &= p(X)p(Y|X) = p(Y)p(X|Y) \end{aligned}$$

- Independency

$$\begin{aligned} p(A, B) &= p(A)p(B) \\ p(X, Y) &= p(X)p(Y) \end{aligned}$$

What conclusion
can you make?

- Conditional independency

$$\begin{aligned} p(A, B|C) &= p(A|C)p(B|C) \\ p(X, Y|Z) &= p(X|Z)p(Y|Z) \end{aligned}$$

Basic Knowledge Review

- Expectation

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

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

- Variance

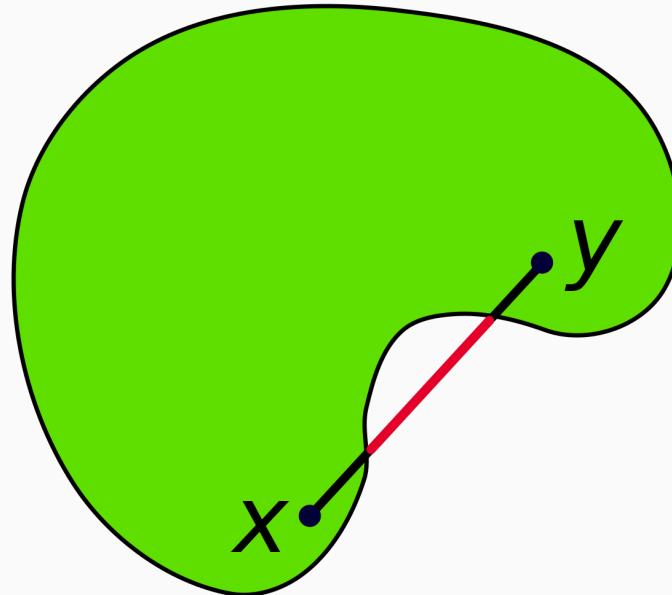
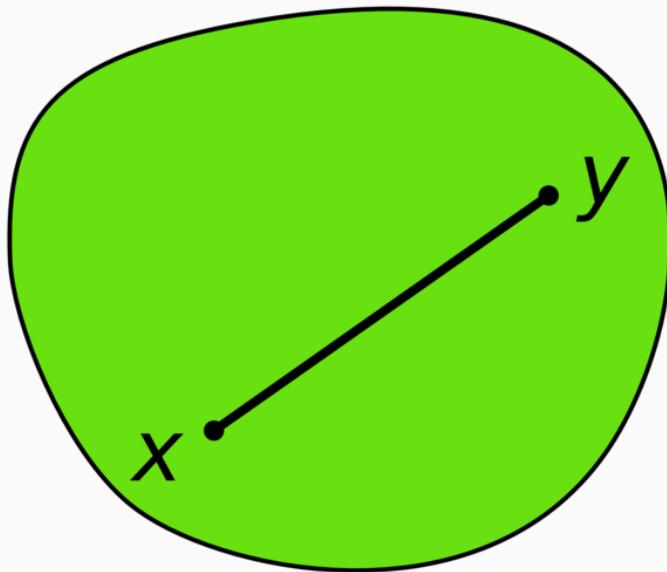
$$Var(X) = E(X^2) - E(X)^2 \geq 0$$

- Conditional Expectation/Variance?

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

Basic Knowledge Review

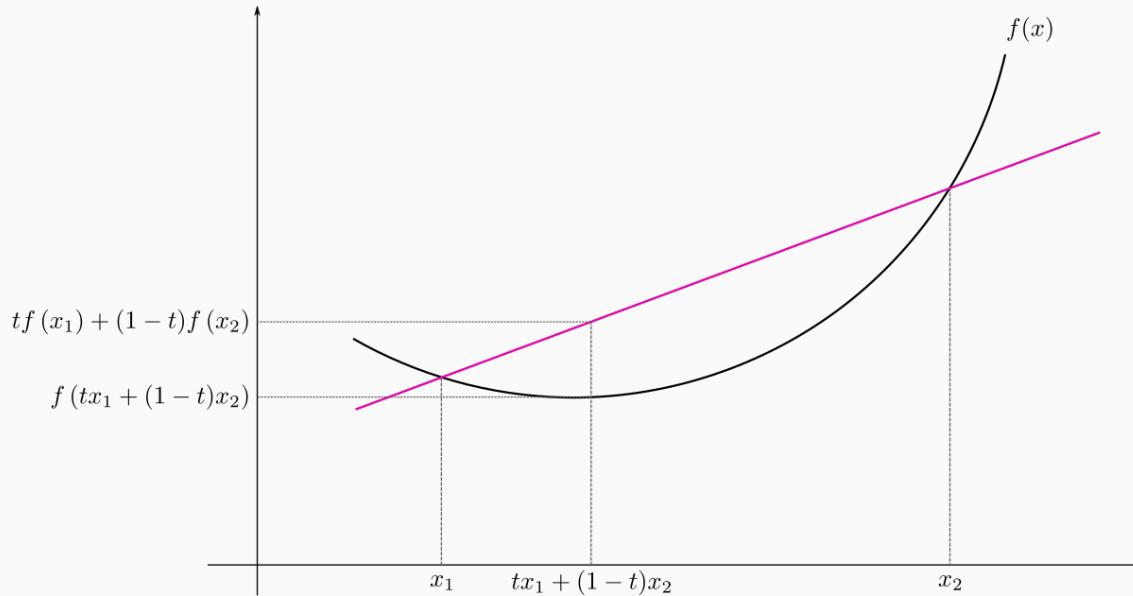
- Convex region/set



Basic Knowledge Review

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

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



Basic Knowledge Review

- Examples of convex functions

$$f(x) = Ax + b \quad f(x) = e^x \quad f(x) = -\log(x)$$

- How to determine a convex function?

When differentiable

$$f(x) \geq f(y) + f'(y)(x - y)$$

When twice differentiable

Second derivatives is non-negative/positive semi-definite

Basic Knowledge Review

- Jensen's inequality (for convex function)

When X is random variable

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

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