CS5489
Lecture 1.1: Machine Learning - Overview

Kede Ma

City University of Hong Kong (Dongguan)

香港城市大學（東莞）
City University of Hong Kong
(Dongguan)

Slide template by courtesy of Benjamin M. Marlin

由**提供

Basic Information

- **Teaching Team**
  - Instructor: Dr. Kede MA
    - kede.ma@cityu.edu.hk, Office: YEUNG-6424
  - **Lead TA**: Mr. Haoyu Chen
    - haoychen3-c@my.cityu.edu.hk
  - **TA**: Ms. Shuyan Zhai
  - **TA**: Mr. Shengzhuang Chen
  - **TA**: Mr. Yunqiao Yang
  - **TA**: Mr. Mian Zou
  - **TA**: Ms. Wen Wen
  - **Tutorial Lead**: Mr. Jinglong Yang
  - **Tutorial Lead**: Dr. Zhan Zhuang
- **Canvas-Based Course Site**
  - It is your own responsibility to check Canvas and University e-mail account regularly for announcements and updates

Mixed-Mode Teaching Activities (In-Person and Zoom)

- **Lecture** (**two** hours per week)
  - Present machine learning algorithms with emphasis on how those algorithms are **mathematically** derived and **practically** applied
- **Tutorial** (eight in total, schedules TBD) To be determined
  - Use machine learning algorithms on small examples to gain better intuition and understanding
- **Homework Assignment** (four in total)
  - Solve mathematical problems in machine learning
- **Course Project** (one)
  - Apply machine learning to solve a real-world problem
  - Up to **three** students per group

Assessment

- **Coursework** (70%)
  - Tutorial exercises (10%) - due one week (after release)
  - Assignments (30%) - due two weeks (after release)
  - Course project (30%) - due Week 13 (tentative) 尝试性的
    - Project report, code implementation, and (optional) presentation
- **Final Exam** (30%)
- **Note**:
  - Must get at least 30% on the final exam and at least 30% on the course project to pass the course

Reference Books

The course will recommend several books freely available from the authors:

- [IMLP]: *Introduction to Machine Learning with Python*. Andreas C. Müller and Sarah Guido
- [PRML]: *Pattern Recognition and Machine Learning*. Christopher M. Bishop
- [CO]: *Convex Optimization*. Stephen Boyd and Lieven Vandenberghe
- [MC]: *The Matrix Cookbook*. Kaare B. Petersen and Michael S. Pedersen
- [LA]: *Linear Algebra and Its Applications*. Gilbert Strang

Course Abstract

- The goal of this course is to introduce students to the field of machine learning
  - Machine learning algorithms allow computers to automatically learn to recognize complex patterns from empirical data, such as text and web documents, sounds, images, and videos 经验性或实验性数据
  - This course is intended to give a broad overview of machine learning from both **theoretical** and **practical** standpoints, with emphasis on applying machine learning algorithms to real-world problems
  - At the end of the course, students will have both mathematical understanding of and practical experience with machine learning algorithms

Therefore, in some occasions, you may feel like you are sitting in some **math** class. In case, you find difficulty in picking up the course, you should drop it during the add drop period. No late drop would be allowed

CILOs

- Identify and explain common machine learning algorithms
- Apply machine learning algorithms to solve real-world problems
- Evaluate the effectiveness of different machine learning algorithms and discuss their advantages and disadvantages
- Understand mathematical aspects of machine learning algorithms

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds

- Linear Algebra (e.g., matrix multiplication, linear independence)
- Calculus (e.g., continuity, differentiability) 可微性
- Probability and Statistics (e.g., random variable, multivariate Gaussian distribution)
- Optimization (e.g., convexity, duality)
  凸性 二相性？？

This course requires the use of Python for programming. Students are expected to learn Python as we go

Self-Evaluation: Am I Ready to Go?

- What is the derivative of $\mathbf{x}^T\mathbf{w}$, w.r.t. $\mathbf{w}$, where $\mathbf{x}^T \in \mathbb{R}^{1\times N}$ is a row vector and $\mathbf{w} \in \mathbb{R}^{N\times 1}$ is a column vector?
- What is maximum likelihood estimation (MLE)? What is maximum a posteriori (MAP) estimation?
  归纳的
- What is gradient descent? What is stochastic gradient descent?
  下降 随机的
- What is principal component analysis (PCA)? What is singular value decomposition (SVD)?
  主成分分析，这俩都是梯度下降算法

3-4: Good to go; **0-2**: Don't worry - we'll get you covered

Academic Honesty

- CityU (Dongguan) has its own **Rules of Academic Honesty**
- Plagiarism...
  - It is serious fraud to plagiarize others' work
  - Punishment ranges from warning to course failure
- How to prevent plagiarism...
  - Finish the assignments by yourself! You have to write the program/solution yourself
    - Okay to talk about how to do the problem with your classmates
    - Protect your solution; don't give it away as a "reference" copy
  - In plagiarism cases, we treat both giver and copier as guilty
  - You hurt your own grades by not reporting cheating
- As instructor...
  - We have responsibility to report academic dishonesty cases so as not to compromise the quality of education
  - We take suspected plagiarism cases very seriously

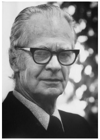
Academic Honesty in the Presence of Generative AI (GenAI)

- We basically follow the department's policy on using GenAI (e.g., ChatGPT, GPT-4, and LLaMA 3.1) in coursework
  - Students are not allowed to use GenAI for programming tasks
    - Programming is something that you shall learn by yourself
  - Students are allowed to use GenAI for helping writing assignments and reports
    - Acknowledgement must be made through proper citation
    - **Lack of acknowledgment (e.g., by simply copying text or ideas) is considered plagiarism**
    - Use with caution as GenAI is good at confabulating (i.e., hallucinating) false yet coherent information
      闲谈
      产生幻觉
- We will introduce the underlying mechanisms of some popular generative models towards the end of this course

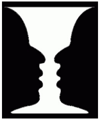
Introduction

What is Learning?

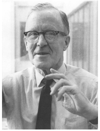
Definitions of Learning



Behaviorism (Skinner, 1900-1950): Learning is a long-term change in behavior due to experience



Cognitivism (Gestalt School, 1920-): Learning is an internal mental process that integrates new information into established mental frameworks and updates those frameworks over time



Connectionism (Hebb, 1949): Learning is a physical process in which neurons join by developing the synapses between them

突触

Introduction

What is Machine Learning?

Views on Machine Learning

Samuel (1959): "Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed"

Mitchell (1997): "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"

Jordan (2015): "It is one of today's rapidly growing technical fields, lying at the intersection 十字路口 of computer science and statistics, and at the core of artificial intelligence and data science"

Samuel's View

- Example: Recognizing handwritten digits in an image
  - 28×28 image $\longrightarrow$ 784-dim vector
  - A lot of variations & permutations 排列序列
  - Difficult to identify rules & code by hand

- ML solution:
  - Gather some example data
  - Train computer to discover differences automatically

Mitchell's View

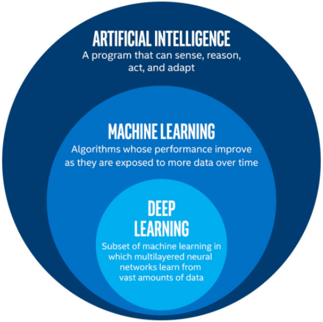
- "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."
  - "class of tasks T": Learning is task-specific (recognition, clustering, etc.)
  - "performance measure P": Optimize a loss function (e.g., error rate), but also prevent overfitting (regularization) 正则化
  - "experience E": Data-driven! More data is better!

Machine Learning, Deep Learning, and Artificial Intelligence (AI)

- Machine learning grew out of early work in AI
  - and other fields: statistics, physics, neuroscience, ...
  - Fueled by more powerful computers and more data

Different Fields in AI

- General artificial intelligence
- Machine learning
- Natural language processing
- Computer vision
- ...



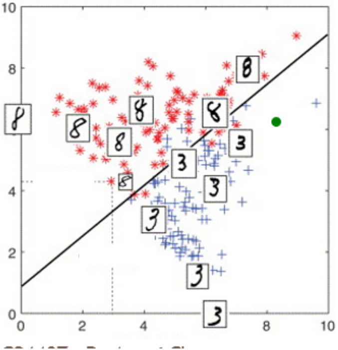
Topics in Machine Learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Learning theory

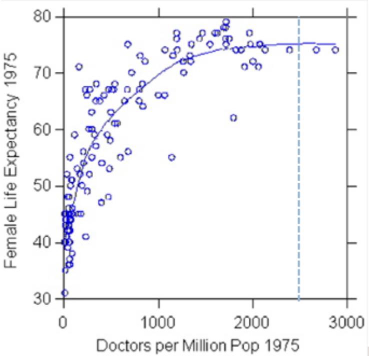
Supervised Learning

- Training data has inputs and outputs
  - E.g., digit recognition (input=image, output=digit)
  - Learn a function mapping inputs to outputs

Classification

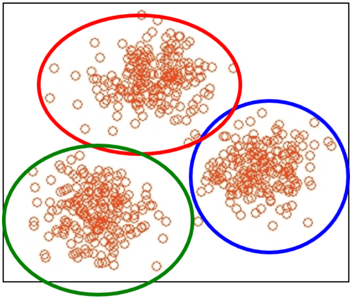


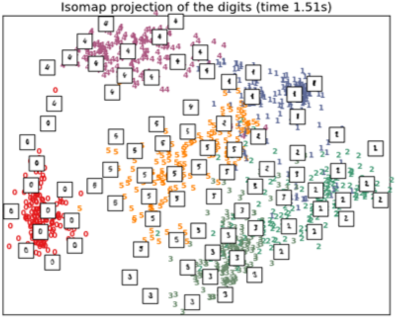
Regression



Unsupervised Learning

- Training data only has inputs (no outputs)
  - Density estimation - construct a probability model over the input
  - Clustering - discover groups of similar examples
  - Dimensionality reduction - project high-dim data to 2 or 3-dimensions (for visualization purposes) 降维度

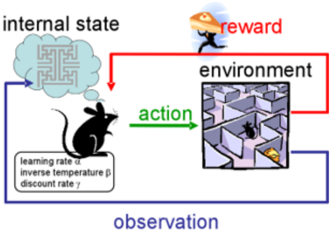




Reinforcement Learning

有点像生物的反馈调节啊

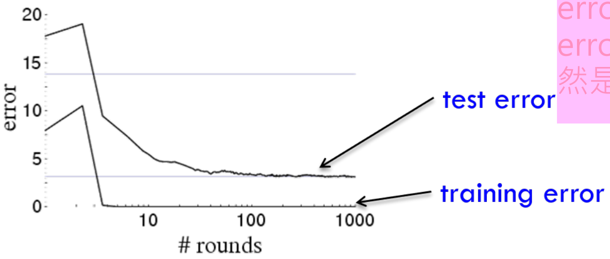
- Make a sequence of actions, given current states
  - E.g. a robot interacting with its environment
  - Maximize the reward
    - At some point, receive a reward or a punishment
    - Actions may also affect future reward



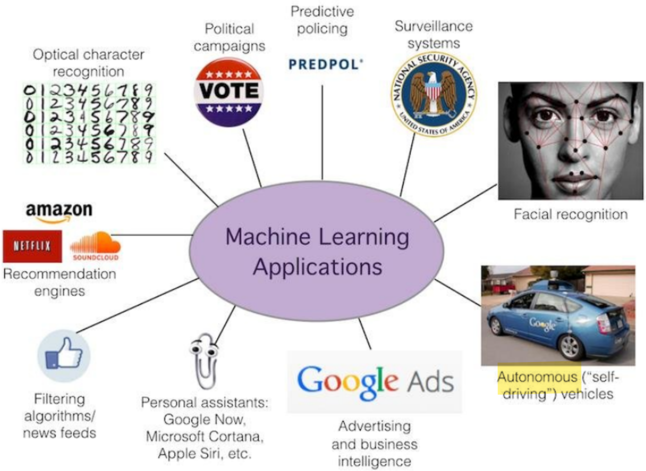
Learning Theory

- Why does machine learning work?
  - Performance guarantees - bounds on the expected test error
  - What types of functions can be represent by an algorithm, and how much data do we need?

这个training error 和test error的gap依然是难解之题



Machine Learning In the Real World



AI Generated Content (AIGC): ChatGPT

M Tell me how ChatGPT works.

ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.

AIGC: Text to Image Models

vibrant portrait painting of Salvador Dalí with a robotic half face

a shiba inu wearing a beret and black turtleneck

贝雷帽 高领绒衣

a close up of a handpalm with leaves growing from it

哦这些应是prompt

an espresso machine that makes coffee from human souls, artstation

浓咖啡

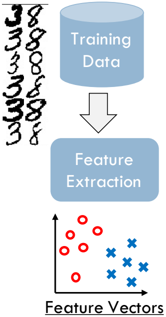
panda mad scientist mixing sparkling chemicals, artstation

a corgi's head depicted as an explosion of a nebula

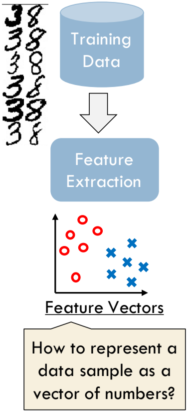
小狗 描绘 星云

Machine Learning Training Pipeline

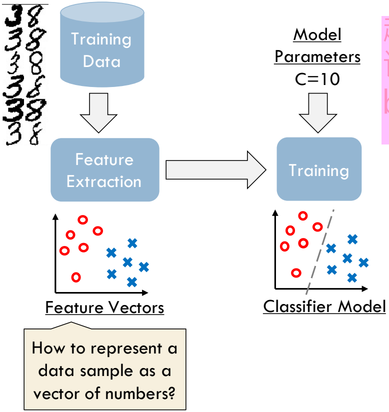
Machine Learning Training Pipeline



Machine Learning Training Pipeline

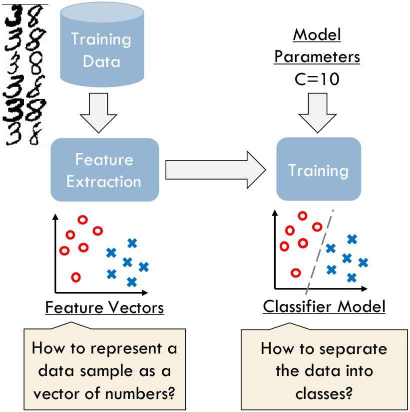


Machine Learning Training Pipeline

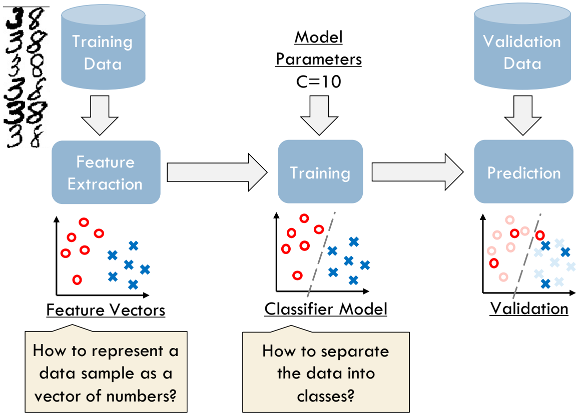


超参hyper parameter，就是自己设定的那种，比如batchsize，learning rate之类

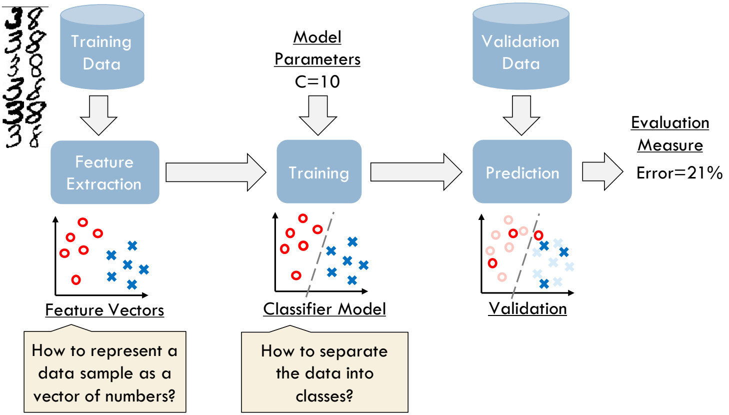
Machine Learning Training Pipeline



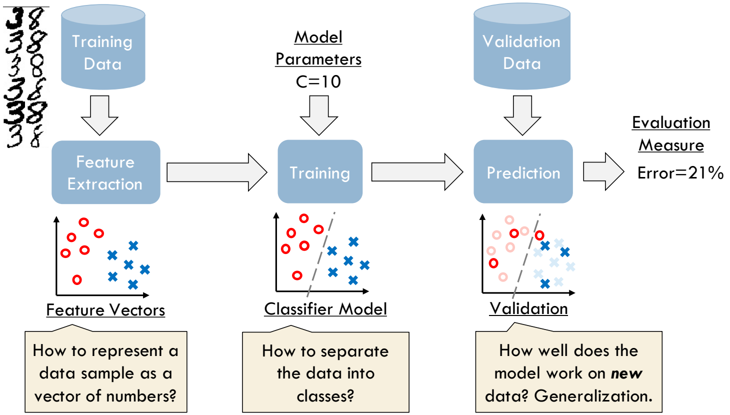
Machine Learning Training Pipeline



Machine Learning Training Pipeline

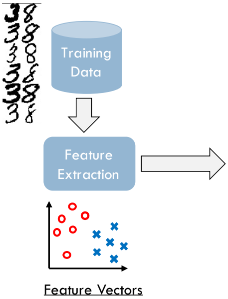


Machine Learning Training Pipeline

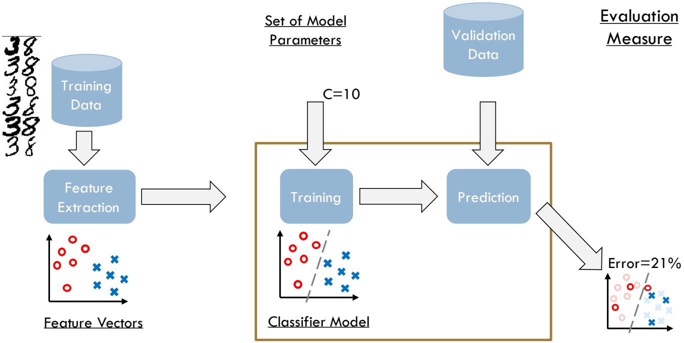


Model Selection Pipeline

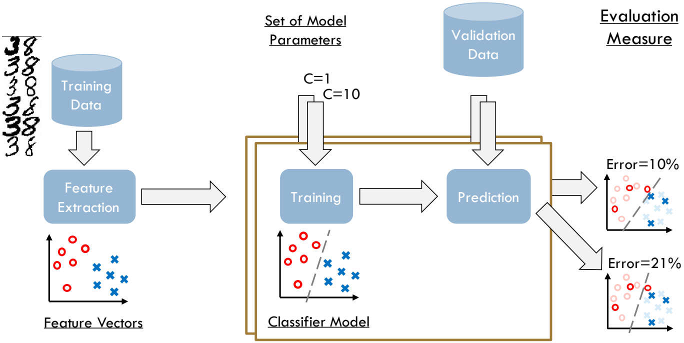
Evaluation Measure



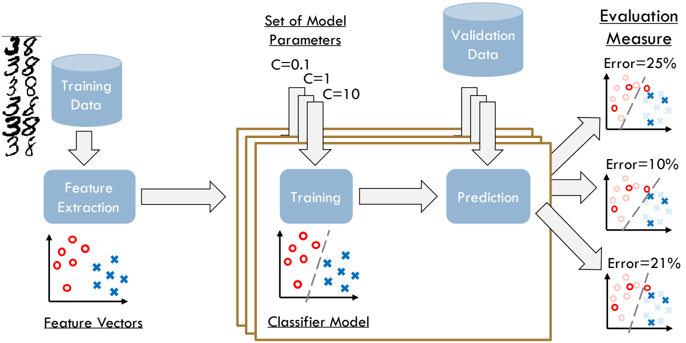
Model Selection Pipeline



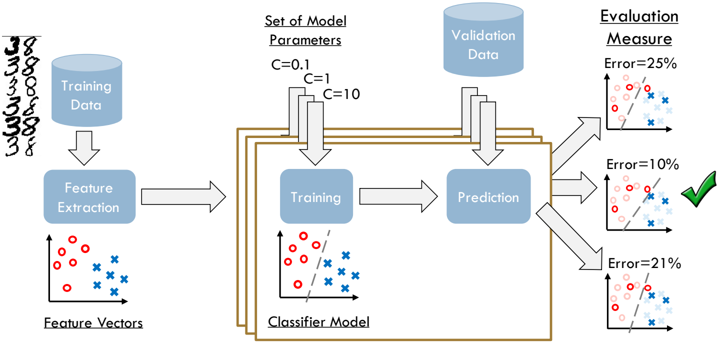
Model Selection Pipeline



Model Selection Pipeline



Model Selection Pipeline



Linear Algebra

Definition: Vector Space

The real vector space $\mathbb{R}^N$ is a set with elements $\mathbf{x} = [x_1, \dots, x_N]^T$ where each $x_j \in \mathbb{R}$. The elements $\mathbf{x}$ are called vectors, and they satisfy the following properties: RN是向量空间是列向量空间，不是矩阵啊

- **Addition:** If $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$, then $\mathbf{x} + \mathbf{y} = [x_1 + y_1, \dots, x_N + y_N]^T \in \mathbb{R}^N$
- **Scalar Product:** If $\mathbf{x} \in \mathbb{R}^N$ and $a \in \mathbb{R}$, then $a\mathbf{x} = [ax_1, \dots, ax_N]^T \in \mathbb{R}^N$
- **Inner Product:** If $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$, then $\mathbf{x}^T\mathbf{y} = \sum_{j=1}^N x_j \cdot y_j$

outer product: x(1*N)yT(N*1)

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$ is rectangular array of elements $x_{ij} \in \mathbb{R}$, $1 \leq i \leq M$, $1 \leq j \leq N$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{pmatrix}$$

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{M \times N}$, $\mathbf{Y} \in \mathbb{R}^{M \times N}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{M \times N}$ and $z_{ij} = x_{ij} + y_{ij}$
- **Scalar Product:** If $\mathbf{X} \in \mathbb{R}^{M \times N}$, $a \in \mathbb{R}$ and $\mathbf{Z} = a\mathbf{X}$, then $\mathbf{Z} \in \mathbb{R}^{M \times N}$ and $z_{ij} = ax_{ij}$
- **Matrix Multiplication:** If $\mathbf{X} \in \mathbb{R}^{M \times N}$, $\mathbf{Y} \in \mathbb{R}^{N \times C}$ and $\mathbf{Z} = \mathbf{XY}$, then $\mathbf{Z} \in \mathbb{R}^{M \times C}$ and $z_{ij} = \sum_{k=1}^{N} x_{ik}y_{kj}$

You should be familiar with basic matrix types (square, diagonal, identity), basic matrix operations (transpose, inverse, trace, etc.), and matrix concepts (eigenvalues, orthogonality, etc.)

爱跟外溜，特征值 正交

Probability

Definition: Probability Distribution

A probability distribution p over a sample space Ω is a **function** from elements/subsets of Ω to real numbers that satisfies the following conditions:

- Non-negativity: $p(\omega) \geq 0$ for all $\omega \subseteq \Omega$
- Normalization: $p(\Omega) = 1$ 归一化
- Additivity: For all $\omega, \omega' \subseteq \Omega$ that are **disjoint** sets, $p(\omega \bigcup \omega') = p(\omega) + p(\omega')$

Random Variable

Definition: Random Variable

A random variable X is defined by a **function** f_X that maps each element ω of the sample space Ω to a value $x = f_X(\omega)$ in a set $\mathcal{X}$ (called the *range* of the random variable)

For each $x \in \mathcal{X}$ the **event** $\{X = x\}$ refers to the subset of the sample space $\{\omega | \omega \in \Omega, f_X(\omega) = x\}$

For each $x \in \mathcal{X}$ the probability $p(X = x) = p(\{\omega | \omega \in \Omega, f_X(\omega) = x\})$

Probability and Random Variable

We can also specify a probability distribution for a random variable X with range $\mathcal{X}$ directly instead of via an underlying sample space Ω. The following conditions must hold:

- **Discrete probability mass function:**
  $p(X = x) \geq 0 \ \ \forall x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(X = x) = 1$
- **Continuous probability density function:**
  $p(X = x) \geq 0 \ \ \forall x \in \mathcal{X}$ and $\int_{\mathcal{X}} p(X = x)dx = 1$

P(X=x)可以大于1，在概率密度函数中，乐，头回想到这个点

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization
- Conditioning
- Bayes' rule
- Expectation
- Classical distributions (Bernoulli, Multinomial, Gaussian)