CS 8501 Advanced Topics in Machine Learning

Lecture 01: Introduction

Yangfeng Ji
Information and Language Processing Lab
Department of Computer Science
University of Virginia
https://yangfengji.net/

Course Information

Course Webpage

Our course webpage:

https://yangfengji.net/uva-advanced-ml/

- Contact information
- Office hours
- Course schedule
- Rubrics and grading policy
- Final project suggestions

Goals of This Course

This course is designed for

- a research-like course environment
- introducing important topics in machine learning, which we do not have time to cover in the introductory course (e.g., CS 6316)
- understanding the recent advances

Goals of This Course (Cont.)

This course is **not** designed for

- Learning machine learning packages
- Working on machine learning projects
- Solve problems on textbooks

Requirements for Students

To make sure this is the course that you would like to spend time on, you should

- be ready to devote time on reading, writing and discussion
- feel comfortable with mathematical notations

Depending on many other random factors, the expected time that you should spend on this course

- 3 -- 8 hours per week, in addition to attending the class lectures

Topics

- Bayesian machine learning
  - Bayesian inference
  - Probabilistic graphical models
  - Approximate inference
- Deep neural networks
  - Variational autoencoders
  - Generative adversarial networks
- Further topics
  - Beyond the IID assumption

Lecture Schedule

You can find the lecture schedule on the course webpage

https://yangfengji.net/uva-advanced-ml/schedule.html

Week	Dates	Topic	Readings
1	Aug. 24	Introduction	Guideline
2	Aug. 29, 31	Generative modeling	Guideline
3	Sept. 5, 7	Bayesian statistics	Guideline
4	Sept. 12, 14	Probabilistic graphical models	Guideline
5	Sept. 19, 21	Probabilistic graphical models (II)	Guideline
6	Sept. 26, 28	Information theory basics	Guideline
7	Oct. 5	Variational inference	Guideline
8	Oct. 10, 12	Variational inference (II)	Guideline
9	Oct. 17, 19	Monte Carlo inference	Guideline
10	Oct. 24, 26	Monte Carlo inference (II)	Guideline
11	Oct. 31, Nov. 2	Variational autoencoders	Guideline
12	Nov. 7, 9	Variational autoencoders (II)	Guideline
13	Nov. 14, 16	Generative adversarial networks	Guideline
14	Nov. 21	Beyond the IID assumption	Guideline
15	Nov. 28, 30	Beyond the IID assumption (II)	Guideline
16	Dec. 5	Summary	

Readings

Reading is the foundation of class participation!

Textbooks

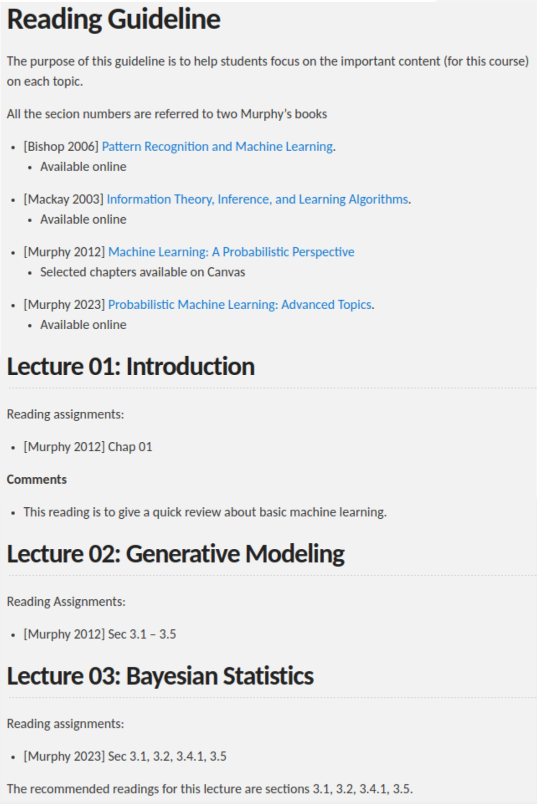
The readings of this course will be selected from the following four books:

- Mackay. Information Theory, Inference, and Learning Algorithms. 2003
- Bishop. Pattern Recognition and Machine Learning. 2006
- Murphy. Machine Learning: A Probabilistic Perspective. 2012
- Murphy. Probabilistic Machine Learning: Advanced Topics. 2023

Plus some research papers.

Reading Guideline

For each topic, we will try to provide some guideline to help you navigate the reading



Reading Guideline

The purpose of this guideline is to help students focus on the important content (for this course) on each topic.

All the secion numbers are referred to two Murphy's books

- [Bishop 2006] Pattern Recognition and Machine Learning.
  - Available online
- [Mackay 2003] Information Theory, Inference, and Learning Algorithms.
  - Available online
- [Murphy 2012] Machine Learning: A Probabilistic Perspective
  - Selected chapters available on Canvas
- [Murphy 2023] Probabilistic Machine Learning: Advanced Topics.
  - Available online

Lecture 01: Introduction

Reading assignments:

- [Murphy 2012] Chap 01

Comments

- This reading is to give a quick review about basic machine learning.

Lecture 02: Generative Modeling

Reading Assignments:

- [Murphy 2012] Sec 3.1 – 3.5

Lecture 03: Bayesian Statistics

Reading assignments:

- [Murphy 2023] Sec 3.1, 3.2, 3.4.1, 3.5

The recommended readings for this lecture are sections 3.1, 3.2, 3.4.1, 3.5.

Assignments

- Questions: 35 points
  - 10×3.5 points
- Discussion: 40 points
  - 10×4.0 points
- Final project
  - Proposal: 7 points
  - Final report: 13 points
- Class attendance
  - 5 points

Question and Discussion Schedule

- Each week, students should finish the reading assignment for next week's topic and submit one question before
- As the lecture, students will also have one week to answer a question from another student

For example,

- Read the assigned materials for **lec01: generative modeling**
- Submit your question by **Sunday Aug 28, 11:59 PM**
- (The instructors will randomly select some questions and post them on the discussion board for next step)
- Submit your answer to another student's question by **Sep 7, 11:59 PM**

Question

- Students are expected to submit one question on the topic before the lecture
- Each question is limited to *100 words*
- Ten questions in total for the semester, which means you can skip four topics based on your preference
- Full points of each question are 3.5

You can find the grading rubric on the course webpage

Discussion

- Students are expected to answer one question from other students within *one week* after the lecture
- Each answer is limited to *150 words*
- Ten questions in total for this semester, similar to the question part, you can skip four of them
- Full points of each answer are 4.0

Grading rubric is available on the course webpage

Organizing Lectures

- Each week will cover one lecture topic (some important topics will span two weeks)
- Class on Monday will discuss the basic idea on each topic (based on the reading materials)
- Class on Wednesday will discuss some selected questions from students

Final Project

The final project consists of three components

- Write a survey paper on a specific machine learning topic
- You can pick the topic that is related to your own research
- Each team should have 2 -- 3 students
- A detailed suggestion will be provided later

Class Attendance

- Attendance will be taken in each class time
- To get full points of class attendance, students should not miss more than one class

Missing classes	Point deduction
1 < n <= 2	-1.0
2 < n <= 4	-2.0
4 < n <= 6	-3.0
6 < n <= 8	-4.0
n > 8	-5.0

- In-class discussion is not part of the performance evaluation, but it is highly encouraged

Student Responsibility

- Read assigned materials
- Submit a question before the lecture
- Answer another question after the lecture

Grade Mapping

Point range	Letter grade
[99 100]	A+
[94 99)	A
[90 94)	A-
[88 90)	B+
[83 88)	B
[80 83)	B-
[74 80)	C+
[67 74)	C
[60 67)	C-
[0 60)	F

Machine Learning: A Quick Review

Categories of Learning Framework

$$f : X \to Y$$

Learning f with

- X and Y provided: supervised learning
- Y not provided: unsupervised learning
- Y as an indirect feedback: reinforcement learning

Some Examples

Image classification

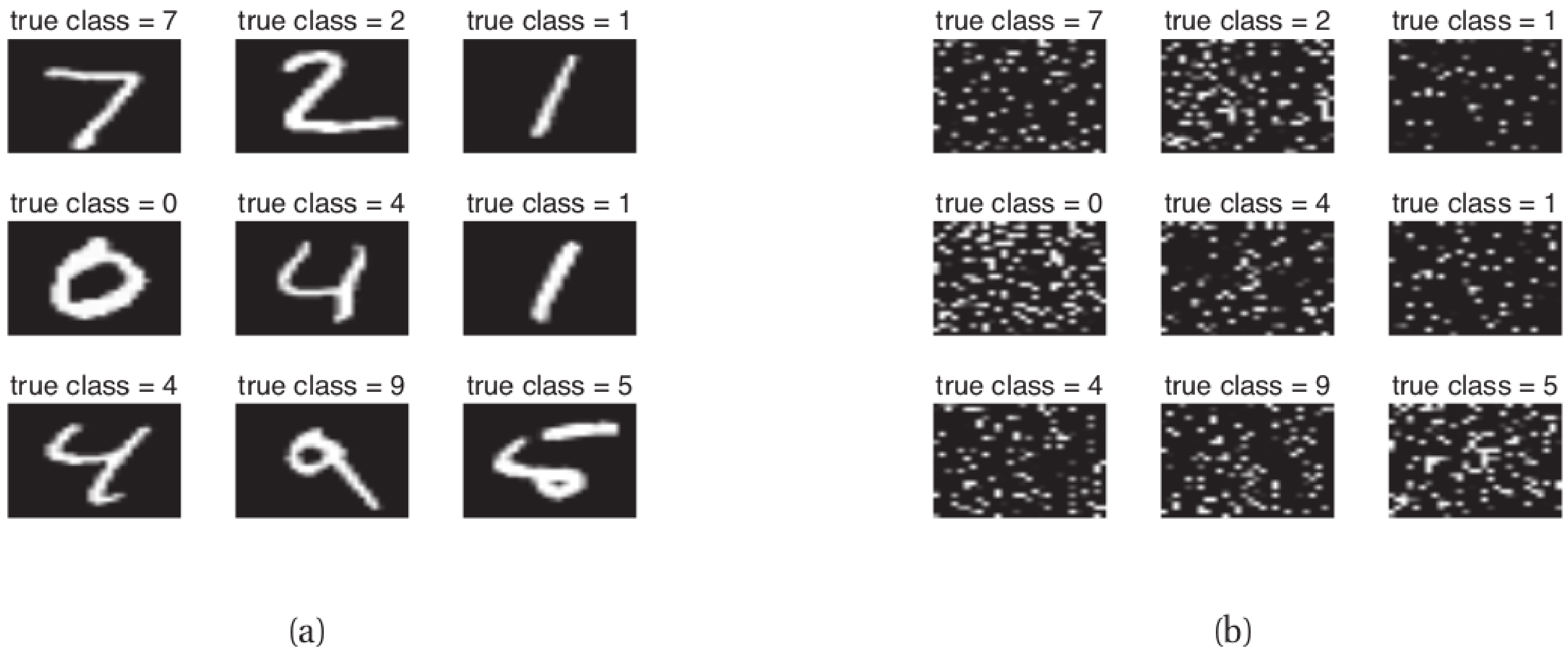
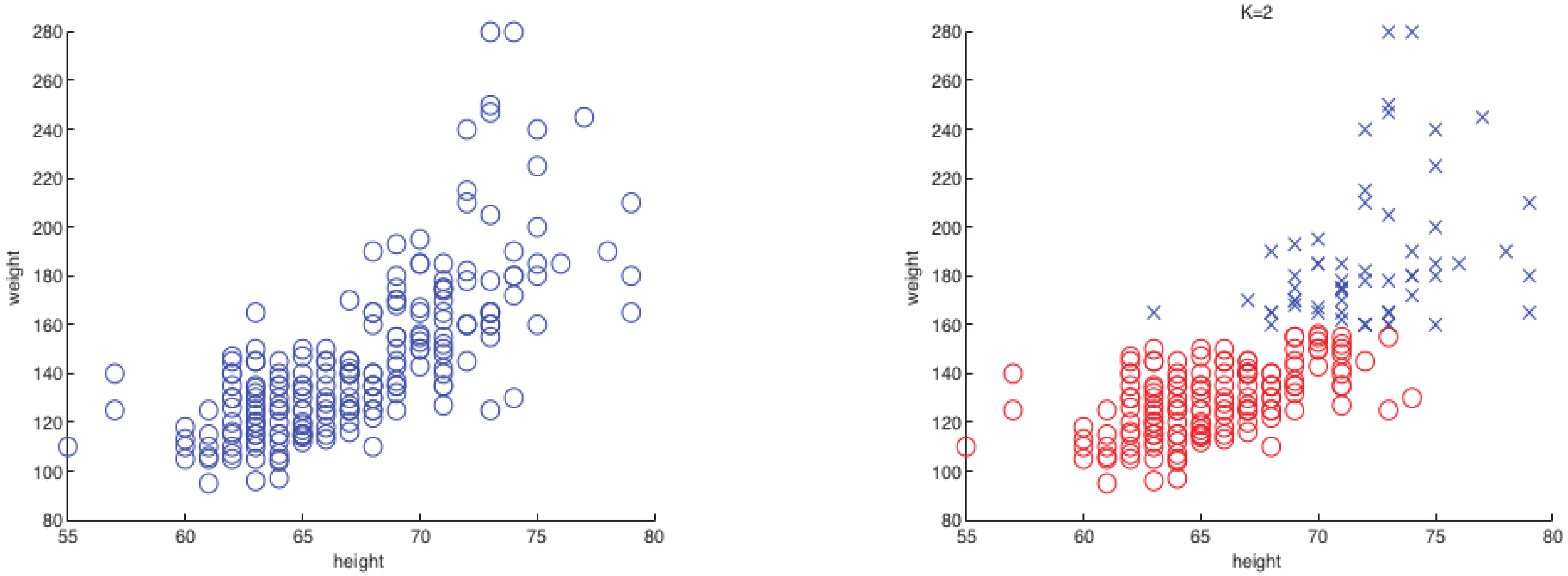


Figure 1.5 (a) First 9 test MNIST gray-scale images. (b) Same as (a), but with the features permuted randomly. Classification performance is identical on both versions of the data (assuming the training data is permuted in an identical way). Figure generated by `shuffledDigitsDemo`.

Some Examples (II)

K-Means clustering



(a) (b)

Figure 1.8 (a) The height and weight of some people. (b) A possible clustering using $K = 2$ clusters. Figure generated by `kmeansHeightWeight`.

Some Examples (III)

Principal component analysis

Figure 1.10 a) 25 randomly chosen 64 × 64 pixel images from the Olivetti face database. (b) The mean and the first three principal component basis vectors (eigenfaces). Figure generated by `pcaImageDemo`.

note @755

Eigenfaces are scary, don't work on that section at night.

Avoid doing the eigenface section of homework at night.
It WILL give you nightmares.
I'm pretty sure eigenface 4 is out there trying to kill me.

hw6

Source: https://twitter.com/alpha_convert/status/1452731497060864001

Generalization

- The ultimate goal is to have the generalization power of f, which means
  - Great performance on training set is not what we want
  - Be careful about which f we should select, when multiple of them are available (model selection)

Parametric Methods

For parametric models, two typical ways of learning model parameter θ are

- MLE

$$\hat{\theta} \leftarrow \max_{\theta} \log p(y \mid x; \theta)$$

- MAP

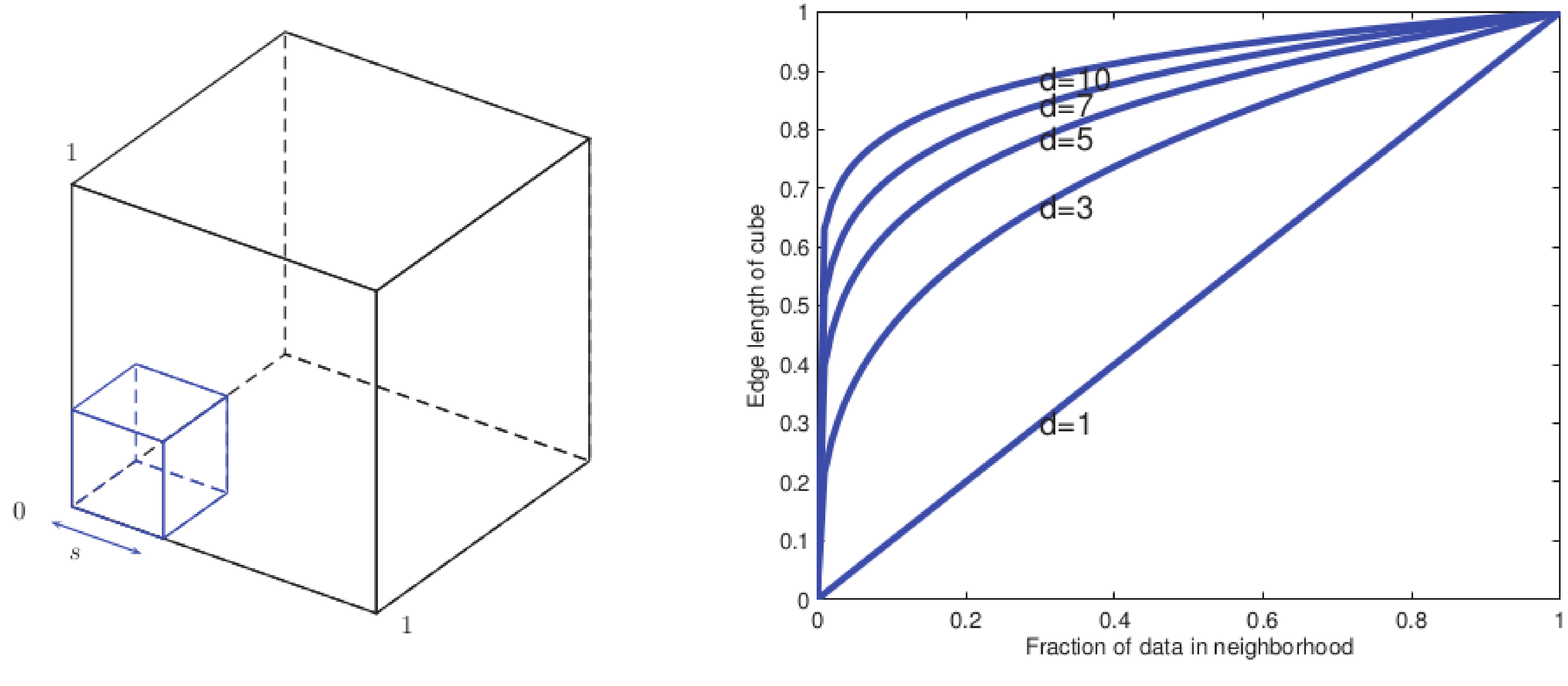
$$\hat{\theta} \leftarrow \max_{\theta} \log p(\theta \mid x, y)$$

  implicitly θ here is a random variable with a prior distribution $p(\theta)$

Non-parametric Methods

- Explanations
  - [Wasserman 2006] "The basic idea of nonparametric inference is to use data to infer an unknown quantity while making as few assumptions as possible ... it is difficult to give a precise definition of nonparametric inference ..."
- Examples
  - k-NN classification
  - Kernel density estimation

Curse of Dimensionality



(a) (b)

Inductive Bias

- [Murphy 2012]: "The main way to combat the curse of dimensionality is to make some assumptions about the nature of the data distribution ..." (page 19)
- [Mackay 2003]: You cannot do inference without making assumptions" (page 26)
  - This statement is introduced under a different context, but it's generally true.

No Free Lunch Theorem

- [Murphy 2012]: "there is no universally best model" (page 24)
- [Shalev-Shwartz and Ben-David, 2014]: Let A be any learning algorithm for the task of binary classification. Let m be any number smaller than $|\mathcal{X}|/2$, representing a training set size. Then there exists a distribution $\mathcal{D}$ over $\mathcal{X} \times \{0,1\}$ such that
  - There exists a function $f : \mathcal{X} \to \{0,1\}$ with $L_{\mathcal{D}}(f) = 0$
  - With a certain probability over the choice of $S \sim \mathcal{D}^m$, we have

$$L_{\mathcal{D}}(A(S)) \geq 1/8$$

- George Box: "All models are wrong, but some are useful"
  - The motto of this course

Thank You!