

# CS484 Introduction to Machine Learning

Instructor: Dr. Binghui Wang

Lecture hours: MW 10:00AM – 11:15AM

Location: John T. Rettaliata Engg Center 104

Office: Stuart Building, 216C

Email: [bwang70@iit.edu](mailto:bwang70@iit.edu)

Teaching Assistants

- TBD

# Course Description

## Topics

- Supervised learning
  - Regression, classification
- Unsupervised learning
  - Clustering, dimen. reduction
- Semi-supervised learning

## Advanced topics (Intro level)

- Federated learning
- Graph neural network (GNN)
- Large language model (LLM)
- Generative AI
- Trustworthy machine learning

**Topics not covered:** Reinforcement learning, Boosting/Ensemble learning, Bayesian learning, Active learning, Sequential models, Sampling methods, etc.

## Textbooks (free E-Book)

- Pattern Recognition and Machine Learning, by Christopher M. Bishop
- Machine Learning: A Probabilistic Perspective, by Kevin P. Murphy

# Grades & Assignments

## Grades

- 5 assignments (50%)
- Final project presentation (15%)
- Final take-home exam (35%)

## Assignments

- Assignment submission: Single PDF for solution or/and separate source code files, NOT using zip
- Late submission: All assignments are due by Sunday at 11:59 PM. 50% of the grade will be deducted for that assignment if it is late within 1 day (24 hours). 0 grade if the assignment is late more than 1 day
- Per the university rule, it is NOT allowed to use GenAI (ChatGPT, Perplexity.AI, Gemini, Claude, etc) to generate your solution

# Final Presentation & Exam

2-3 students form a group

An initial project topic (along with the group members) is due at the **end of Week 10**

- Let me/TAs know as early as possible if you decide to change the topic after the due date

The final presentation includes a **6 min project representation**

- **ALL group members should clearly clarify their contributions in the project**

**One week** is allocated for the final take-home exam

- **Due Dec 2<sup>th</sup> (Monday NOT Sunday)**

# Academic Integrity & Course Recordings

Can discuss assignments with classmates, but all final work **MUST** be your own

Academic dishonesty of any kind may result in

- a 0 grade on the assignment
- a reduction in final grade
- and/or referral to the Dean

IIT code of Academic Honesty

- <https://www.iit.edu/student-affairs/student-handbook/fine-print/code-academic-honesty>

Class recordings are the intellectual property of the university or instructor and are reserved for use only by students in this class and only for educational purposes

Students **SHOULD** not post or otherwise share the recordings outside the class in any form

# What is Machine Learning?

Arthur Samuel: Machine Learning is a field of study that gives  
**computers the ability to learn** without being explicitly programmed

Shapire: Machine learning studies how to **automatically learn** to make predictions based on **past observations**

Kevin Murphy: Machine learning as a set of methods that can automatically **detect patterns in data**, and then use the uncovered patterns to **predict future data**

# Learning from Past Data/Observations



Future Data



Predict

"car"



Train

**Past Data**  
(perhaps with **labels**)

# Machine Learning Draws Inspiration & Concepts from Many Scientific Fields

**Statistics:** Inference from data, probabilistic models, learning theory...

**Mathematics:** Optimization theory, numerical methods, tools for theory...

**Engineering:** Signal processing, robotics, control, information theory...

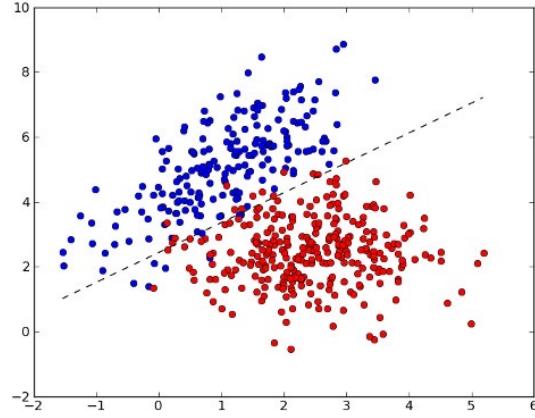
**Economics:** decision theory, operations research, econometrics...

**Psychology/Cognitive science:** Computational linguistics, learning, reinforcement learning, movement control...

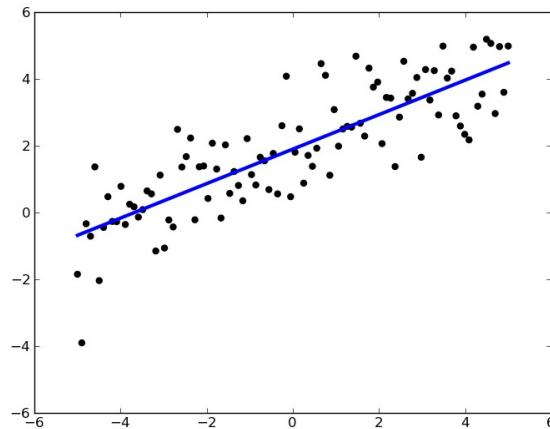
**Physics:** Energy minimization principles, entropy...

**Computational neuroscience:** Neural networks, principles of neural information processing, ...

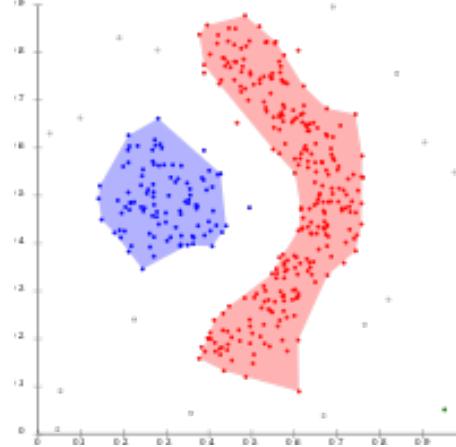
# Machine Learning Tasks



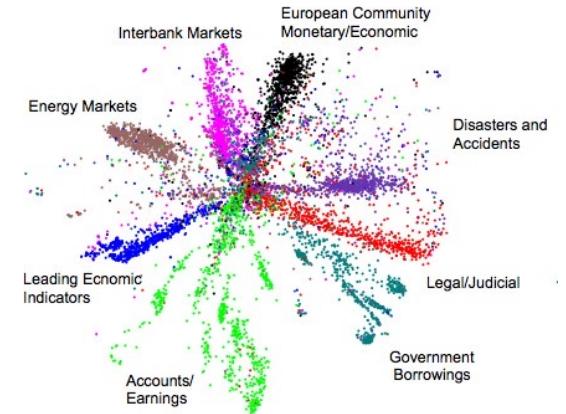
## Classification



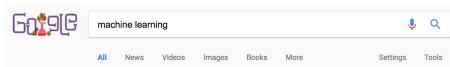
## Regression



## Clustering



## Dimensionality Reduction



About 289,000,000 results (0.55 seconds)

IBM Machine Learning - Watson APIs and Solutions

Discover New Opportunities, Fuel Growth & Beat Competitors.  
Brands: Watson Developer Cloud, Watson Internet of Things, Watson APIs, Watson SDKs  
What's Watson Working On? - IBM Watson Products - Search IBM Marketplace

Machine Learning Online Course - Udacity Nanodegree® Program

Learn To Build Machine Learning Models & Start Making Predictive Models  
Courses: Machine Learning, Intro To Programming, Android Developer, Self-Driving Car Engineer, Data...  
Want a high-paying job? Get a Nanodegree - CNED

Udacity For Businesses - View All Courses - Learn Self-Driving Cars - Nanodegree Programs

Adobe Sensei - Inspire Move Teach Transform - adobe.com

4.7 ★★★★★ rating for adobe.com

Adobe Sensei is the magic behind the world's best digital experiences.

AI for Marketing Cloud - Adobe PR News - Sensei - AI for Creative Cloud - AI for Document Cloud

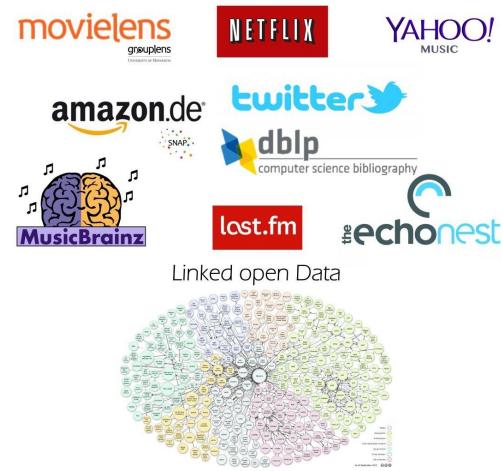
Machine Learning Guide - SAS Best Practices eBook - sas.com

Free Download! "The Machine Learning Primer". Get Your Copy Now!

Business Analytics - Statistical Analytics - Predictive Analytics

Machine learning - Wikipedia

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959 while at IBM. Machine learning - Outline of machine learning - Machine Learning (journal) - Torch



## Ranking

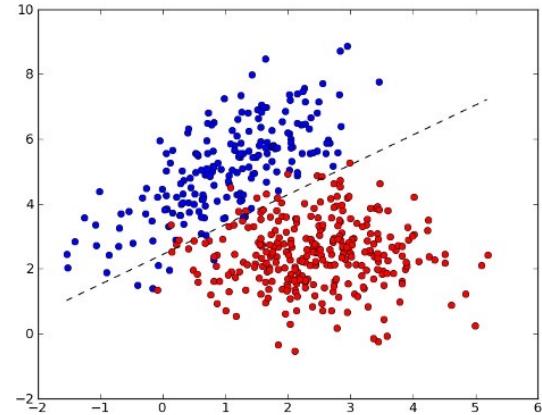


## Recommendation

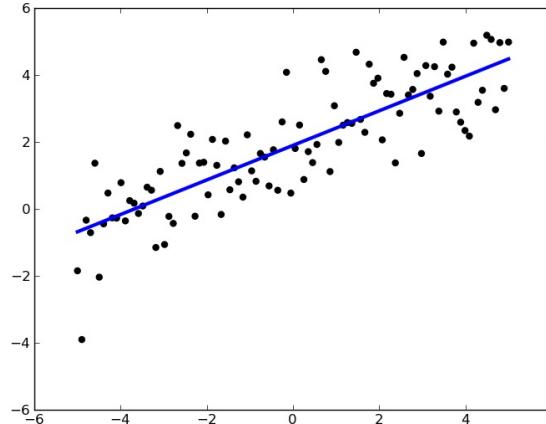


## Generative model

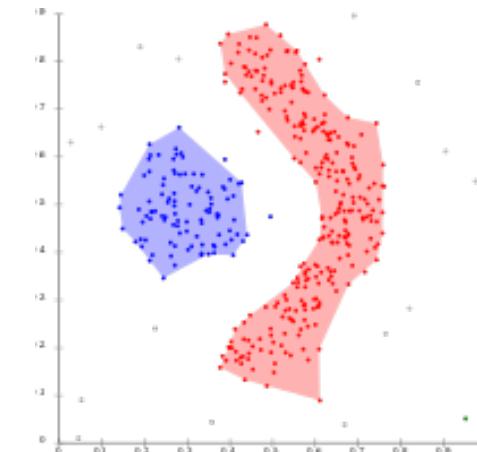
# Machine Learning Tasks



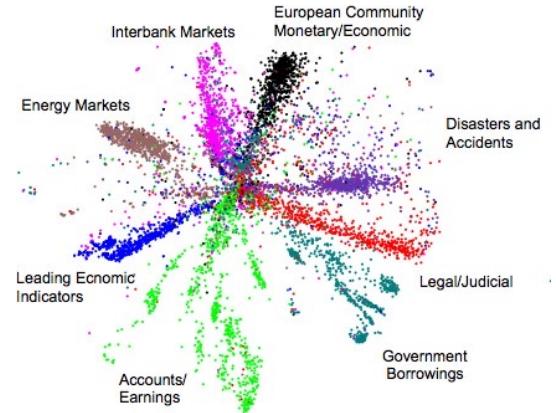
Classification



Regression



Clustering



Dimensionality  
Reduction

# Classification: Spam Detection

Screenshot of a Gmail inbox showing search results for "in:spam".

The search bar at the top shows "in:spam". The results list 1-50 of 133 messages.

A message in the list states: "Messages that have been in Spam more than 30 days will be automatically deleted. [Delete all spam messages now](#)".

The inbox sidebar shows the following counts:

- Inbox: 107
- Starred: 0
- Important: 0
- Sent: 0
- Drafts: 0
- Spam: 131
- Categories: 0
- Updates: 8,664
- Forums: 2,136
- Promotions: 19,344
- ML Mastery: 92
- Notes: 0
- 旅行相关: 0

The main list of 133 spam messages includes:

- IJECER Journal: Publish your paper without Publication Fee - Dear researcher, International Journal of Electrical and Computer Engineering Research (IJECER) is an academic journal that publishes research articles and r... (5:12 AM)
- IJECER Journal: Publish your paper without Publication Fee - Dear researcher, International Journal of Electrical and Computer Engineering Research (IJECER) is an academic journal that publishes research articles and r... (Jan 8)
- Home Chef: Oven-Ready meals with minimal prep - Enjoy \$100 Off! (Jan 8)
- HigherEdJobs Agent: 7 New Jobs - FacultyJobPosition - Your HigherEdJobs Agent for 01/08/2022 has returned 7 jobs that meet the search criteria you specified. 7 new jobs that match your criteria Visiting Assistant P (Jan 8)
- aws-marketi...@amazon...: Learn what the cloud can do for your institution - Join this webinar to learn about cybersecurity, data analytics, and more! Webinar – Higher Education: Thinking Out Cloud Tuesday, January 18, 2022 11:0... (Jan 7)
- Colonial Van Lines.: How to Fight Packing Procrastination - Fight Packing Procrastination View this email in your browser Call For Savings (888) 866-5606 Are you a procrastinator when it comes to packing? We totally get (Jan 7)
- HigherEdJobs Agent: 4 New Jobs - FacultyJobPosition - Your HigherEdJobs Agent for 01/07/2022 has returned 4 jobs that meet the search criteria you specified. 4 new jobs that match your criteria Full Time Faculty: C (Jan 7)
- ICMSSP Committees |: 🌟 ICMSSP' 22: The 7th International Conference on Multimedia Systems & Signal Processing 🌟 - ICMSSP 2021 | ACM | ISBN: 978-1-4503-9037-8 | EI Compendex & Scopus 2022 7th International Confer... (Jan 7)
- Blue Nile: Explore Studs & Play It By Ear This Year! - Shop Classics Earrings & More! View web version DIAMONDS ENGAGEMENT Jewelry GIFTS Whether you're looking to impress your dinner guests or a Zoom co... (Jan 6)
- Colonial Van Lines.: Moving Out of State? Save \$100 with Colonial's Moving App - Moving App View this email in your browser CALL FOR SAVINGS! (888) 866-5606 ARE YOU MOVING OUT OF STATE? Save \$100 with Colonial... (Jan 6)
- LinkedIn: You appeared in 7 searches this week - You appeared in 7 searches this week You appeared in 7 searches this week You were found by people from these companies See all searches This email was inte... (Jan 6)

# Classification: Spam Detection

Sir / Madam,

We invite you to submit your manuscript(s) for publication. The journals include research papers, review articles, technical projects and short communications containing new insight into any aspect of the covered scope of the journal. Our objective is to inform authors of the decision on their manuscript(s) within weeks of submission. After acceptance, the paper will be published in the current issue immediately.

**Keywords:** English, Literature, Science, Economics, Engineering, Management, Agriculture, Horticulture, Environment .....

[International Journal of Advanced Engineering Research and Science \(IJAERS\)](#) ISSN: 2456-1908(O) | 2349-6495 (P)

DOI (CrossRef): [10.22161/ijaers](https://doi.org/10.22161/ijaers)

Thomson Reuters ResearcherID: P-3738-2015

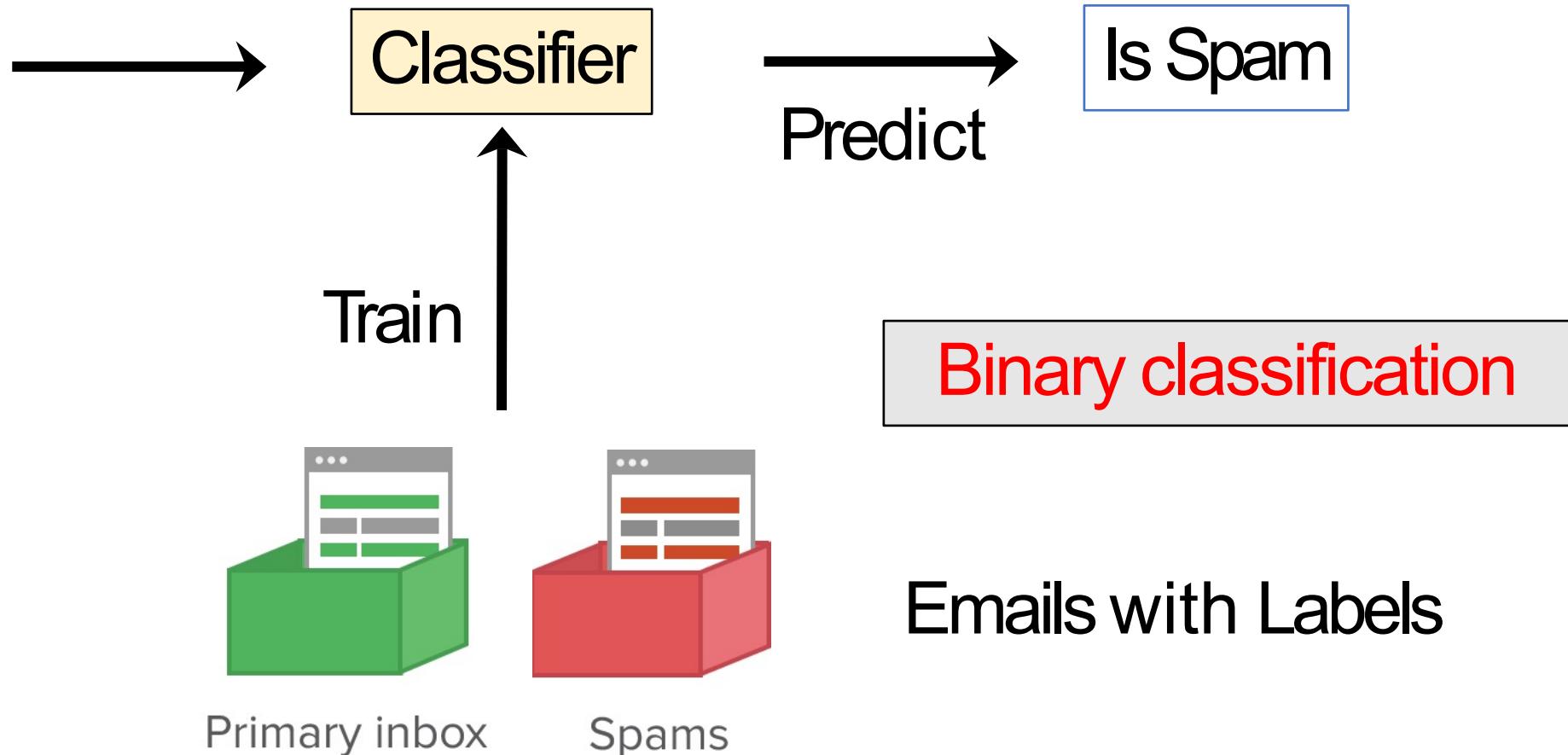
Impact Factor: 4.192, SJIF: 4.072, IBI: 3.2, PIF: 2.465, ISRA-JIF: 1.317,

Website: <http://www.ijaers.com>

Kindly submit research articles to <http://ijaers.com/submit-paper/> or mail us at [editor.ijaers@gmail.com](mailto:editor.ijaers@gmail.com)

[International Journal of English, Literature and Science \(IJELS\)](#)  
ISSN: 2456-7620

## New Email



# Classification: Face Recognition



Train



multi-class

Faces & Names

# Classification: Image Recognition



Classifier



Predict

“Dog”  
“Woman”

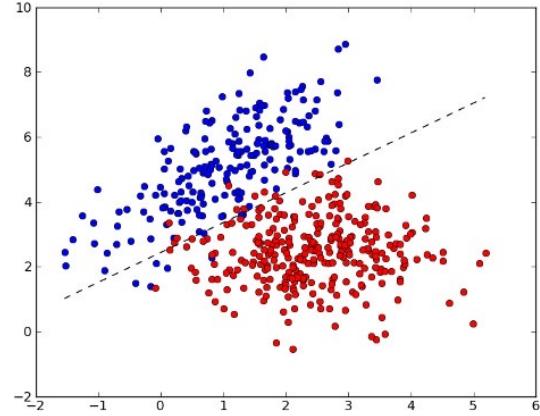


Train

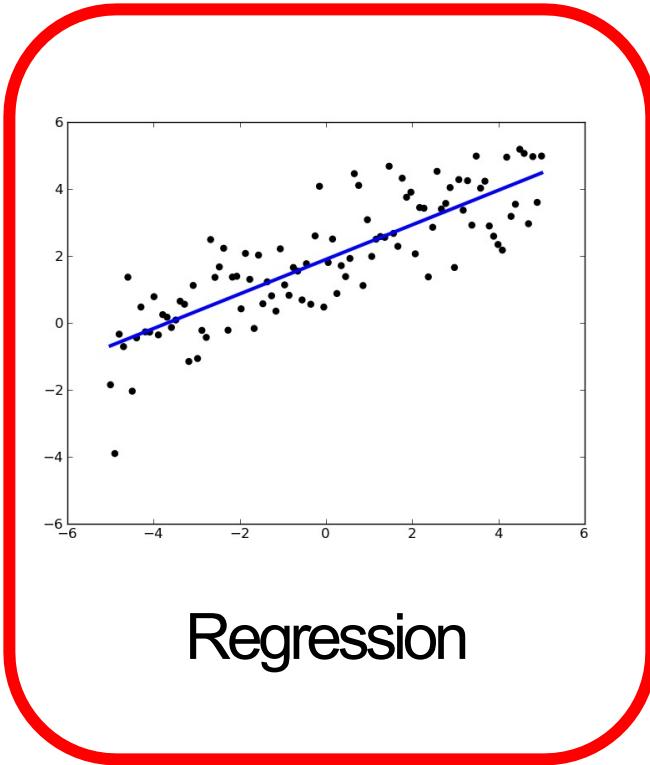
multi-class, multi-label

Images & Labels

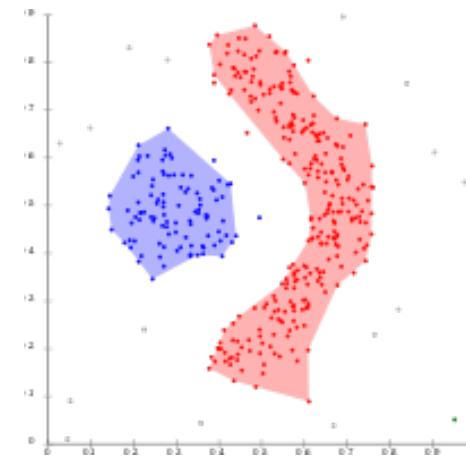
# Machine Learning Tasks



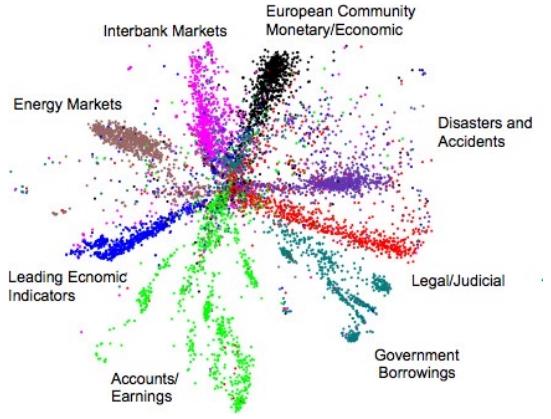
Classification



Regression



Clustering



Dimensionality  
Reduction

# Regression: Housing Price



Features of a House



# Regression VS. Classification

- Regression: labels are continuous and ordered

House Prices:

\$324K

<

\$521K

<

\$1.2M



# Regression VS. Classification

- **Classification: labels are categorical**

**Categories:**

Class1

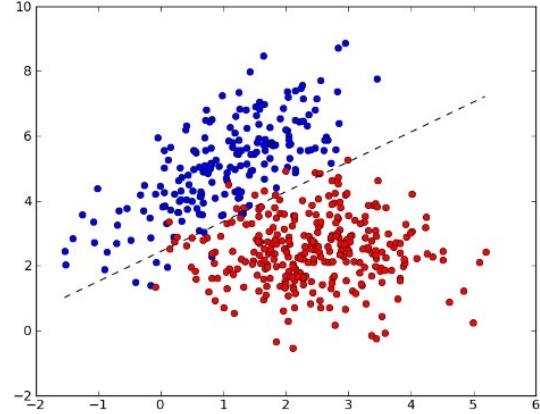
Class 2

Class 3

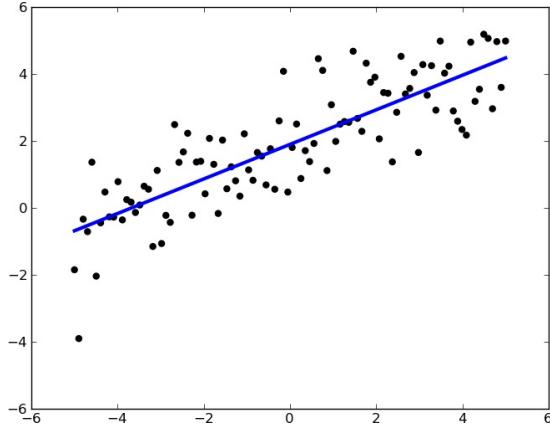


“Class 1” and “Class 3” are not ordered and cannot be compared!

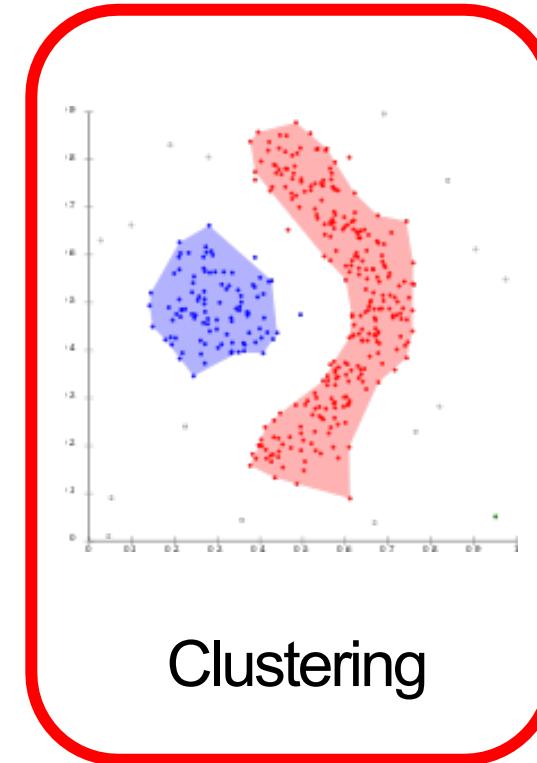
# Machine Learning Tasks



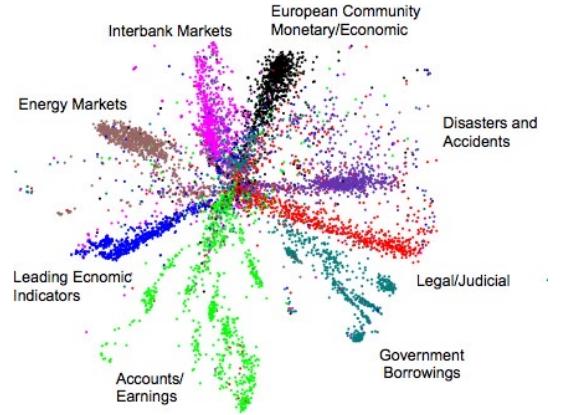
Classification



Regression



Clustering



Dimensionality  
Reduction

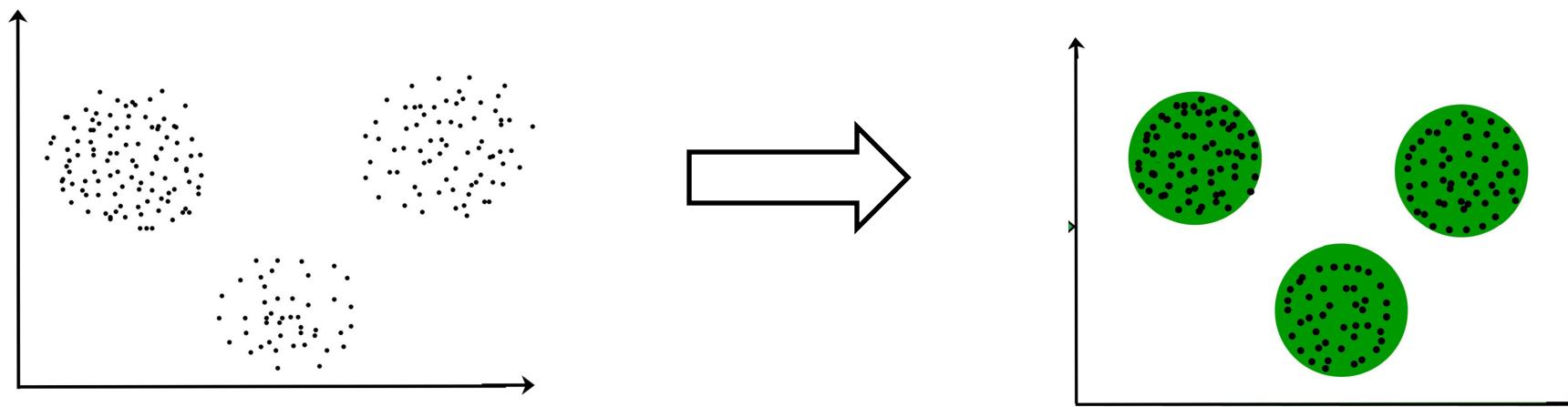
# Clustering

**Task:** divide a set of data points into a number of groups such that

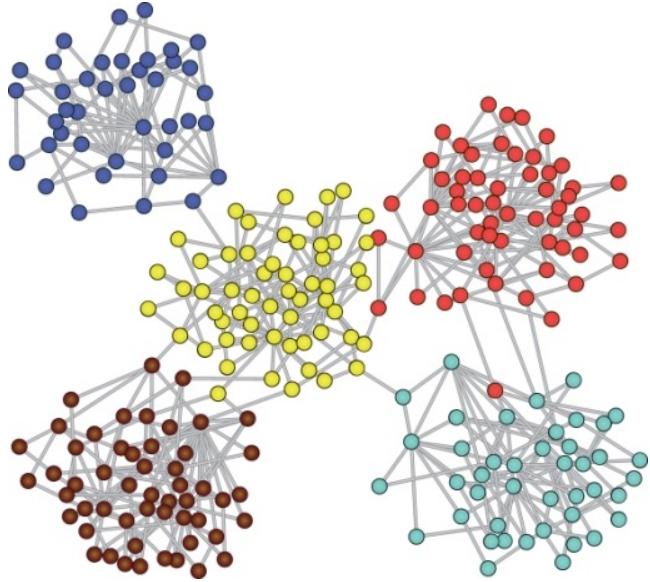
- data points in the same groups are similar to each other
- data points from different groups are dissimilar to each other

**Input:** data points  $x_1, x_2, \dots, x_n$

**Output:** predict clusters  $y_1, \dots, y_n \in \{1, \dots, k\}$ ,  $k$  is #clusters



# Clustering: Applications



Community detection

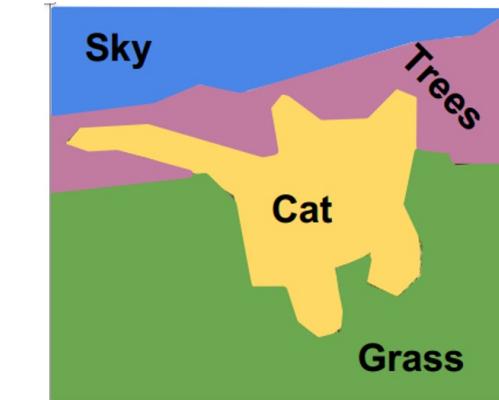
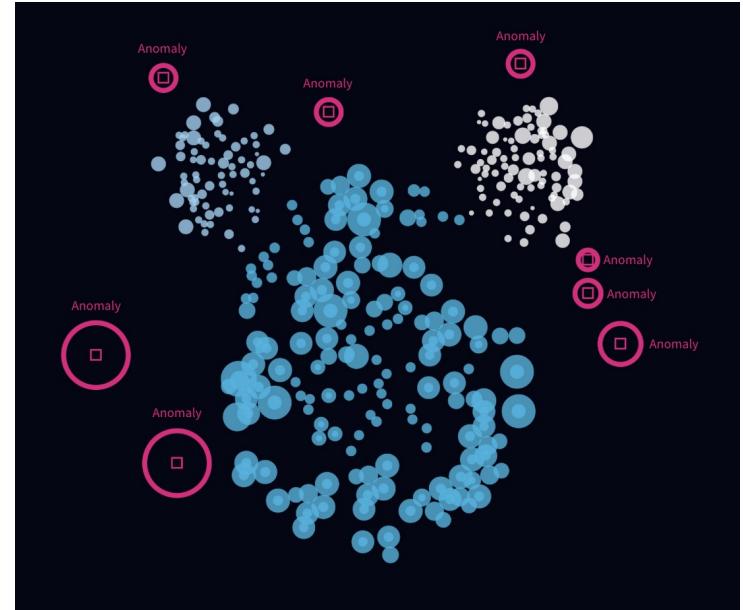
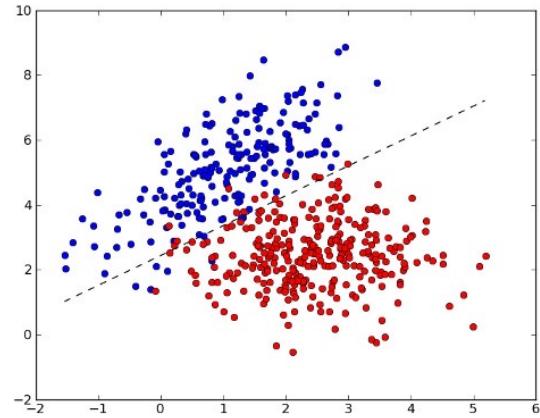


Image segmentation

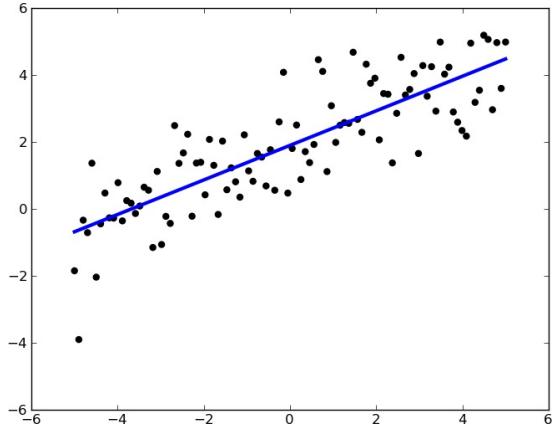


Anomaly detection

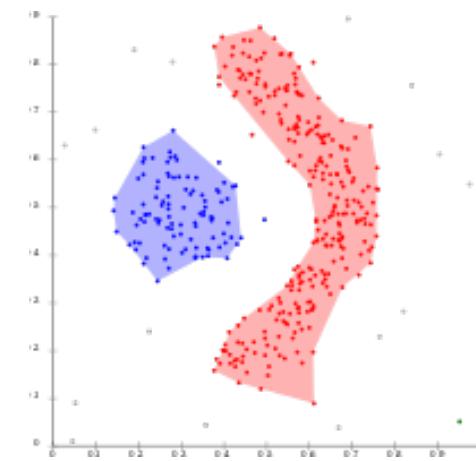
# Machine Learning Tasks



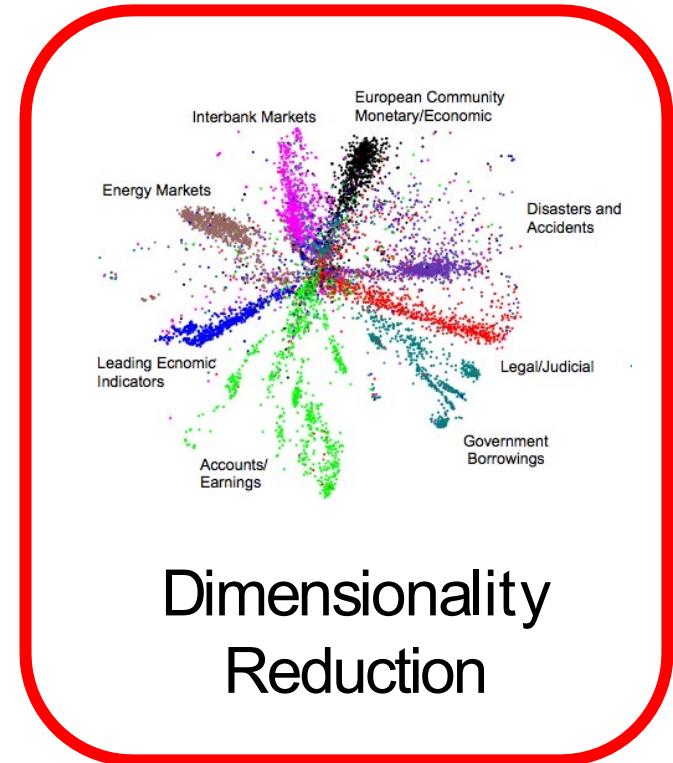
Classification



Regression



Clustering



Dimensionality  
Reduction

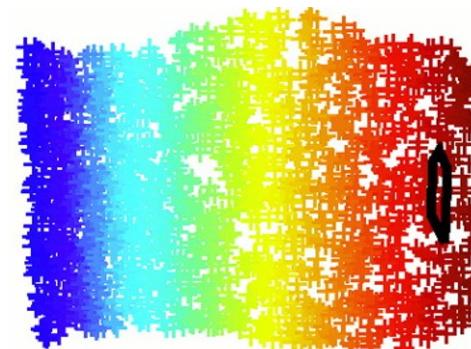
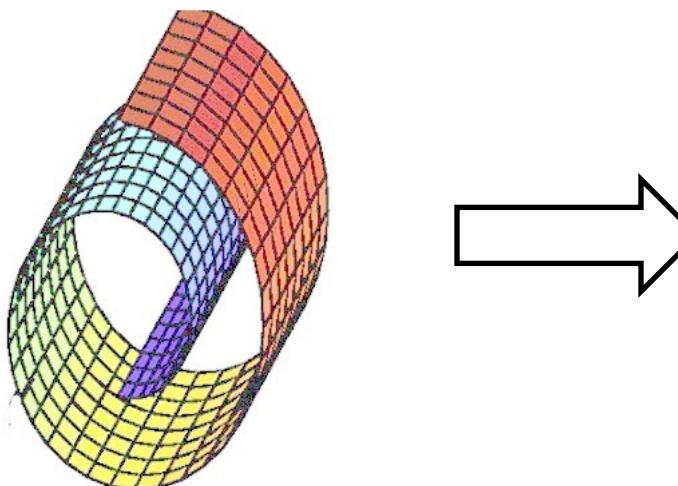
# Dimensionality Reduction

**Task:** map high-dim data points into a low-dim space such that

- Global information of data points (e.g., variance) can be kept
- Local information of data points (e.g., neighborhood) can be kept

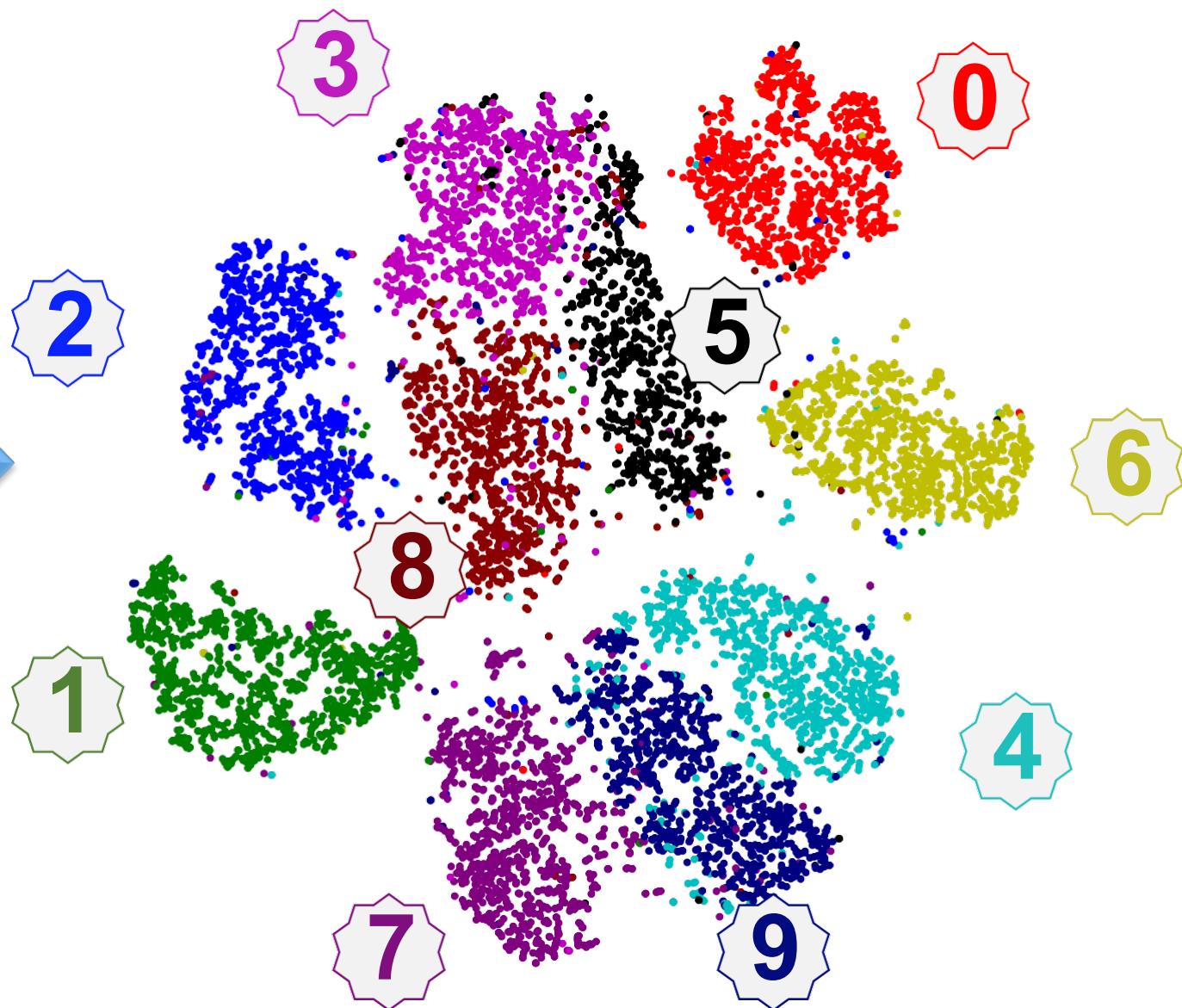
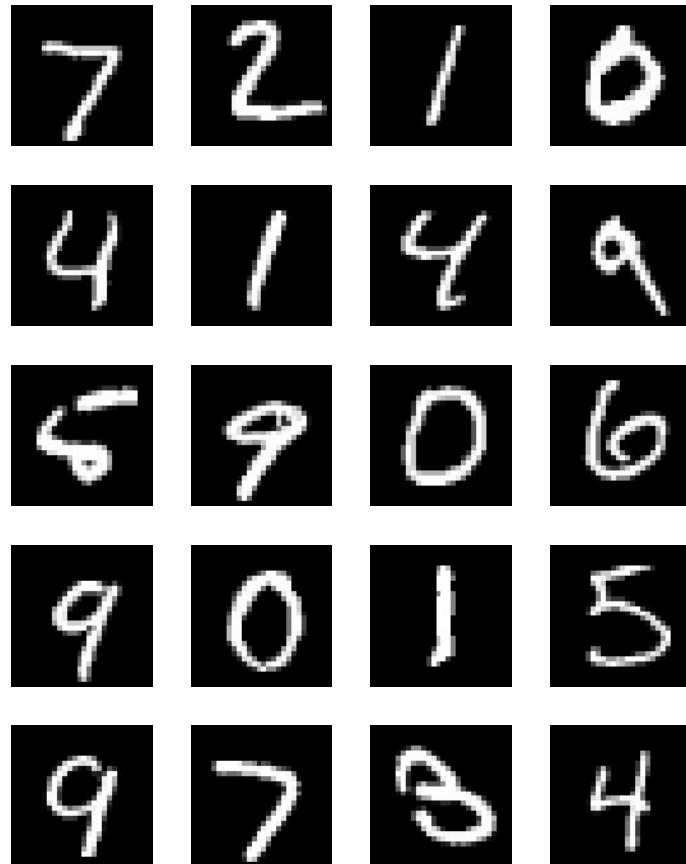
**Input:** high-dim data points  $x_1, x_2, \dots, x_n$

**Output:** low-dim data representations  $y_1, y_2, \dots, y_n$



Visualization  
Data compression  
Data preprocessing

# Visualization: Handwritten Digit Data



60K images (size 28×28)

# Intelligent Transportation Systems (ITS)

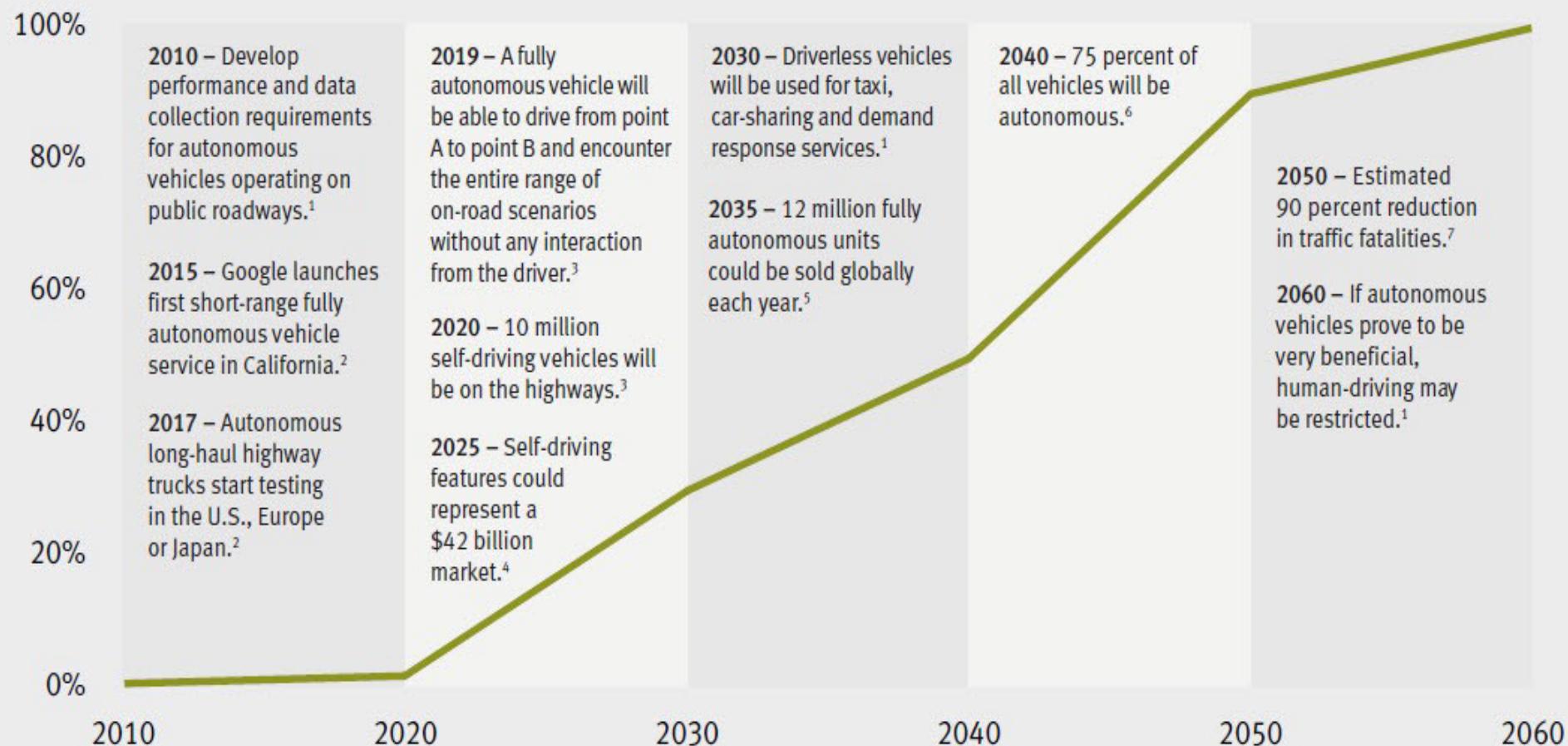


# Autonomous Vehicles

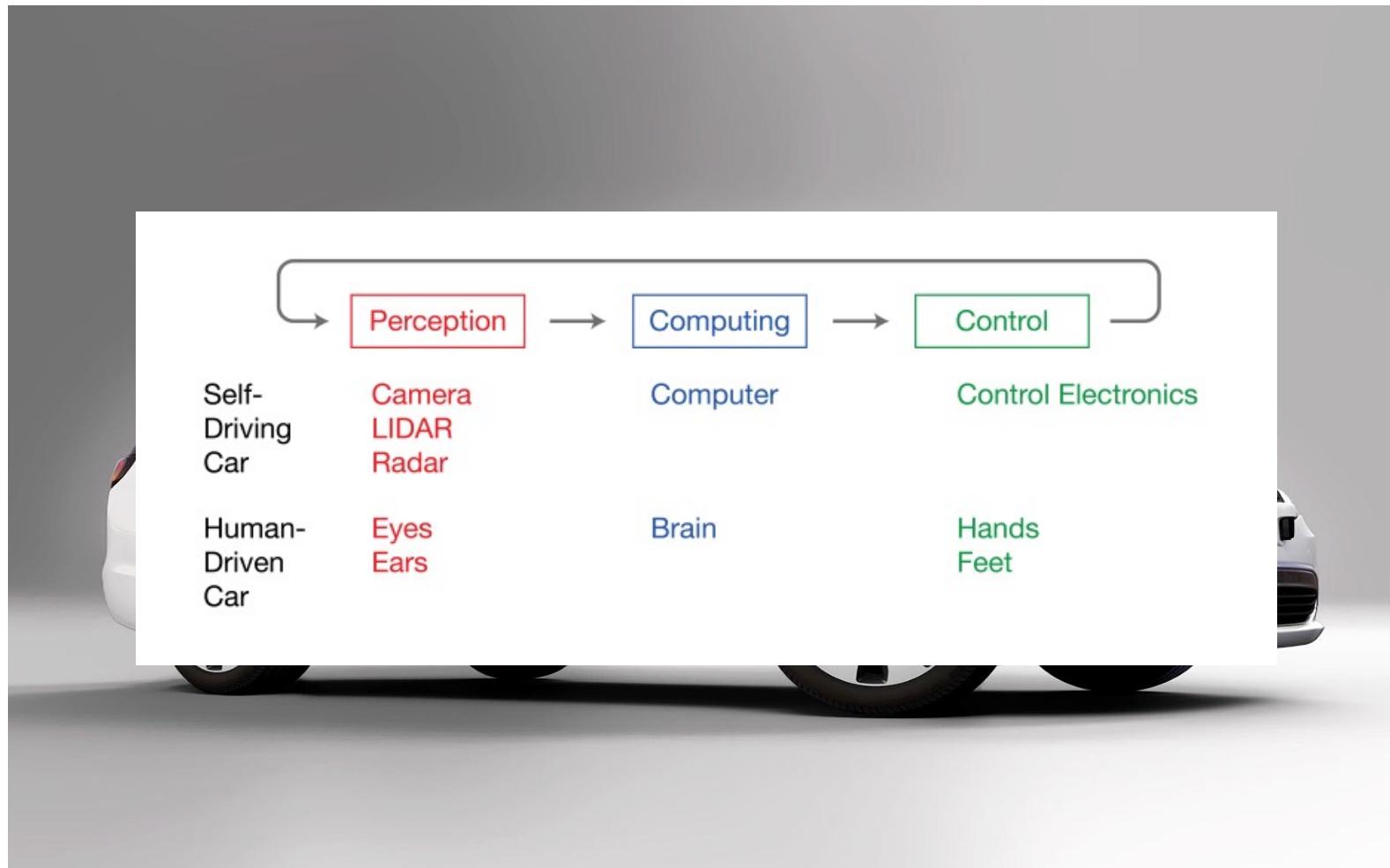


# Autonomous Vehicles

## Estimated Percentage of Autonomous Vehicle Adoption, and Key Milestones



# Autonomous Vehicles



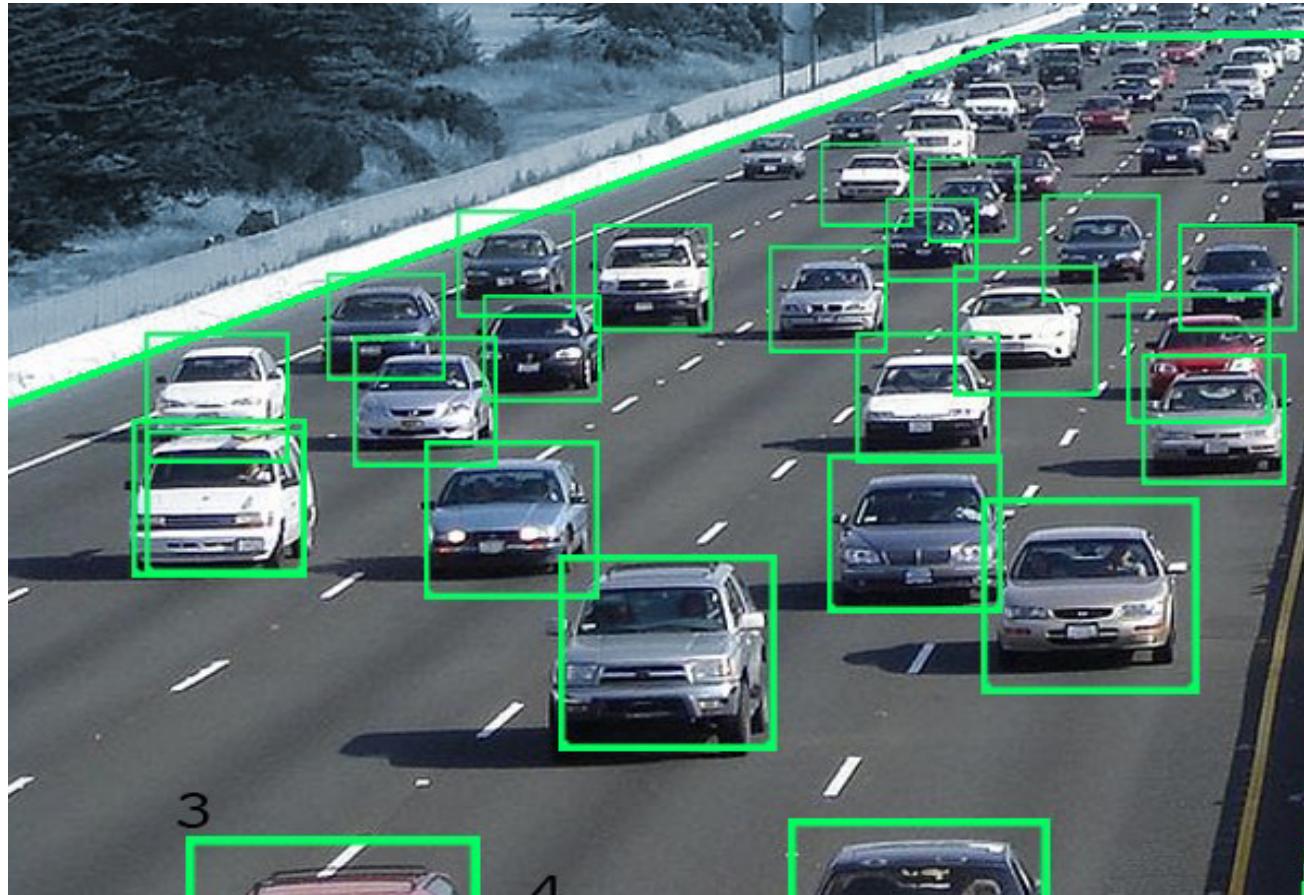
# Machine Learning for ITS

A primary use of ML in ITS is for intelligent perception

Some key tasks

- Object detection
- Multi-object tracking
- Activity recognition

# Object Detection

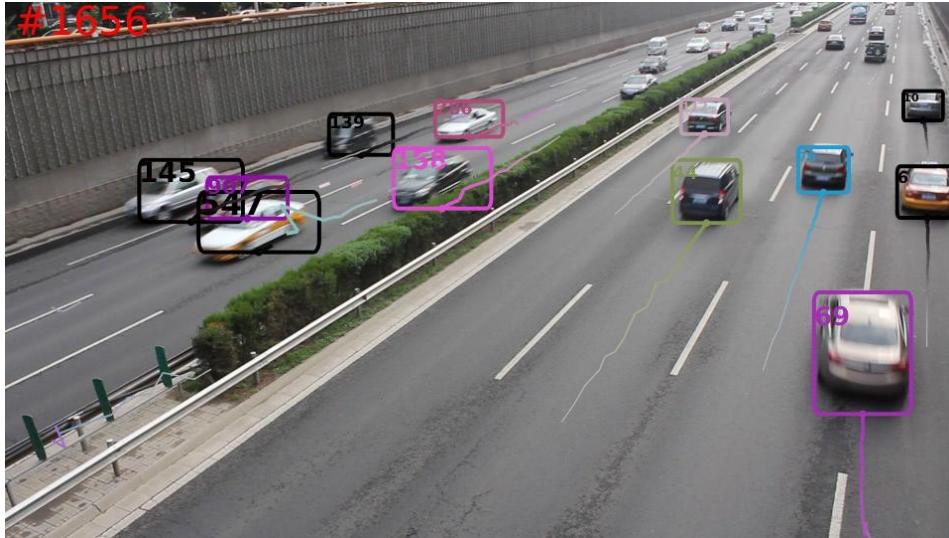


Where are the interesting objects within my field of view?

What are the object classes (pedestrian, bicyclist, sedan, ...)?

# Multi-Object Tracking (MOT)

The goal is to estimate the trajectories of all objects in a dynamic scene



MOT from a stationary traffic cam



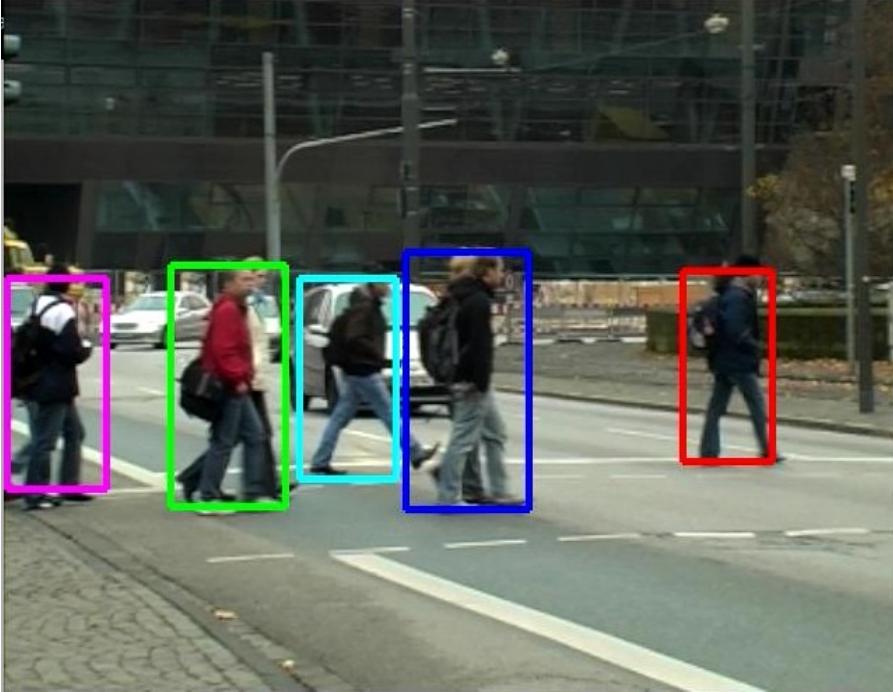
MOT using LiDAR from an AV

# Activity Recognition

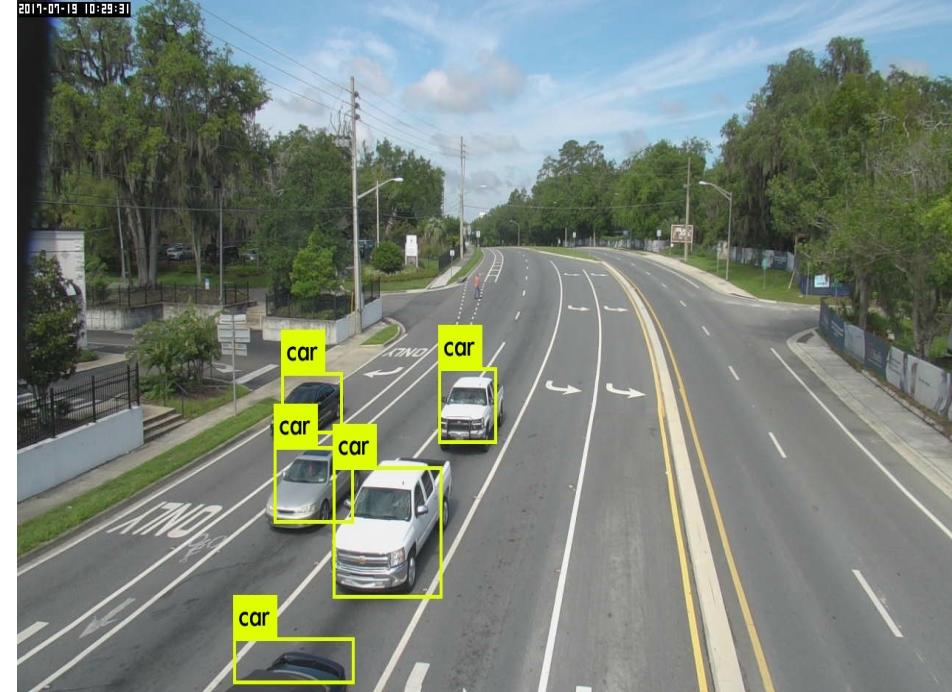
Using object detections and trajectories, can we then extract patterns at the level of behaviors?

- Pedestrian safety
  - whether a person is walking/about to walk into the street
- Vehicle collision prediction

# Pedestrian/Car Safety

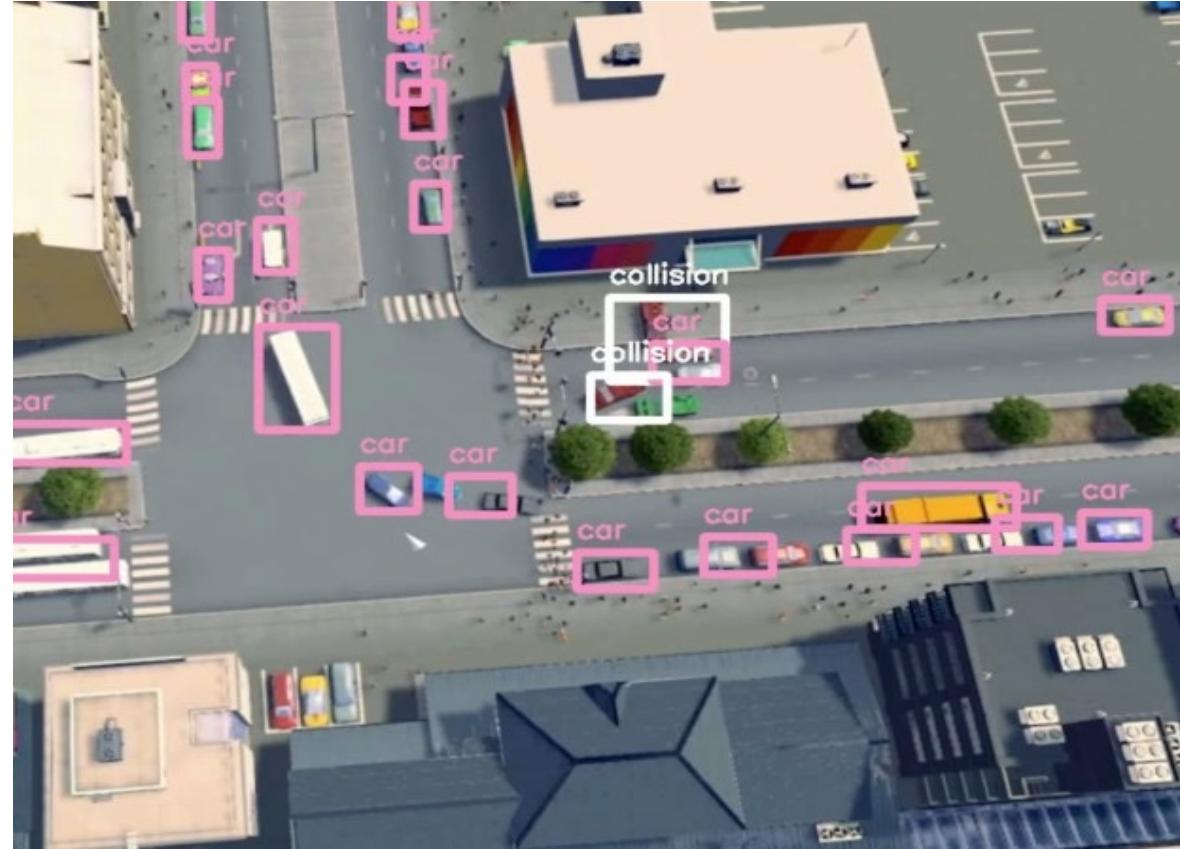


Are pedestrians crossing?



Any driving the wrong way?

# Collision Prediction



# Machine Learning Tasks & Methods

# Tasks

# Methods

Regression

Classification

Clustering

Dim Reduction

# Tasks

# Methods

Regression

Classification

Clustering

Dim Reduction

# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

Linear Regression

Polynomial Regression

Ridge Regression

LASSO

Elastic Net



# Tasks

# Methods

Regression

Classification

Clustering

Dim Reduction

# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

Logistic Regression

Naïve Bayes

Neural Networks

SVM

Nearest Neighbor

Decision Tree / Boosting



# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

DBSCAN

K-means

GMM

Spectral Clustering

Affinity Propagation

Sparse sub. Clustering



# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

PCA/SVD

Nonnegative MF

LLE/LE

Deep Belief Net

Random Projection



# Supervised vs. Unsupervised Learning

**Supervised learning** : learn from **labeled** data

- Classification and regression are supervised learning tasks

**Unsupervised learning** : learn from **unlabeled** data

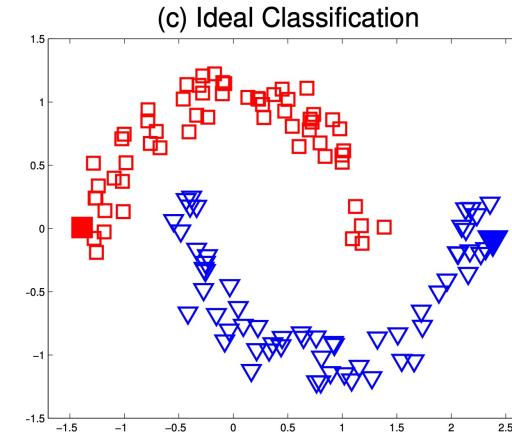
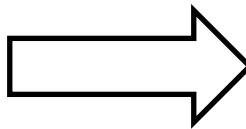
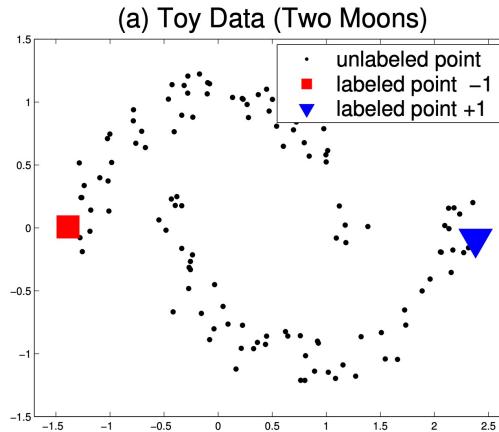
- Clustering and dimensionality reduction are unsupervised learning tasks

# Semi-supervised Learning

**Supervised learning** : learn from **labeled** data

**Unsupervised learning** : learn from **unlabeled** data

**Semi-supervised learning** : learn from both **labeled** and **unlabeled** data



# Why/When Semi-supervised Learning?

Labeled data is insufficient/expensive, unlabeled data is sufficient/cheap

- Imagine the cost/time of tagging millions of images!

Speech analysis

- Switchboard dataset
- **400 hours** annotation time for 1 hour of speech

Natural Language Processing

- Penn Chinese Treebank
- **2 years** of 4,000 sentences

# Semi-supervised Learning Methods

- Self-Training
- Co-Training
- Semi-supervised SVM
- Generative methods, mixture models
- Mixup & MixMatch
- Graph-based methods
  - Label propagation
  - Belief propagation
  - Graph neural networks

# Semi-supervised Learning Methods

- Self-Training
- Co-Training
- Semi-supervised SVM
- Generative methods, mixture models
- Mixup & MixMatch
- Graph-based methods
  - Label propagation
  - Belief propagation
  - Graph neural networks

# Project Lists

## Deep learning theory

- Information bottleneck, Neural Tangent Kernel, ReduNet, Causality

## Unsupervised representation learning

- Disentangled representation learning, causal representation learning, contrastive learning, self-supervised learning, variational methods

## Semi-supervised learning

- Graph neural networks, graph embedding, mixup, self-training, co-training

## Trustworthy machine learning (DNN, graph learning, federated learning, etc.)

- Security attacks (evasion, data/model poisoning, backdoor);
- Empirical defense (adversarial training, robust optimization); Provable defense (randomized smoothing, IBP)

## Privacy-preserving machine learning (DNN, graph learning, federated learning, etc.)

- Privacy attacks (model stealing/inversion, property/attribute inference)
- Privacy preserving (differential privacy, Crypto, information theory)

## Machine learning for security

- Blockchain security, Network security, Software/hardware security, Cyber-Physical System security

# Project Lists (Continue)

Large-scale machine learning (High-dimensional/massive data)

- Randomized algorithms
- Streaming, sketching, compressive sampling, ...

Federated learning

- Communication efficient, computational-efficient, personalization, fairness

Meta learning

- Model-agnostic meta learning

Interpretable machine learning

Machine unlearning

(Deep) compressive sensing/sparse coding/dictionary learning

Generative AI

Large language model (LLM)

Other topics you may be interested in...

# Tasks

Regression

Classification

Clustering

Dim Reduction

# Methods

Linear Regression

Polynomial Regression

Ridge Regression

LASSO

Elastic Net

# Acknowledgement

Some slides are from **Shusen Wang**

<https://github.com/wangshusen/DeepLearning>