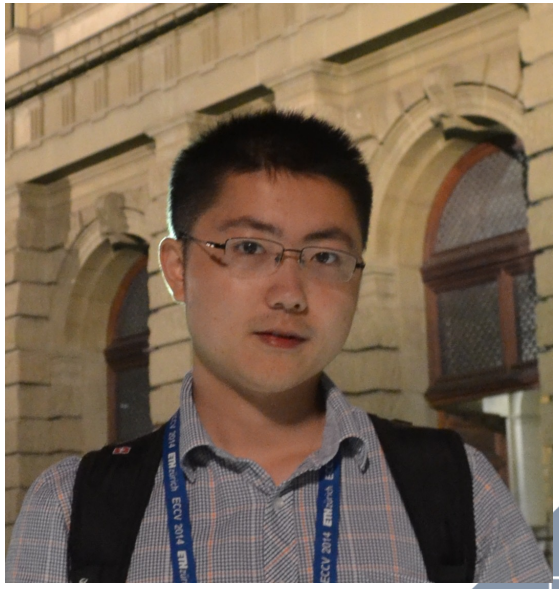
CSE 176 Introduction to Machine Learning
Lecture 1: Course Introduction

Some materials from Miguel Carreira-Perpiñán and Pascal Poupart

About the instructor

- Meng Tang, Assistant Professor in Computer Science and Engineering
- Please call me Meng (No Prof. Dr.)
- 10+ years of experience in machine learning and computer vision
- Ex-Meta, Amazon, Borealis AI, Disney
- www.mengtang.org



UCMERCED

What is Machine learning?

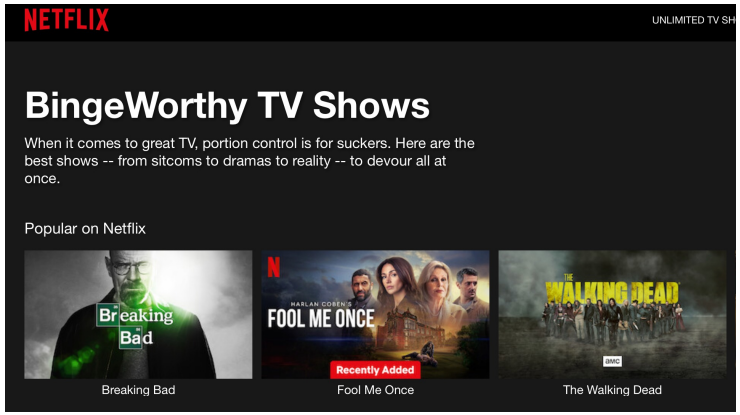
- Traditional computer science
  - Program computer for every task

- Machine Learning
  - Arthur Samuel (1959): **Machine learning** is the field of study that gives computers the ability to learn without being explicitly programmed.
  - The capability of a machine to learn from data to imitate intelligent human behavior.

Applications of Machine learning

Autonomous driving

Speech recognition



Recommendation system

UCMERCED

Machine learning ≠ Artificial Intelligence

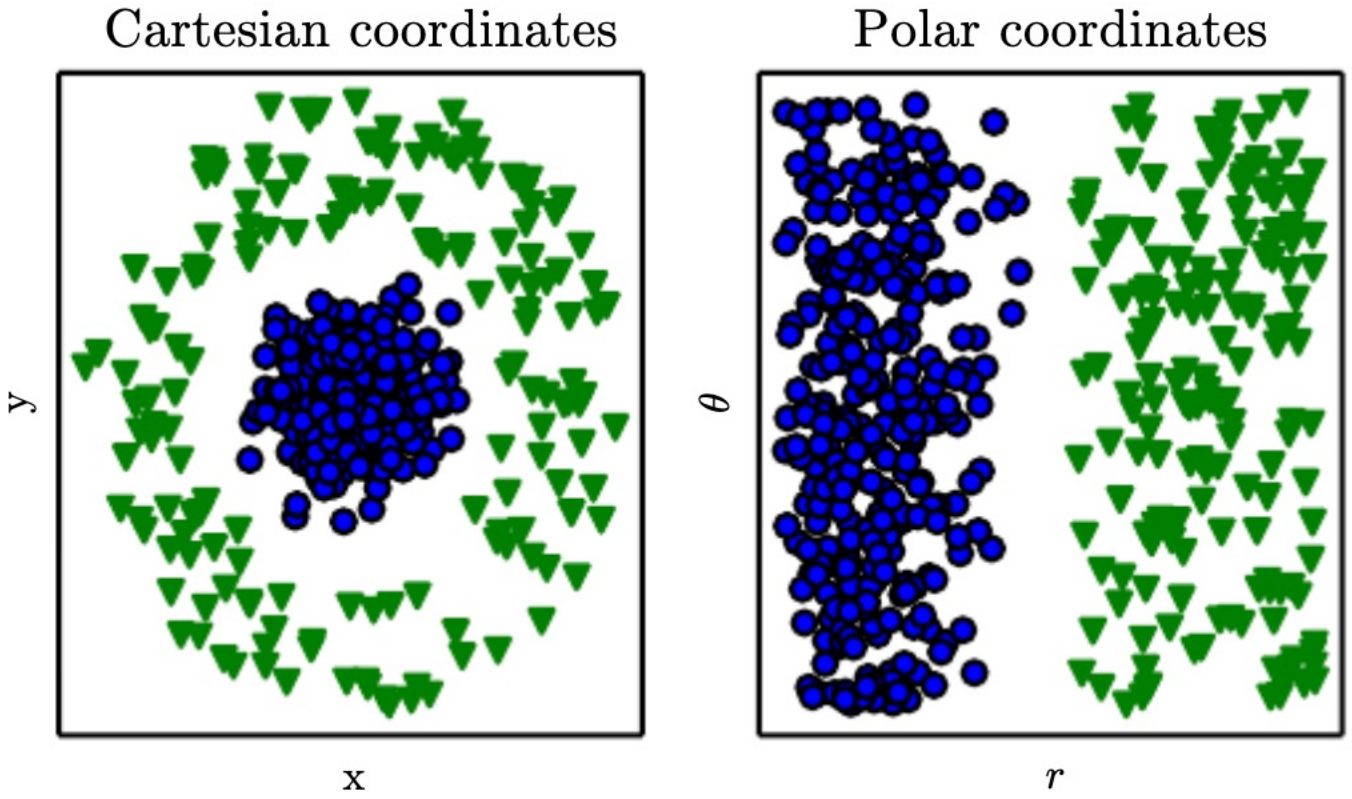
- Machine learning is a specific approach within the broader field of artificial intelligence.
- Machine learning specifically involves the use of algorithms and statistical models for learning from data.
- AI encompasses a wider range of techniques, including
  - rule-based systems
  - expert systems
  - machine learning
  - and more

Machine learning ≠ Artificial Intelligence

- IBM DeepBlue (1997)
- Chess game
- sophisticated algorithm, expert-system, and brute-force computation
- Not machine learning

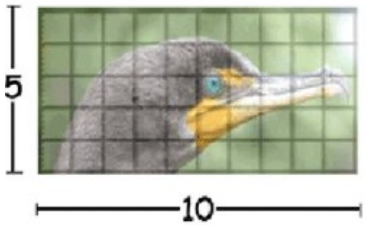
- Deepmind AlphaGo (2016)
- Go game (more complex)
- Use deep neural networks
- Trained on datasets of expert go game

Examples of representations

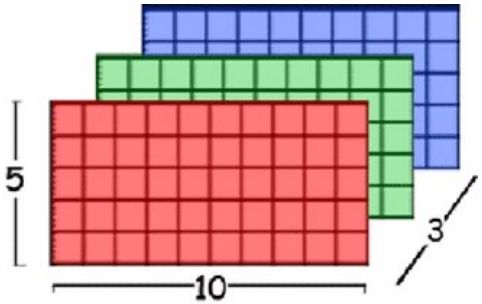


How to represent raw sensory data?

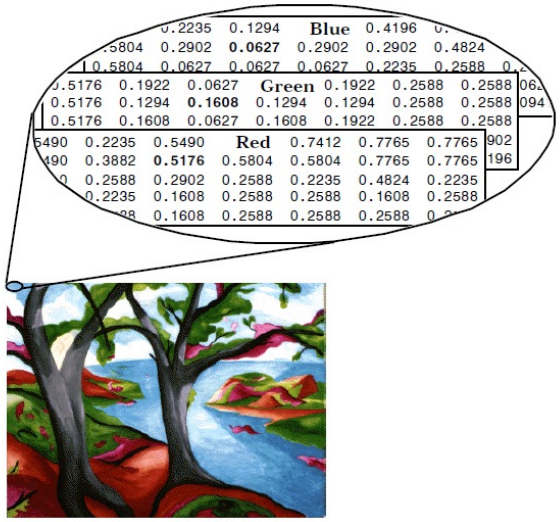
- ❑Cannot hand-craft representation
- ❑Representation learning



Original Color Image

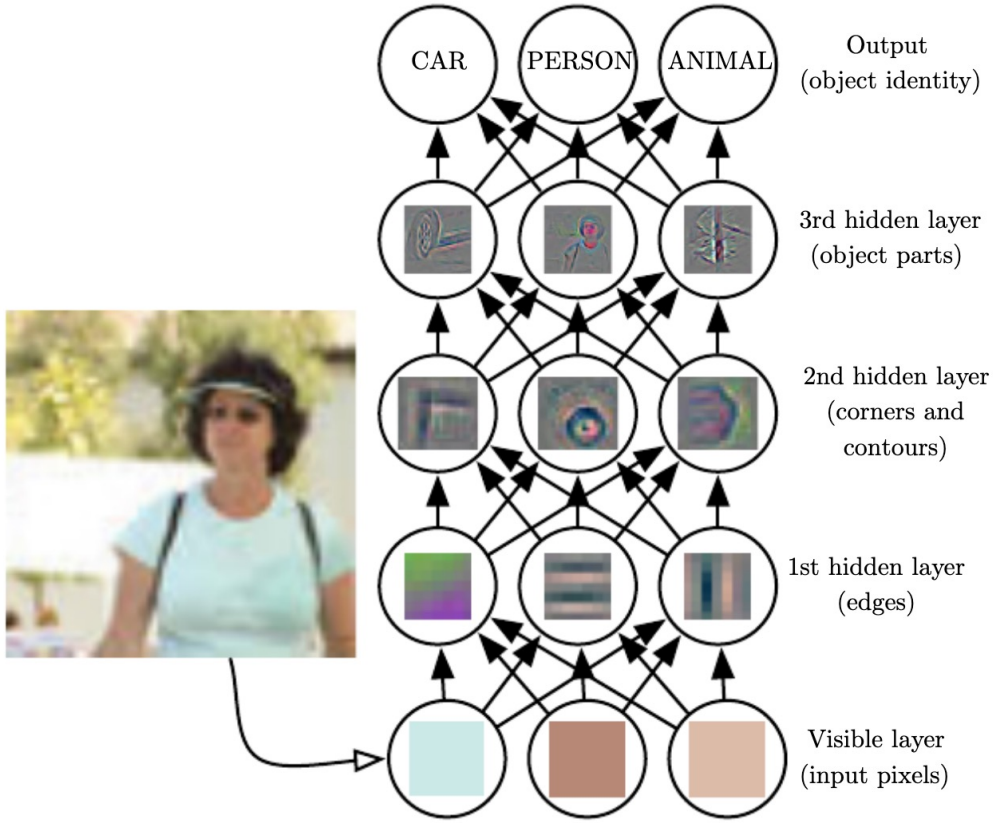


Matlab RGB Matrix

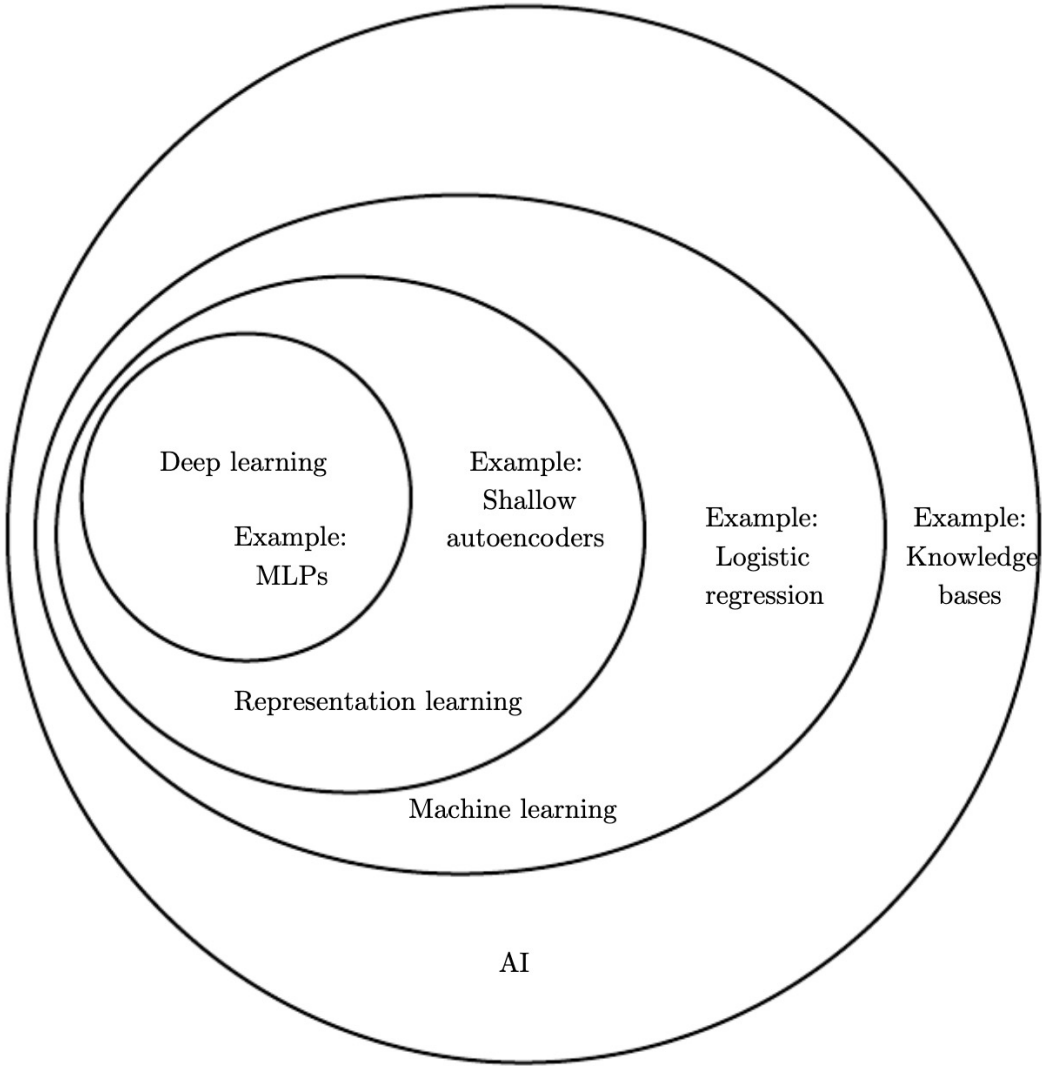


A deep learning model

- Deep neural networks learn hierarchical representation

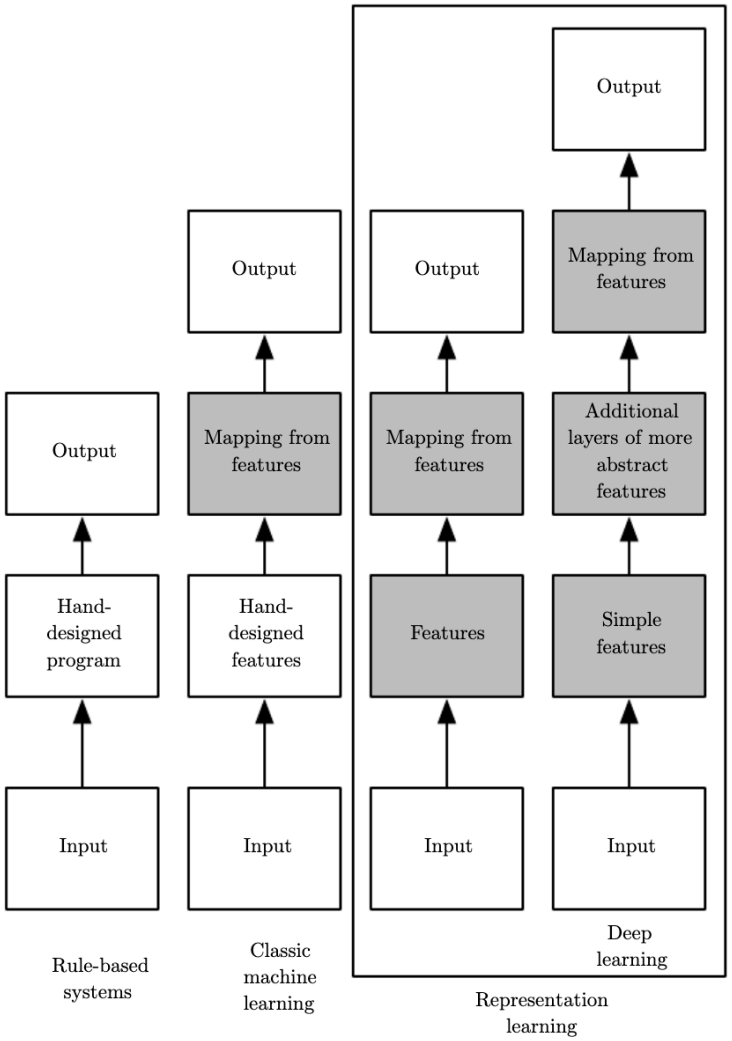


Deep Learning



UCMERCED

Different AI systems



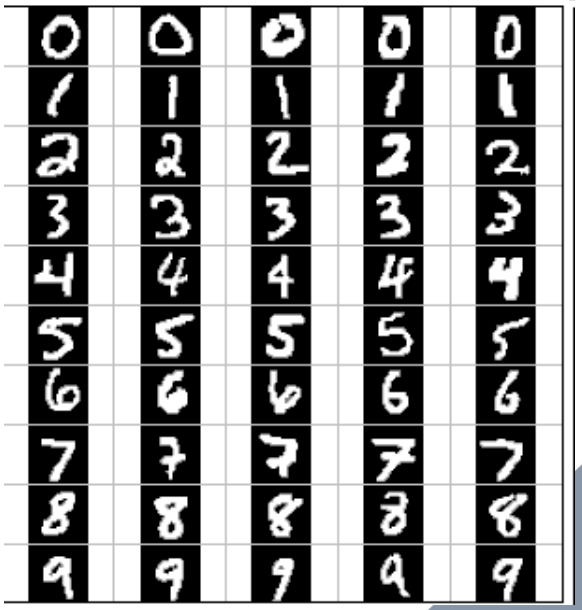
Major Types of Machine Learning

- ❑Supervised Learning
- ❑Unsupervised Learning
- ❑Reinforcement Learning

Supervised Learning

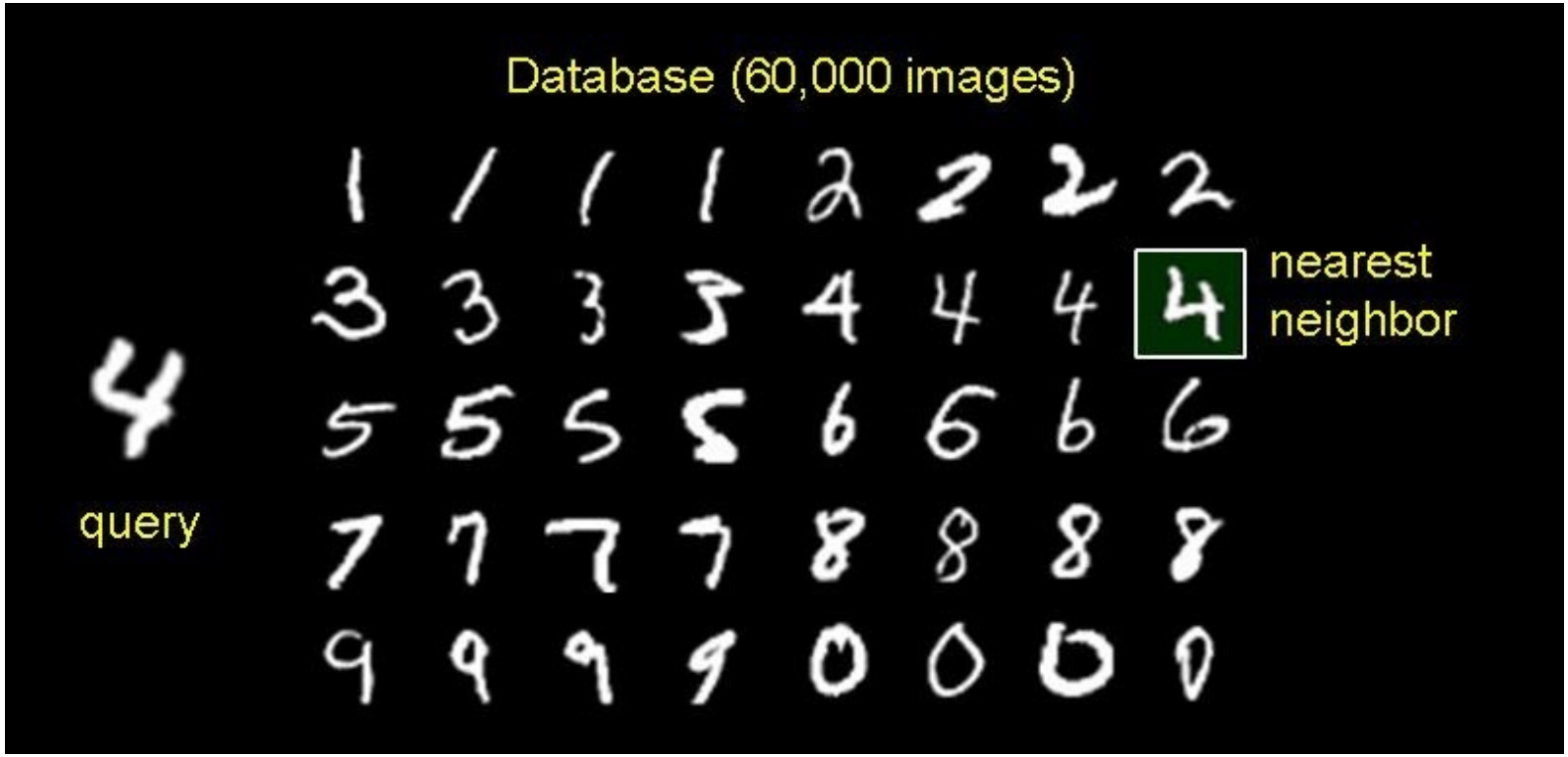
- Given pairs of input-output, learn to map the input to output
  - Image classification
  - Speech recognition
  - Regression (continuous output)
- Example: Digit recognition

- Simplest approach: Memorization



Supervised learning

- Nearest neighbor

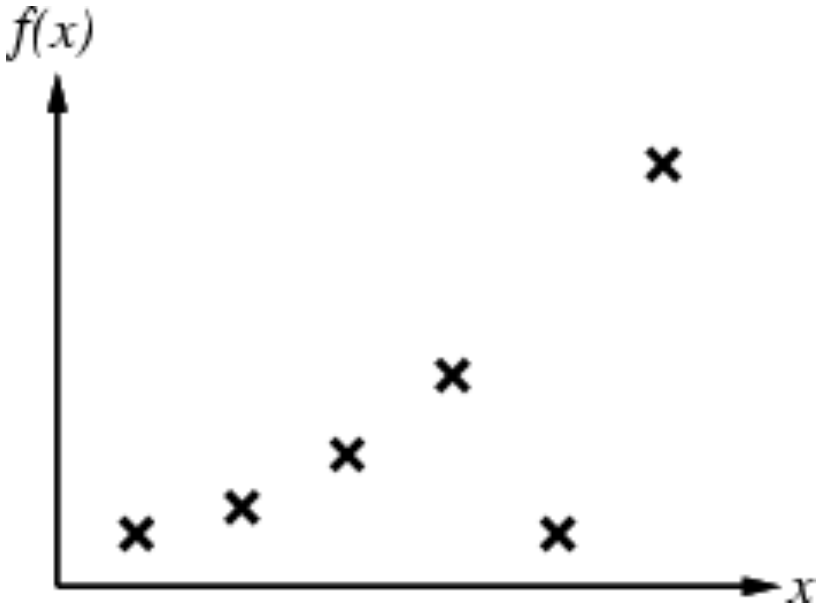


More formally

- Supervised learning: Given a training set of examples of the form $(x, f(x))$
  - x is the input, $f(x)$ is the output
- Return a function h that approximates f
  - h is called the hypothesis

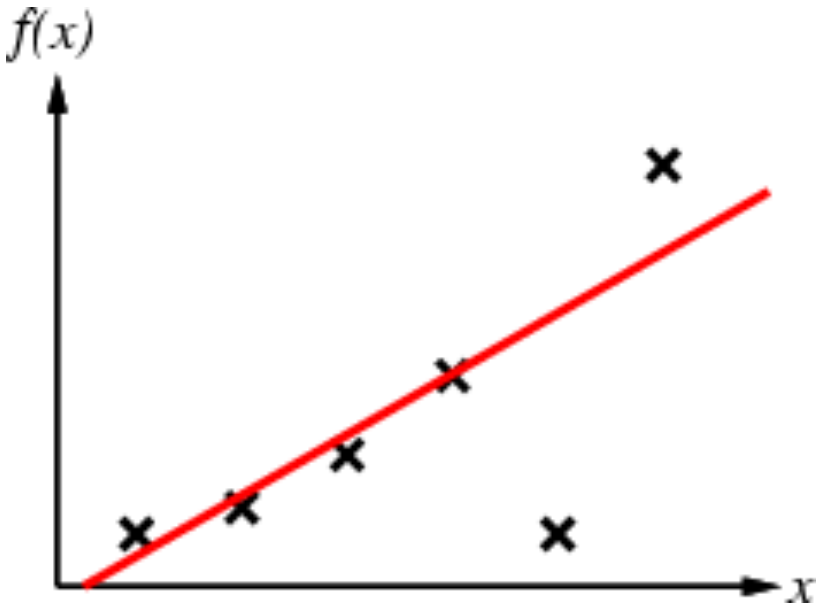
Prediction

- Find function h that fits f at instances x:



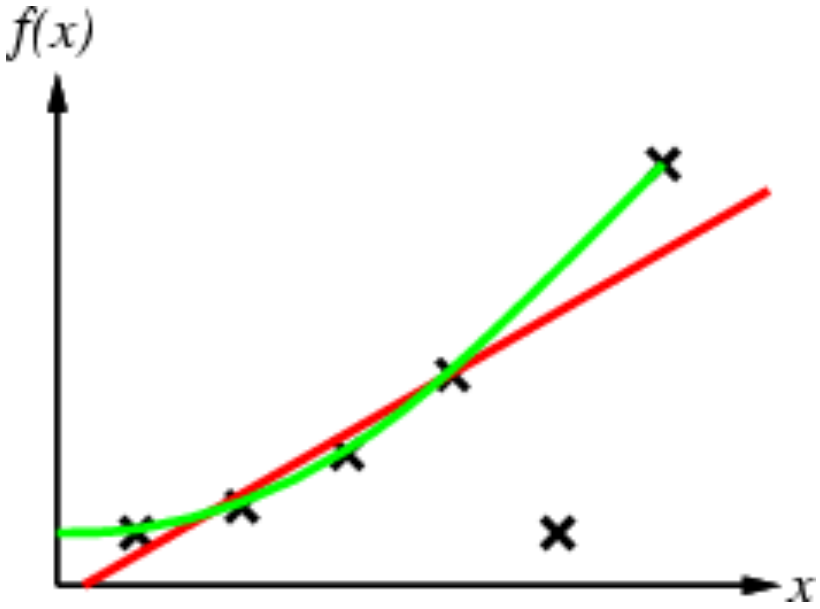
Prediction

- Find function h that fits f at instances x:



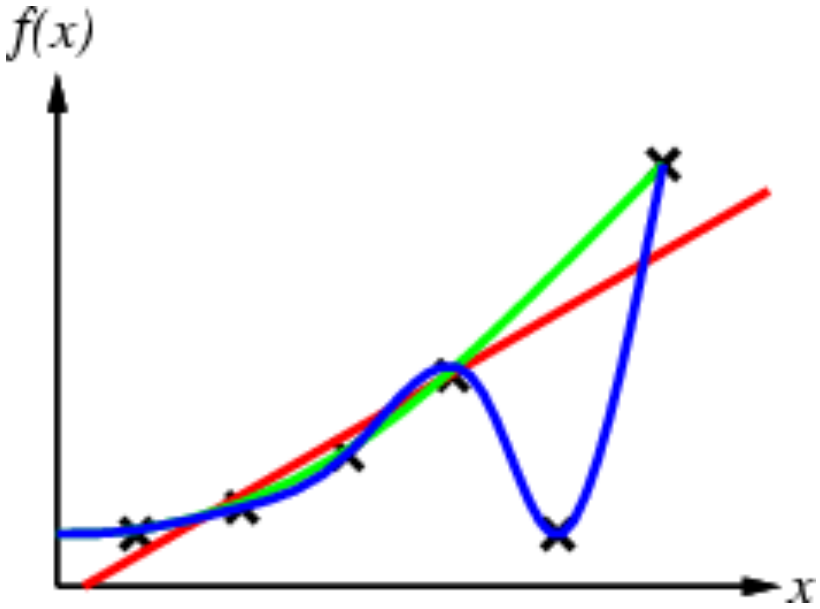
Prediction

- Find function h that fits f at instances x:



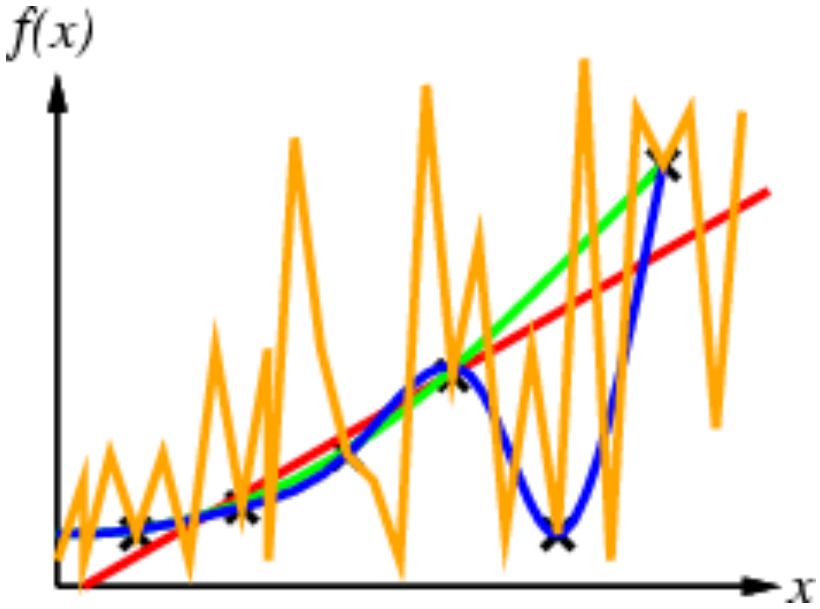
Prediction

- Find function h that fits f at instances x:



Prediction

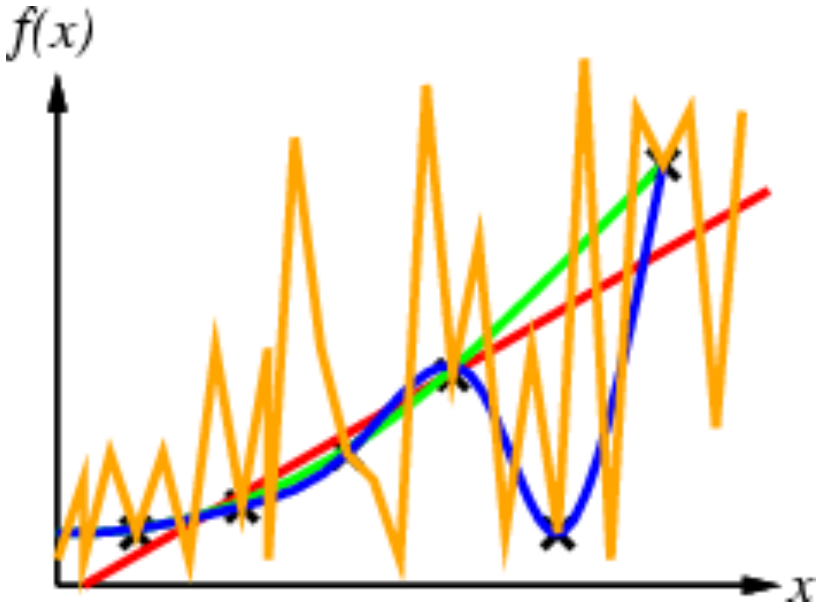
- Find function h that fits f at instances x:



Generalization

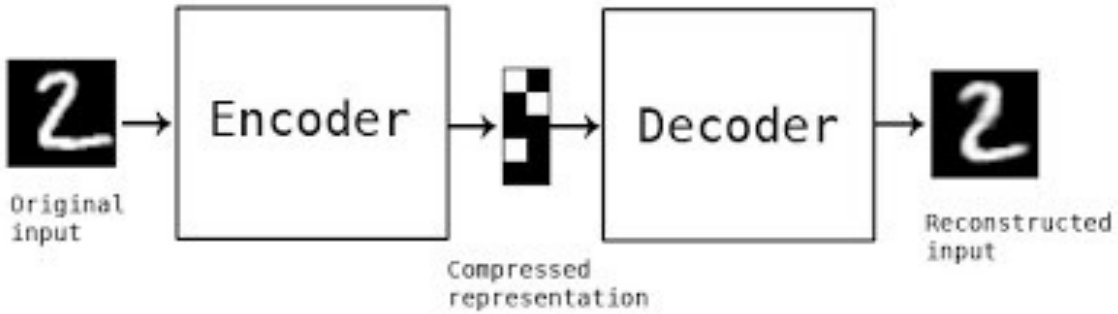
- Key: a good hypothesis will generalize well (i.e., predict unseen examples correctly)

- Ockham's razor: prefer the simplest hypothesis consistent with data



Unsupervised learning

- Given unlabeled data, uncover the underlying structure or distribution of the data
- In other words, find model that explains the data
- Examples:
  - Clustering
  - Dimensionality reduction
  - Generative models



Unsupervised image generation

- Which images are real? And which ones are fake?

Real

CelebA (Liu et al., 2015)

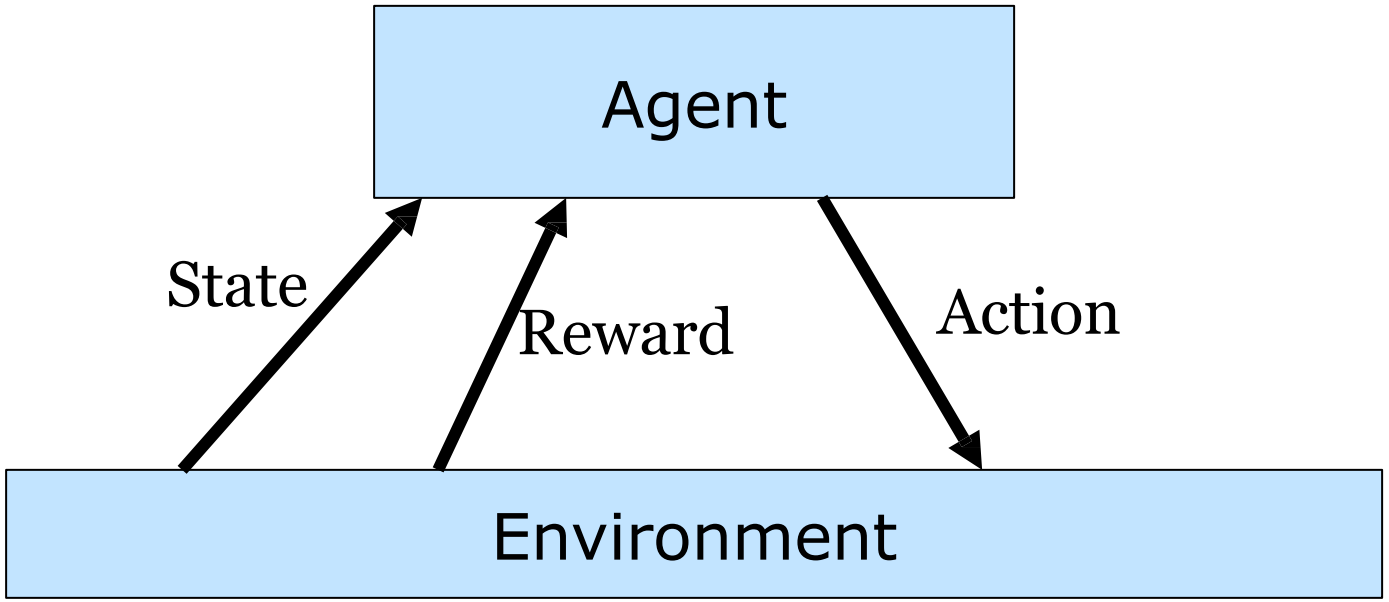
StyleGAN2 (Karras et al., 2020)

Fake!

- Image generation: variational autoencoders, generative adversarial networks, diffusion models

Reinforcement learning

- Training an agent to make decisions within an environment to maximize a cumulative reward
  - Game playing (e.g., AlphaGo)
  - Robot control

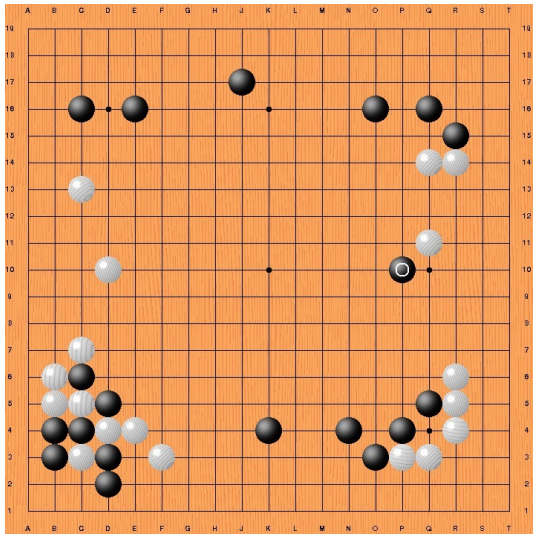


Goal: Learn to choose actions that maximize rewards

Game Playing

❑Example: Go (one of the oldest and hardest board games)

- **Agent:** player
- **Environment:** opponent
- **State:** board configuration
- **Action:** next stone location
- **Reward:** +1 win / -1 loose



2016: AlphaGo defeats top player Lee Sedol (4-1) Game 2 move 37: AlphaGo plays unexpected move (odds 1/10,000)

Other Types of machine learning

- Semi-supervised learning: Learning from a combination of labeled data and unlabeled data
- Transfer learning: applying knowledge learned from one task to another related task

This course

- **Supervised** and **unsupervised** machine learning
- But not **reinforcement** learning
- This course will discuss both algorithms and applications

Applications of Machine Learning

- Speech recognition: Siri, Cortana
- Natural Language Processing: Machine translation, dialog systems
- Computer vision: Image and video analysis
- Robotic Control: Autonomous vehicles
- Computational finance: Stock trading, portfolio optimization

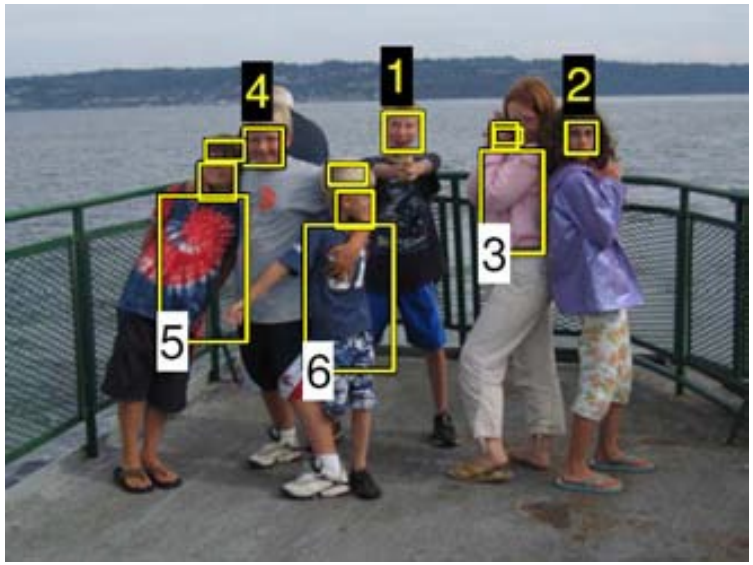
Computer Vision

- Object detection and segmentation

[Detectron2]

Computer Vision

- Face Detection

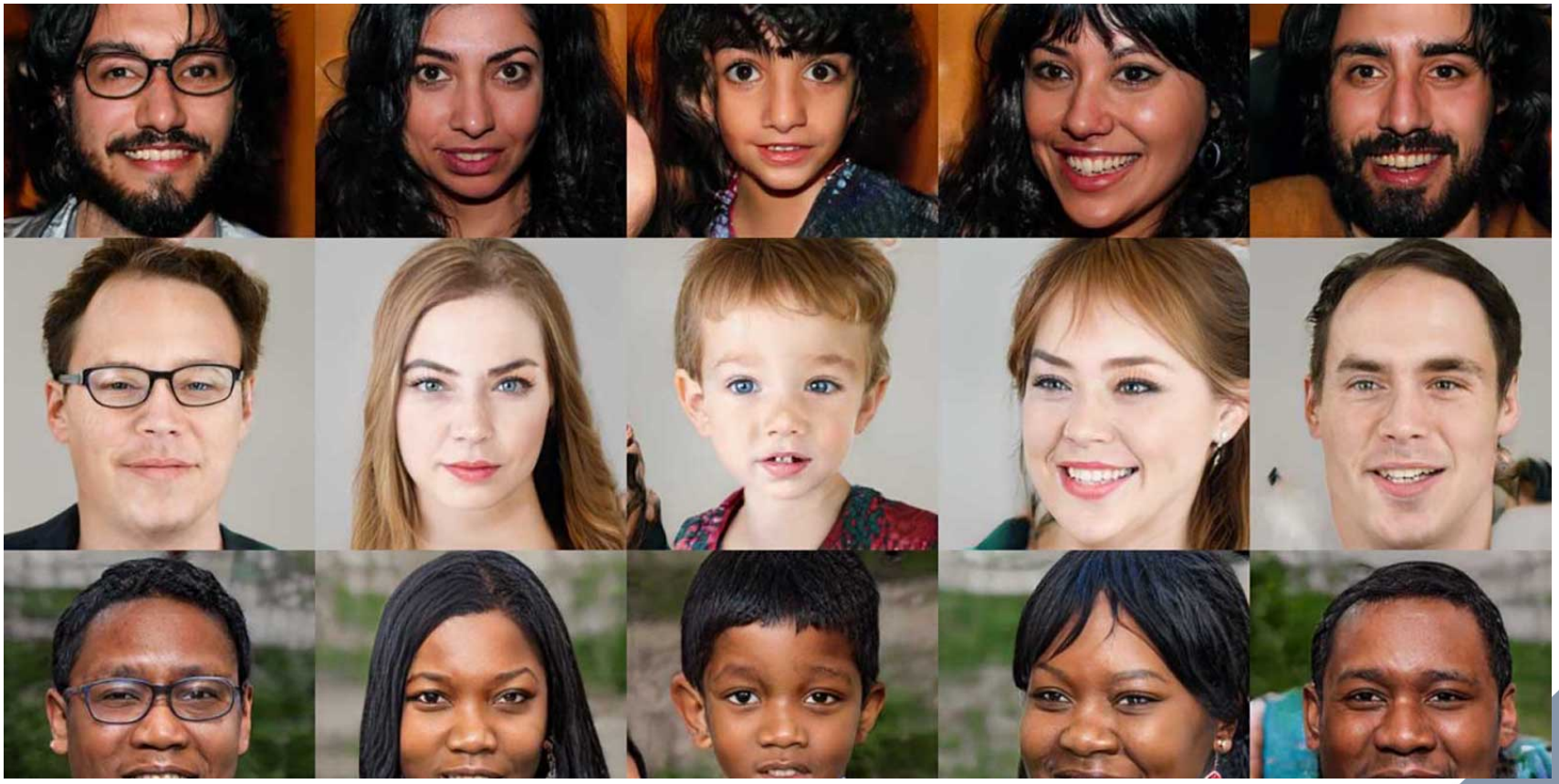


face detection around 00's

and now

Image Synthesis

- Generative Adversarial Network (GAN)



UCMERCED

Image Synthesis from Text

- ❑DALLE-2
- ❑IMAGEN

"A photo of an astronaut riding a horse."

"A modern, sleek Cadillac drives along the Gardiner expressway with downtown Toronto in the background, with a lens flare, 50mm photography."

"A man walking through the bustling streets of Kowloon at night, lit by many bright neon shop signs, 50mm lens."

UCMERCED

ChatGPT- https://chat.openai.com/

How to obtain A+ grade in undergraduate level machine learning course?

Memory updated

Achieving an A+ in a machine learning course requires a mix of solid understanding, consistent effort, and strategic studying. Here's a plan that might help you get there:

1. Understand the Basics Deeply

- **Math Foundation:** Ensure you're comfortable with linear algebra, calculus, probability, and statistics, as these are crucial in understanding machine learning algorithms.
- **Conceptual Clarity:** Focus on truly understanding core concepts like supervised vs. unsupervised learning, overfitting vs. underfitting, and bias-variance tradeoff.

2. Attend Lectures and Participate Actively

- **Engagement:** Attend all lectures, and participate in discussions. Ask questions if something isn't clear.
- **Lecture Notes:** Take detailed notes during lectures and review them regularly.

3. Practice Coding Regularly

- **Programming Skills:** Practice coding in Python, and get comfortable with libraries like NumPy, pandas, Scikit-learn, TensorFlow, or PyTorch.
- **Projects:** Work on small projects or exercises that reinforce the algorithms and techniques discussed in class.

ChatGPT- https://chat.openai.com/

4. Complete Assignments Meticulously

- **Timeliness:** Start your assignments early to avoid last-minute rushes.
- **Thoroughness:** Ensure your assignments are thorough and well-documented. Understand every step you take.
- **Extra Credit:** If there's an opportunity for extra credit, take it!

5. Work on a Research Project

- **Deep Dive:** If the course offers a research project or a capstone, choose a topic that genuinely interests you, and dive deep into it.
- **Novelty:** Try to contribute something new or a unique perspective in your project.

6. Form Study Groups

- **Collaboration:** Join or form a study group. Teaching others is a great way to reinforce your own understanding.
- **Peer Review:** Get your peers to review your work and offer to review theirs.

7. Stay Consistent with Review and Practice

- **Regular Review:** Consistently review topics covered in class and revisit any areas where you're weak.
- **Practice Problems:** Regularly solve problems from textbooks or online platforms like Kaggle, Coursera, or LeetCode.

ChatGPT- https://chat.openai.com/

8. Prepare for Exams Strategically

- **Study Guide:** Create a study guide that covers all the key concepts, formulas, and algorithms.
- **Practice Exams:** Take practice exams under timed conditions to simulate the real exam environment.

9. Seek Help When Needed

- **Office Hours:** Utilize your professor's office hours to clarify doubts.
- **Online Resources:** Leverage online resources like Stack Overflow, YouTube tutorials, and MOOCs if you're stuck.

10. Focus on Time Management

- **Schedule:** Create a study schedule that balances course workload with review and practice.
- **Prioritization:** Prioritize tasks based on their importance and deadlines.

By staying disciplined and curious, you'll be well on your way to earning that A+!

Topics

- Brief review of linear algebra, probability and statistics
- Machine Learning Theory and Algorithms
  - Statistical Learning and Bayesian decision theory
  - Density estimation
  - Clustering and Mixture of Gaussians
  - Dimensionality reduction
  - Logistic Regression and generalized linear models
  - Perceptron and multilayer neural networks
  - Convolutional Neural Networks
  - Decision trees and random forests
  - Ensemble learning: bagging and boosting
  - Kernel machines (support vector machines, SVMs)
- Machine Learning Applications
  - Computer Vision
  - Natural Language Processing (Large Language Model)

Text books

- Hal Daume III, Course in Machine Learning (2017), http://ciml.info
- Kevin Murphy, Probabilistic Machine Learning: An Introduction (2022), https://probml.github.io/pml-book/book1.html
- Understanding Deep Learning by Simon J.D. Prince Published by MIT Press 2023. https://udlbook.github.io/udlbook

Prerequisites

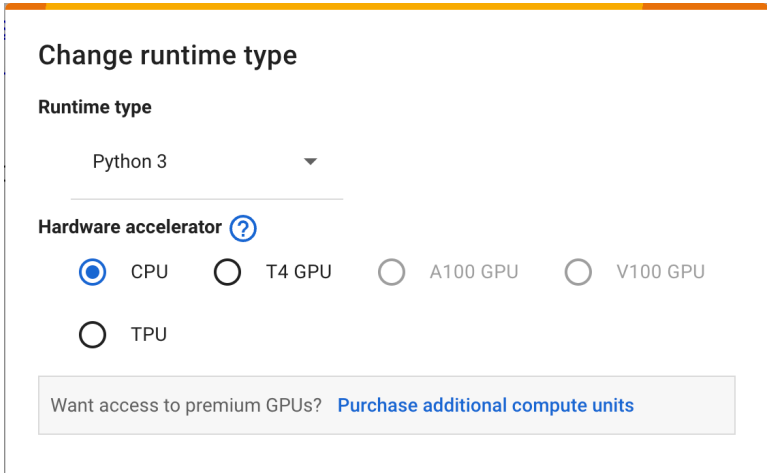
- ❑Linear Algebra
- ❑Linear Analysis
- ❑Probability and Statistics
- ❑Data structure and algorithms
- ❑Python programming
  - ❑Will use Pytorch or Tensorflow framework for deep learning related topics

Grading

- Exams (40%)
  - Midterm:20%
  - Final exam: 20%
- Assignment (40%)
  - 4 assignments
  - 5%, 10%, 10%, 15%
- Labs (20%)

GPU

- Use Python notebook via Google Colab for some labs and assignment



Course Materials

- Course webpage
  - https://ucmercedcse176.github.io/
  - Syllabus
  - Lecture notes
  - Assignments and labs
- Reference for background study:
  - Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron Courville https://www.deeplearningbook.org/
  - UvA Deep Learning Tutorials https://uvadlc-notebooks.readthedocs.io/en/latest/

Looking for Undergraduate Student Researcher

- Sponsored by Google-CAHSI research program
  - The goal of the CAHSI-Google IRP program is to build competitive research capacity of faculty and Hispanic students at CAHSI institutions.
- https://cahsi.utep.edu/cahsi-google-irp/
- Send CV and transcript to Meng (mtang4@ucmerced.edu)
  - Background in computer vision is required
- Prof. Hyeran Jeon is also hiring for a project on machine learning and computer security.

UCMERCED