Machine Learning Theory

In progress

Contents

Chapter 1 Introduction

Definition 1.1 (Probability triple)

A **probability triple** is $(\Omega, \mathcal{F}, \mathbb{P})$ where:

- The **sample space** Ω is any non-empty set.
- The **σ-algebra** $\mathcal{F}$ is a collection of subsets of Ω satisfies:
  1. Containing Ω and $\varnothing$.
  2. (closed under the complements) For any $A \in \mathcal{F}$, $A^C \in \mathcal{F}$.
  3. (closed under the countable unions) For any countable (or finite) collection $\{A_i\} \subseteq \mathcal{F}$, the union $\bigcup_i A_i \in \mathcal{F}$.
- The **probability measure** $\mathbb{P}$ is a mapping from $\mathcal{F}$ to $[0, 1]$ satisfies:
  1. $\mathbb{P}(\Omega) = 1$.
  2. (**Countable additivity**) For a countable disjoint collection $\{A_i\} \subseteq \mathcal{F}$, $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$.

♣

Remark In general $\mathcal{F}$ might not contain all subsets of Ω.

Corollary 1.1

- *$\mathcal{F}$ is closed under the countable intersections.*
- $\mathbb{P}(\varnothing) = 0$.
- $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.
- *(**Monotonicity**) $\mathbb{P}(A) \leq \mathbb{P}(B)$ whenever $A \subseteq B$.*
- *(**Principle of inclusion-exclusion**) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*
- *(**Countable subadditivity**) For a countable collection $\{A_i\} \subseteq \mathcal{F}$, $\mathbb{P}(\bigcup_i A_i) \leq \sum_i \mathbb{P}(A_i)$.*

♡

Theorem 1.1

Let

- *Ω be a finite or countable non-empty set.*
- *$p : \Omega \to [0, 1]$ be ab function satisfying $\sum_{\omega \in \Omega} p(\omega) = 1$,*

then there is a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F} = \mathbb{P}(\Omega)$, and for $A \in \mathcal{F}$, $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$.

♡