Introduction to Machine Learning

Week 1

Lecture 1

Prof. Subrahmanya Swamy Peruru
Department of Electrical Engineering
IIT Kanpur

Machine Learning: Learning from Data

Data

Cat

Cat

Dog

Dog

Training

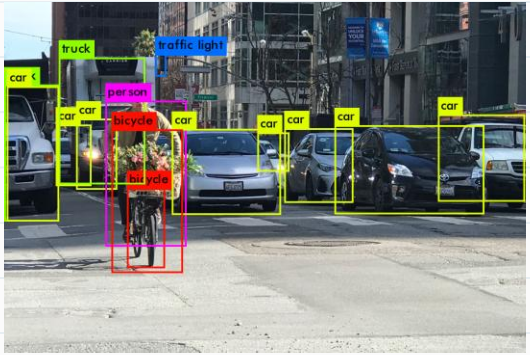
Trained model

Dog / Cat ?

Example: Image Classification

Machine Learning Everywhere

Self-Driving Cars



ChatGPT

Probability & Statistics

Basics

Lecture Contents

01 Elements of Probability

- Sample space
- Events
- Axioms of Probabaility

02 Properties of Probability

- Monotonicity
- Union bound
- Inclusion-Exclusion

03 Conditional Probability

- Conditional Probability Definition
- Independence of Events
- Bayes Rule

01 Elements of Probability

Sample Space

Ω - The set of all the outcomes of a random experiment

$\omega \in \Omega$ denotes an element of the sample space

Examples

Coin Toss

$\Omega = \{\mathrm{H}, \mathrm{T}\}$

6-faced Die

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Finite Sample Space

Sample Space

Ω - The set of all the outcomes of a random experiment

$\omega \in \Omega$ denotes an element of the sample space

Examples T T H $\Rightarrow \omega = 3$

Toss until first head

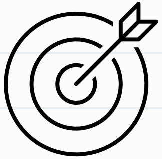
Experiment:
Toss a coin repeatedly until a Head appears

Outcome:
The number of times the coin is tossed.

What is the sample space?

$\Omega = \{1, 2, 3, \ldots\}$ **Countable**

Dart Throw Experiment



Experiment:
Throw a dart at a board and observe where it lands.

Outcome:
Distance of the dart from the center of the board.

What is the sample space?

$\Omega = [0, R]$ **Uncountable**

Sample Space is an Infinite Set

A set is called *countable* if its elements can be arranged in a sequence $x_1, x_2, x_3, \ldots$.

Event

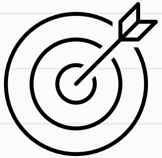
- Often, we are not interested in the exact outcome of our random experiment
- We only care whether an event of interest occurs

6-faced Die

$\Omega = \{1, 2, 3, 4, 5, 6\}$ Did the die land on an even face?

$E = \{2, 4, 6\}$ -- Even face?
$H = \{4, 5, 6\}$ -- At least 4 or more?

Dart Throw Experiment



$\Omega = [0, R]$ Has the dart landed inside the innermost circle? $A = [0, 1]$

$A \subseteq \Omega$ - An event is a subset of the sample space[1]

1. For a rigorous definition of event, see standard measure-theoretic probability texts.

Collection of Events

- $\mathcal{F}$ – The collection of all the events of interest[2]

- Our goal is to quantify how likely an event is to occur (i.e., define probability for all events in $\mathcal{F}$)

Examples:

6-faced Die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\boldsymbol{\mathcal{F}} = \{\emptyset, \boldsymbol{\Omega}, \{\mathbf{2}, \mathbf{4}, \mathbf{6}\}, \{\mathbf{1}, \mathbf{3}, \mathbf{5}\}\}$$

This collection suffices if we are only interested in whether an even face is observed.

Why do we include 4 events if we only care about the occurrence of an even face? (Read about σ –algebra[2])

1. Sample Space should be included: $\Omega \in \mathcal{F}$
2. Complements of events: $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. Countable unions: $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$

2. Formally, the collection of events must be a σ-algebra (see a measure-theoretic probability text)

Collection of Events

- $\mathcal{F}$ – The collection of all the events of interest[2]
- Our goal is to quantify how likely an event is to occur (i.e., define probability for all events in $\mathcal{F}$)

Examples:

6-faced Die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\mathcal{F} = \{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\}$$

This collection suffices if we are only interested in whether an even face is observed.

When the **sample space is finite**, typically we **include all possible subsets** in our collections

$\mathcal{F} = 2^{\Omega}$ → Power set (Collection of all possible subsets of Ω)

2. Formally, the collection of events must be a σ-algebra (see a measure-theoretic probability text)

Probability Axioms

Given a *Sample Space* $\boldsymbol{\Omega}$, and

a *collection of events* $\mathcal{F}$,

the probability measure $\boldsymbol{P} : \mathcal{F} \rightarrow [0,1]$ assigns a number between zero and one to each event

such that:

$A_i \cap A_j = \emptyset$ for any $i \neq j$

- **Normalization:** $\boldsymbol{P(\Omega) = 1}$
- **Countable Additivity:** For any set of disjoint events $\boldsymbol{A_1, A_2, A_3, \ldots,}$

$$\boldsymbol{P(\cup_i A_i) = \sum_i P(A_i)}$$

Probability Measure: Examples

Coin Toss

Sample Space $\Omega = \{\mathrm{H}, \mathrm{T}\}$

Collection of Events $\mathcal{F} = \{\emptyset, \Omega, \{H\}, \{T\}\}$

Probability Measure

- $P(\emptyset) = 0$
- $P(\Omega) = 1$
- $P(\{H\}) = 0.5$
- $P(\{T\}) = 0.5$

Is P a valid probability measure?

Proof:

- $A_1 = \{\mathrm{H}\},\ A_2 = \{T\}$ are disjoint
- $A_1 \cup A_2 = \Omega \Rightarrow P(A_1 \cup A_2) = P(\Omega) = 1$
- $P(A_1) + P(A_2) = 0.5 + 0.5 = 1$

Probability Measure: Discrete

- When the sample space is discrete (countable),
- specifying probabilities of individual outcomes
- uniquely determines the probability of every event via countable additivity.

Example

- Discrete sample: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Specify probabilities for individual outcomes: $P(\{i\}) = \frac{1}{6}$ **for** $i \in \Omega$
- Event probabilities through Countable additivity:

$$A = \{2, 5\}$$

$$P(A) = P(\{2\} \cup \{5\}) = P(\{2\}) + P(\{5\}) = \frac{2}{6}$$

Probability Measure: Continous

Continuous case (Ω is uncountable)

- $P(\{\omega\}) = 0$ for all $\omega \in \Omega$
- Typically Probabilities assigned to intervals/regions

Example: Uniform probability measure on $[\mathbf{0}, \mathbf{1}]$

- $P(\{0.34786\}) = 0$ (a particular element in sample space)
- $P([0.2, 0.5]) = 0.3$ (Interval of length 0.3)

Properties (Derived from Axioms)

- **Empty Set:** $P(\emptyset) = 0$
- **Complement:** $P(A^c) = 1 - P(A)$
- **Monotonicity:** $A \subseteq B \Rightarrow P(A) \leq P(B)$
- **Inclusion-Exclusion:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$$

Verify for Disjoint events: $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Union Bound: $P(A \cup B) \leq P(A) + P(B)$

Conditional Probability

- Suppose that in a certain city, 20% percent of the days are rainy

$$\boldsymbol{P}(Rain) = \mathbf{0.2}$$

- The probability that it rains given that it is cloudy might be:

$$\boldsymbol{P}(Rain\ \boldsymbol{given}\ Cloudy) = \mathbf{0.6}$$

Conditional Probability: If $P(B) > 0$, The probability of A **given** B is defined as

$$\boldsymbol{P}(\boldsymbol{A}\ \boldsymbol{given}\ \boldsymbol{B}) = \boldsymbol{P}(\boldsymbol{A} \mid \boldsymbol{B}) = \frac{\boldsymbol{P}(\boldsymbol{A} \cap \boldsymbol{B})}{\boldsymbol{P}(\boldsymbol{B})}$$

Key Identity: $P(A \cap B) = P(A)P(A \mid B)$

Chain Rule: $P(A \cap B \cap C) = P(A)\, P(B \mid A)\, P(C \mid A \cap B)$

Example:

- Given that the die shows a number ≤ 3, what is the probability that it is even?

- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Given:** $B = \{\leq 3\} = \{1, 2, 3\}$
- **To Find:** $A = \{even\} = \{2, 4, 6\}$

Compute $P(A \mid B) = \dfrac{\boldsymbol{P(A \cap B)}}{\boldsymbol{P(B)}}$

$$\boxed{\begin{gathered} A \cap B = \{2\} \\ P(A \cap B) = 1/6 \\ P(B) = 3/6 \\[1em] P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3} \end{gathered}}$$

Independent Events

- Events A and B are **independent** if

$$\boxed{\boldsymbol{P(A \cap B) = P(A)\,P(B)}}$$

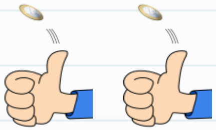
- Equivalently, if $P(B) > 0$:

$$\boldsymbol{P(A \mid B) = P(A)}$$

Intuition: **Knowing that B occurred does not change the probability of A**

Event B provides no information about event A.

Exercise: Consider a random experiment of throwing a fair coin twice and show that the outcome of **first toss is independent of the second toss**.



Proof Outline:

- Model the experiment using an appropriate sample space Ω
- Define the events A_1: Head occurs on first toss, A_2: Head occurs on second toss
- Use the definition of independence and verify $P(A_1 \cap A_2) = P(A_1)\,P(A_2)$

Independence (3 events)

- Events A, B, and C are **(mutually) independent** if

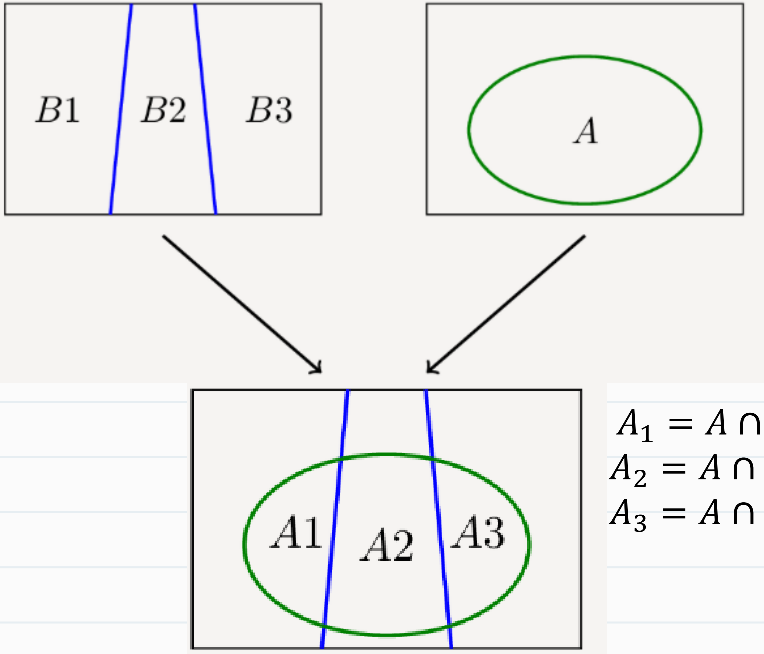
$$\boldsymbol{P(A \cap B) = P(A)P(B)}$$
$$\boldsymbol{P(B \cap C) = P(B)P(C)}$$
$$\boldsymbol{P(A \cap C) = P(A)\,P(C)}$$

and

$$\boldsymbol{P(A \cap B \cap C) = P(A)P(B)P(C)}$$

Pairwise independence alone is not sufficient

Law of Total Probability (LoTP)



$A_1 = A \cap B_1$
$A_2 = A \cap B_2$
$A_3 = A \cap B_3$

$$\boldsymbol{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A \mid B_3)P(B_3)}$$

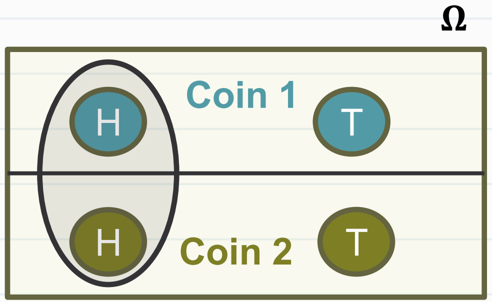
$$P(A) = P(A_1) + P(A_2) + P(A_3) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$
$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A \mid B_3)P(B_3)$$

LoTP: Example

Suppose there are **two coins**:

- Coin 1 lands heads with probability 0.9
- Coin 2 lands heads with probability 0.1

- **One of the two coins is chosen** *uniformly at random*,
- and then the **chosen coin is tossed once**.



Question: What is the **probability of observing a head**?

LoTP: $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$

$$P(Head) = P(Head \mid Coin\ 1)\ P(Coin\ 1) + P(Head \mid Coin\ 2)\ P(Coin\ 2)$$
$$= 0.9 \times \frac{1}{2} + 0.1 \times \frac{1}{2} = 0.5$$

Bayes Rule

- Consider our two-coin experiment discussed before

Question: Given the outcome is a head, what is the probability that coin 1 was chosen?

$$P(Coin\ 1 \mid Head) = \frac{P(Coin\ 1 \cap Head)}{P(Head)}$$

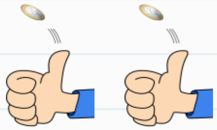
$$= \frac{P(Head \mid Coin\ 1)\ P(Coin\ 1)}{P(Head \mid Coin\ 1)\ P(Coin\ 1) + P(Head \mid Coin\ 2)\ P(Coin\ 2)} \quad \textbf{(LoTP)}$$

Prior: $P(Coin\ 1)$

Likelihood: $P(Head \mid Coin\ 1)$

Posterior: $P(Coin\ 1 \mid Head)$

Conditional Independence



- When you toss a coin twice
- Outcome of **first toss is independent of the second** toss
    - $\boldsymbol{H_1}$: Head in *first* toss
    - $\boldsymbol{H_2}$: Head in *second* toss
    - $P(H_1 \cap H_2) = P(H_1)P(H_2)$

Definition

Let $\mathrm{P(C) > 0}$.

Then A, B are conditionally independent given C if

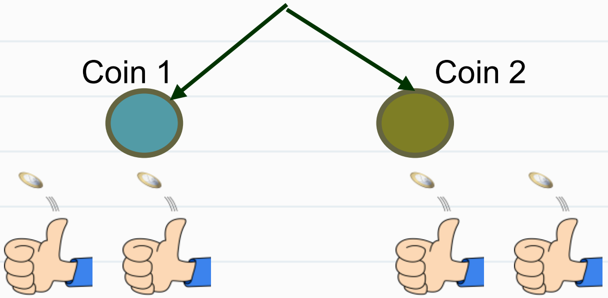
$$P(A \cap B|C) = P(A\,|C)\;P(B\,|C)$$

Another Experiment:

Suppose there are **two coins**:

- **Coin 1** lands heads with probability 0.9
- **Coin 2** lands heads with probability 0.1

Select one of the two coins uniformly at random



Toss the selected coin twice

Exercise:

Are H_1 and H_2 independent in this new experiment?

Are H_1 and H_2 conditionally independent given that Coin 1 was chosen?

References

- *Chapter 1* of the book **H. Pishro-Nik, "Introduction to probability, statistics, and random processes",** available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

Thank You