

CS 446/ECE 449: Machine Learning

Lecture 1: Introduction, Nearest Neighbor

Han Zhao
01/16/2024

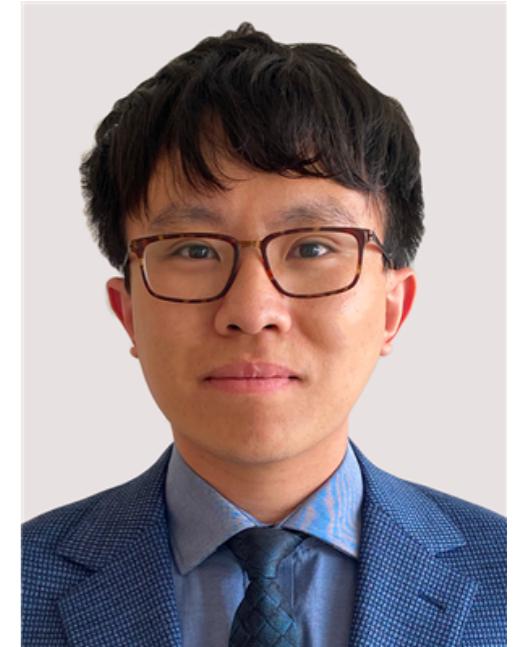


Lecture Today

- Course Information
- Introduction to Machine Learning
- Nearest Neighbor

Instructors

- Name: Han Zhao
- Current position: Assistant Professor @ CS
- Email: hanzhao@illinois.edu
- Research Interests: Machine Learning
 - Robustness
 - Fairness
 - Privacy
 - Interpretability
 - Probabilistic Graphical Models
- Office: SC 3320
- Office Hour: Thu 2-3pm



Instructors

- Name: Shenlong Wang
- Current position: Assistant Professor @ CS
- Email: shenlong@illinois.edu
- Research Interests: Vision & Robotics
 - Robotics
 - 3D Computer Vision
 - Autonomous driving



- Office: SC 3336
- Office Hour: Thu 2-3pm



Teaching Assistants

- Name: Amnon Attali
- Current position: 5th year PhD in CS
- Research Interests: Robotics
 - Representation learning for robotics
 - Guided planning
- Office: Lounge outside SC 3102
- Office Hour: Tues 10-11am



Teaching Assistants

- Name: Jane Du
- Current position: 3rd year PhD in CS
- Research Interests: Machine Learning
 - Fairness
 - Sparsity
 - Optimization
- Office: Lounge outside SC 3102
- Office Hour: Mon 10-11am



Teaching Assistants

- Name: Yubin Ge
- Current position: 5th year PhD in CS
- Research Interests: Natural Language Processing
 - Trustworthy machine learning
- Office: Lounge outside SC 3102
- Office Hour: Wed 3-4pm



Teaching Assistants

- Name: Jing Wen
- Current position: 2nd year PhD in CS
- Research Interests: Computer Vision
 - 3D vision
- Office: SC 3307A
- Office Hour: Fri 3-4pm



Logistics

- Course website: <https://courses.grainger.illinois.edu/cs446/sp2024/>



[Home](#) [Lectures](#) [Homework](#)

CS446/ECE449: Machine Learning (Spring 2024)



Course Information

The goal of machine learning is to develop algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed for a particular task. In this course we will cover three main areas, (1) supervised learning, (2) unsupervised learning, and (3) reinforcement learning. In particular we will cover the following: perceptron, decision trees, Naive Bayes, Gaussian Bayes, linear regression, logistic regression, support vector machines, learning theory, deep learning, structured methods, kMeans, Gaussian mixtures, expectation maximization, VAEs, GANs, Markov decision processes, Q-learning and Reinforce.

Logistics

- Discussion forum:
 - campuswire: <https://campuswire.com/p/G47CE41F1/>

The screenshot shows the CampusWire class feed interface for CS 446 / ECE 449: Machine Learning. The left sidebar includes links for Notifications, DMs, Calendar, Search, Class feed (highlighted), Rooms, Files, Grades, and Settings. The main area displays the 'Class feed' for the course, showing a message that 'No one has created a post in this feed yet' and a 'Create a new post' button. To the right, a 'Class feed at a glance' section shows statistics for Pinned posts (0), Unresolved questions (0), and Unread messages (0). A decorative palm tree icon and the message 'All questions have been resolved. Enjoy the rest of your day.' are also present.

Logistics

- Homework submission: Gradescope
 - Link: <https://www.gradescope.com/courses/690539>
 - Entry code: 3P22VV

 by Turnitin

CS446/ECE449

Machine Learning

- [Dashboard](#)
- [Assignments](#)
- [Roster](#)
- [Extensions](#)
- [Course Settings](#)

Instructors

- Han Zhao
- Shenlong Wang

Course Actions

- [Unenroll From Course](#)

CS446/ECE449 | Spring 2024
Course ID: 690539

Description

Machine learning has been widely applied in many areas, such as computer vision, natural language processing, speech recognition, robotics, and computational biology. This course will cover the fundamental concepts, theory and algorithms in machine learning, including (1). supervised learning, (2). unsupervised learning, and (3). reinforcement learning. We will also discuss the theoretical foundations of machine learning, such as PAC learning theory. The course will be self-contained, and existing knowledge about machine learning algorithms is preferred but not required. Prerequisites include probability and statistics, linear algebra and calculus. We will cover the following aspects: the perceptron algorithm, linear regression, logistic regression, support vector machines, deep learning (multilayer perceptron, convolutional neural network and recurrent neural network), generative models (generative adversarial networks, variational auto-encoders, diffusion process, transformers), PAC learning theory, K-means, Gaussian mixtures, expectation maximization, Markov decision processes, Q-learning and Reinforce.

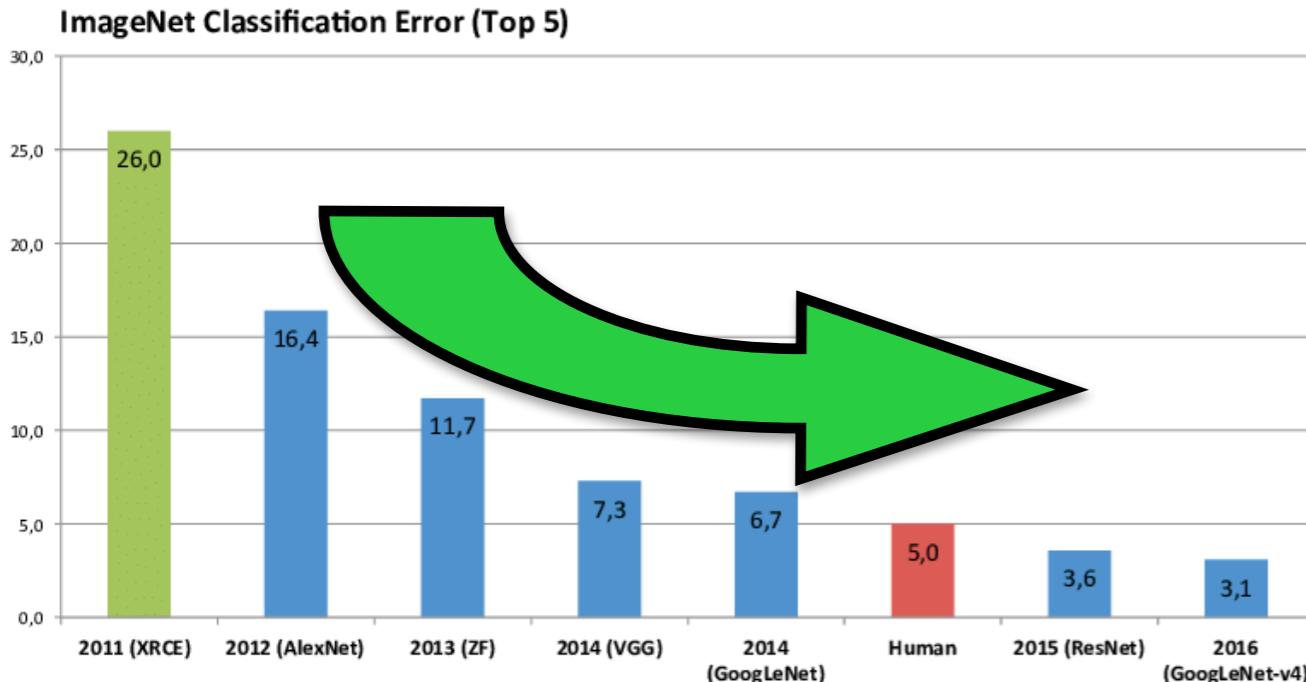
Things To Do

- Add students or staff to your course from the [Roster](#) page.
- Create your first assignment from the [Assignments](#) page.

Active Assignments Released Due (CST) ▾ Submissions % Graded ▾ Published Regrades

Introduction to Machine Learning

What is Machine Learning and why should we care?

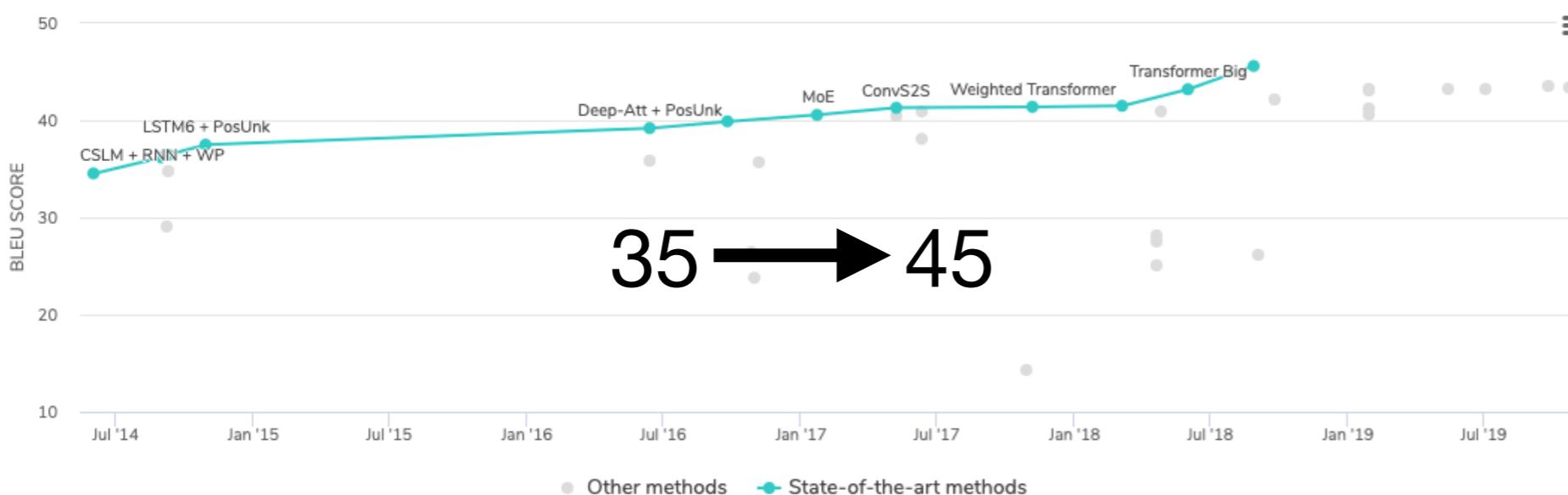


ImageNet: ~1M images, ~1K classes [Deng et al. 09]

Machine Translation on WMT2014 English-French



AlphaGo vs Lee Sedol [Silver et al.'16]



Machine Translation, ~3M parallel sentences [Cho et al.'14; Devlin et al.'14]



ChatGPT [OpenAI' 22]

Introduction to Machine Learning

Question Answering

Training phase:



Input

Who is the president of the US?

In which U.S. state is the University of Illinois Urbana-Champaign located?

Label

Joe Biden

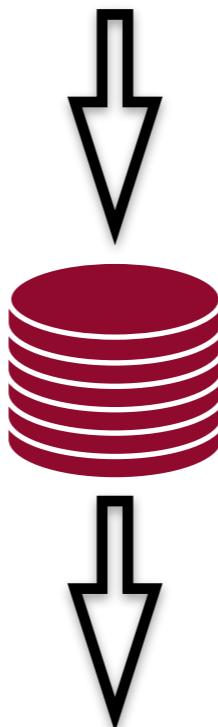
Illinois



ChatGPT [OpenAI' 22]

Test phase:

What is the capital city of Japan



Tokyo

Introduction to Machine Learning

But, what is a more formal definition of ML?

Wikipedia:

Machine learning gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959).^{[1][2]} It is a subfield of computer science.^[3]

The idea came from work in artificial intelligence.^[4] Machine learning explores the study and construction of algorithms which can learn and make predictions on data.^[5] Such algorithms follow programmed instructions, but can also make predictions or decisions based on data.^{[6]:2} They build a model from sample inputs.

ChatGPT 3.5 ▾

ChatGPT:

 You
What is machine learning

 ChatGPT
Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed. In other words, machine learning involves creating systems that can automatically learn and improve from experience.

There are three main types of machine learning:

1. **Supervised Learning:** The algorithm is trained on a labeled dataset, where the input data is paired with the corresponding output. The goal is to learn a mapping from inputs to outputs so that the algorithm can make predictions on new, unseen data.
2. **Unsupervised Learning:** The algorithm is given unlabeled data and must find patterns or structure within it. Clustering and dimensionality reduction are common tasks in unsupervised learning.
3. **Reinforcement Learning:** The algorithm learns by interacting with an environment. It receives feedback in the form of rewards or penalties based on the actions it takes, and its goal is to learn a policy that maximizes the cumulative reward.

Course Overview

Schedule:

Lecture Description	Topics
Part I: Supervised Learning	
1 Course overview	<ul style="list-style-type: none">• k-Nearest Neighbor
2 Linear Parametric Models (Generative)	<ul style="list-style-type: none">• (Discrete & Continuous) Naive Bayes
3-4 Linear Parametric Models (Discriminative)	<ul style="list-style-type: none">• Linear Regression, Logistic Regression
5 Linear Parametric Models (Discriminative)	<ul style="list-style-type: none">• Support Vector Machines
6 Non-linear Parametric Models (Discriminative)	<ul style="list-style-type: none">• Kernel Methods
7-8 Non-parametric Models	<ul style="list-style-type: none">• Decision Trees, Random Forests• AdaBoosting
9-10 Learning Theory	<ul style="list-style-type: none">• PAC Learning Theory
11-15 Deep Learning	<ul style="list-style-type: none">• The Perceptron Algorithm• Multilayer Perceptron, Backpropagation Algorithm• Convolutional Neural Network• Sequential Models, Recurrent Neural Network• Attention, Transformer
16	Midterm Exam
Part II: Unsupervised Learning	
17-18 Clustering	<ul style="list-style-type: none">• k-Means
19 Dimensionality Reduction	<ul style="list-style-type: none">• Gaussian Mixture Models, Expectation-Maximization• Principal Component Analysis• Singular Value Decomposition• Auto-encoders
20-25 Generative Models	<ul style="list-style-type: none">• Introduction to Information Theory• Variational Auto-encoders• Generative Adversarial Networks• Diffusion Models / Score Matching• Self-supervised Learning / Contrastive Learning• Language Modelling / Foundation Models
Part III: Reinforcement Learning	
26-28 Reinforcement Learning	<ul style="list-style-type: none">• Markov Decision Process• Q-Learning, Deep Q-Network• Policy Gradient
29-30	Final exam

Course Overview

Lecture-based course:

- 6 Homeworks
- Two sections (P3 & P4)
- One mid-term & one final exam

	3-credit	4-credit
Homework 0	0%	0%
Homework 1	12%	10%
Homework 2	12%	10%
Homework 3	12%	10%
Homework 4	12%	10%
Homework 5	12%	10%
Homework 6	12%	10%
Mid-term Exam	20%	20%
Final Exam	20%	20%

Course Overview

Homework today:

- Sign up for Campuswire and Gradescope
- Take a look at the course syllabus
- Homework 0

FAQs

What technical background do I need to succeed in this course?

Prerequisites:

- Probability and statistics
- Linear algebra
- Comfortable with programming in Python (Numpy, TensorFlow, PyTorch, etc)

FAQs

What if I hate math?

- This course will mainly be focusing on the theory and algorithms of machine learning techniques.
- Doing well in this course will require comfort with math, including derivations and proofs.
- If you prefer applying ML techniques to your favorite applications (biology, astronomy, chemistry, finance, etc), consider Applied Machine Learning (CS 441)

FAQs

What if I want to learn more math of ML?

- CS 540: Deep Learning Theory
- ECE 543: Statistical Learning Theory
- We will have two lectures covering the basic framework of learning theory, but that's only an introductory glimpse

FAQs

What is the difference between CS 446 (Machine Learning) and CS 440 (Artificial Intelligence)?

- CS 446: ML exclusively, with a focus on theory and algorithms under different learning paradigms
- CS 440: Broad introduction to AI beyond ML, including computer vision (CV), natural language processing (NLP), search and planning (A^* algorithm etc), constrained satisfaction problems, game theory, etc.

FAQs

The class if full but I'd still like to take the class, is there a waiting list?

- No; Check for space to become available

Will this course be video recorded?

- No; in-person lecture by default, unless otherwise notified by the instructors

Can I audit this course?

- Yes, but no need to submit HWs. You can also sign up for campuswire, though

Introduction to Machine Learning

Running Example: ImageNet classification



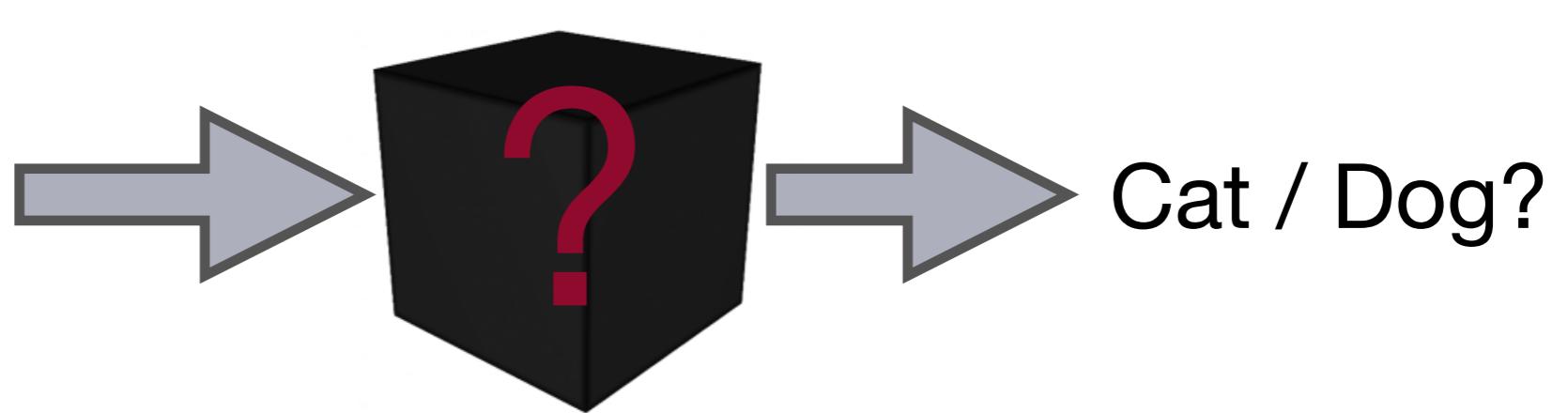
- Input: image, or a 3d array of size 224x224x3
- Output: 1 out of the potential 1000 classes (plane, cat, dog, vehicle, etc)

Introduction to Machine Learning

Running Example: ImageNet classification

- Input: image, or a 3d array of size 224x224x3
- Output: 1 out of the potential 1000 classes (plane, cat, dog, vehicle, etc)

Our goal: build a model/predictor that takes an input and predicts an output label (class)



Introduction to Machine Learning

How to obtain such a model/predictor? Let's recall the definition of machine learning:



“ Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience.

~ Tom Mitchell,
Machine Learning, McGraw Hill, 1997

Carnegie Mellon University
Machine Learning

MACHINE
LEARNING



TOM M. MITCHELL

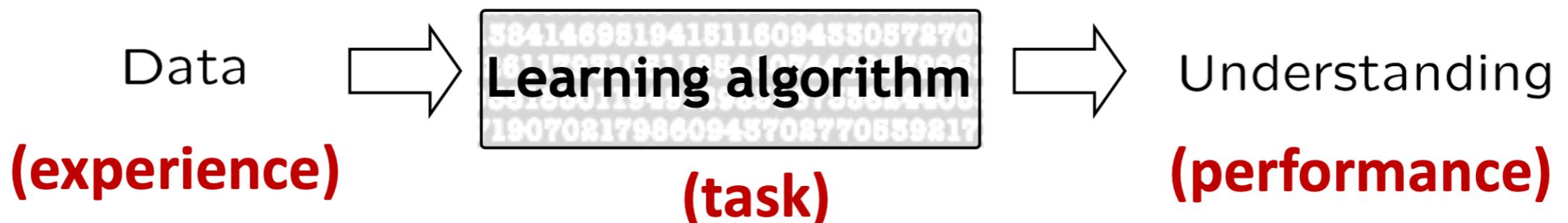
Introduction to Machine Learning

How to obtain such a model/predictor? Let's recall the definition of machine learning:

“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 M. Mitchell

- **E: data**
- **T: task of interest**
- **P: objective function**



Introduction to Machine Learning

Running Example: ImageNet classification

- Input: image, or a 3d array of size 224x224x3
- Output: 1 out of the potential 1000 classes (plane, cat, dog, vehicle, etc)

Formulation & terminology:

- Input data/vector: $x \in \mathbb{R}^d$
- Output label/class: $y \in [k] := \{1, \dots, k\}$
- Dataset: $\mathcal{D} := \{(x^{(i)}, y^{(i)})\}_{i=1}^n$

(Supervised) Machine Learning algorithm:

$$M = \mathcal{A}(\mathcal{D})$$

- The learning algorithm \mathcal{A} takes the dataset as input and return a model M
- The model $M : \mathcal{X} \rightarrow \mathcal{Y}$ can then be used for prediction: $\hat{y} = M(x)$

Introduction to Machine Learning

Machine learning algorithms can be characterized by:

- Type of data available for learning
 - Supervised: $\mathcal{D} := \{(x^{(i)}, y^{(i)})\}_{i=1}^n$
 - Unsupervised: $\mathcal{D} := \{x^{(i)}\}_{i=1}^n$
- Complexity of model:
 - Fixed complexity as data grows: parametric
 - Linear vs nonlinear
 - Increasing complexity as data grows: non-parametric
- Structure of output: independent vs structured
- Type of output: classification vs regression
 - Classification: $\mathcal{Y} = [k] := \{1, \dots, k\}$
 - Regression: $\mathcal{Y} = \mathbb{R}$

Introduction to Machine Learning

Typical examples of supervised learning:

- Classification:
 - Spam filter: $\mathcal{Y} = \{0,1\}$
 - ImageNet: $\mathcal{Y} = [1000]$
 - Sentiment analysis: $\mathcal{Y} = \{1,2,3,4,5\}$
- Regression:
 - Stock price prediction: $\mathcal{Y} = \mathbb{R}_+ := \{t : t \geq 0\}$
 - Portfolio return prediction: $\mathcal{Y} = \mathbb{R}$
 - Temperature forecast: $\mathcal{Y} = [-100,100] \subseteq \mathbb{R}$

Nearest Neighbor

Running Example: ImageNet classification

- Input: image, or a 3d array of size 224x224x3
- Output: 1 out of the potential 1000 classes (plane, cat, dog, vehicle, etc)

We have a huge dataset from ImageNet that contains 1M images:
 $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{1,000,000}$. Now here comes a new image x , how can we predict/determine its label \hat{y} ?



?

Cat / Dog?

Nearest Neighbor

We have a huge dataset from ImageNet that contains 1M images:
 $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{1,000,000}$. Now here comes a new image x , how can we predict/determine its label \hat{y} ?

A simple idea: compare the similarity of x with all the existing images, and use the label ($y^{(c)}$) of the closest image $x^{(c)}$ as the prediction:

$$\hat{y} = y^{(c)} \text{ where } c = \arg \min_{i \in [n]} \|x - x^{(i)}\|_2$$



?

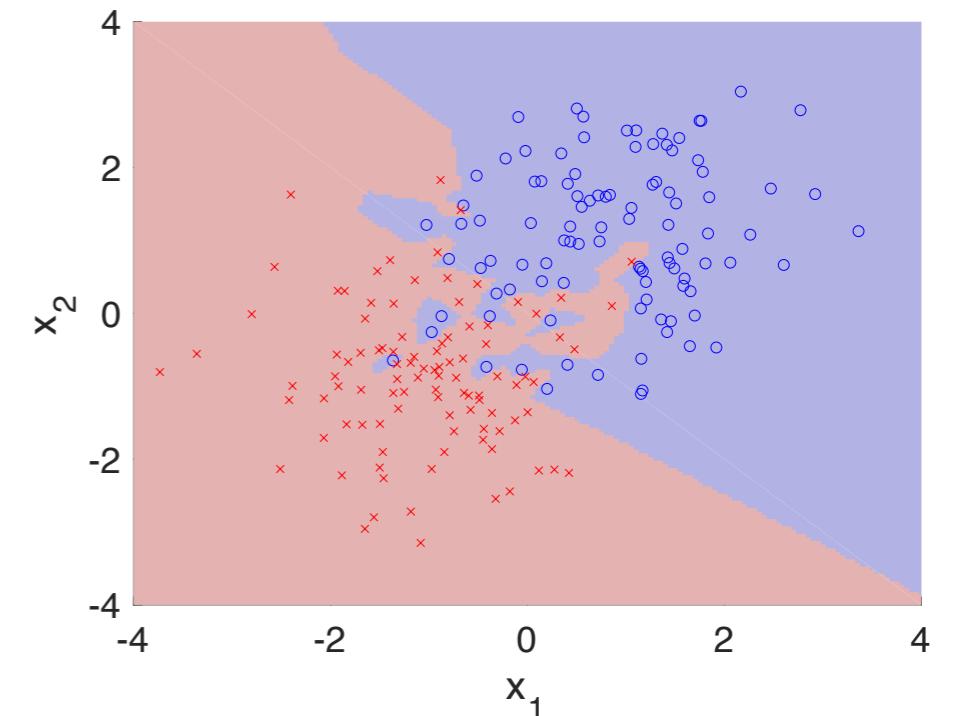
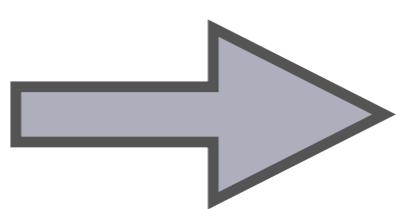
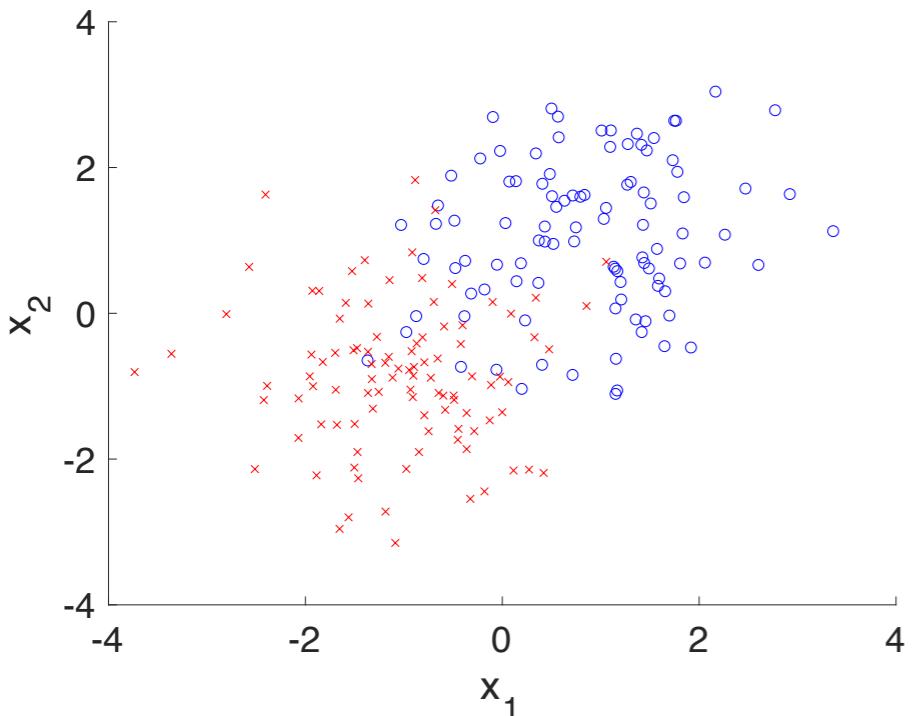
Cat / Dog?

Nearest Neighbor

We have a huge dataset from ImageNet that contains 1M images:
 $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{1,000,000}$. Now here comes a new image x , how can we predict/determine its label \hat{y} ?

A simple idea: compare the similarity of x with all the existing images, and use the label ($y^{(c)}$) of the closest image $x^{(c)}$ as the prediction:

$$\hat{y} = y^{(c)} \text{ where } c = \arg \min_{i \in [n]} \|x - x^{(i)}\|_2$$



Nearest Neighbor

What are the potential shortcomings of the Nearest Neighbor method?

In-class discussion

Nearest Neighbor

Categorization of the Nearest Neighbor method:

- Supervised learning algorithm
- Nonlinear (decision boundary)
- Non-parametric (complexity grows as the data grows)
- Can be extended/generalized to k -Nearest Neighbor

Limitations:

- Prediction complexity also grows as the data grows
- Need to maintain all the training data
- Need to specify a meaningful distance metric

Later we will discuss many more supervised learning algorithms to mitigate the above issues

Summary

- Course overview, syllabus, covered topics
- Nearest neighbor Classifier