ML INFOSEC
2: Probability Theory

January 22, 2019

Notation and Set Theory I

- $\emptyset$: the empty set.
- $\mathbb{N} = \{1, 2, 3, \dots\}$: the set of natural numbers.
- $\mathbb{R}$: the field of real numbers.
- $x \in A$: x is element of A.
- $x \notin A$: x is not element of A.
- $A \subseteq B$: A is subset of B.
- $2^A = \{B \mid B \subseteq A\}$: power set of A.
- $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$: the union of A and B.
- $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$: the intersection of A and B.
- A and B are disjoint if $A \cap B = \emptyset$.

Notation and Set Theory II

- $A \setminus B = \{x \mid x \in A, x \notin B\}$ set difference.
- If $A \subseteq X$, $A^c = X \setminus A$ complement of A in X.
- $X = A \cup A^c$, $A \cap A^c = \emptyset$.
- $(A^c)^c = A$.
- $(A \cap B)^c = A^c \cup B^c$, $(A \cup B)^c = A^c \cap B^c$.
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- $|A|$ (or $\#A$) denotes the (cardinal) number (of elements) of a set A.
- $A \times B = \{(x, y) \mid x \in A, y \in B\}$ the Cartesian product of A and B.

Finite probability spaces I

Let Ω be finite, non-empty set.

Definition

A function $P : 2^{\Omega} \to [0, 1]$ is called a probability measure (distribution) on Ω if

- $P(\emptyset) = 0,\ P(\Omega) = 1$
- If A and B are disjoint subsets of Ω, then $P(A \cup B) = P(A) + P(B)$

Definition

A pair (Ω, P) with a probability measure P on Ω is called a probability space. Subsets of Ω are called events.

Finite probability spaces II

Lemma

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

Proof.

$A \cup B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \emptyset$, thus

$$P(A \cup B) = P(A) + P(B \setminus A).$$

Moreover,
$B = (B \setminus A) \cup (A \cap B)$ and $(B \setminus A) \cap (A \cap B) = \emptyset$, so

$$P(B) = P(B \setminus A) + P(A \cap B).$$

□

Probability vectors

Let $\Omega = \{1, \ldots, n\}$, P a probability measure on Ω and define

$$p_i = P(\{i\}), 1 \leq i \leq n.$$

Then $0 \leq p_i \leq 1$, $\sum_i p_i = 1$ and

$$P(A) = \sum_{i \in A} p_i.$$

Definition

A row-vector $p = (p_1, ..., p_n) \in \mathbb{R}^n$ is called a probability vector if

- $0 \leq p_i \leq 1$ for all $i \in \{1, \ldots, n\}$,
- $\sum_{i=1}^n p_i = 1$.

Examples

1. Fair dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$. For an event $A \subseteq \Omega$ we have

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6};$$

$p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ = uniform distribution on Ω.

2. Card game: Let Ω be set of all cards of a standard 52-card deck and A be the set of all kings. Then

$$P(A) = \frac{4}{52} = \frac{1}{13}.$$

Conditional probability I

Definition

If $P(B) > 0$, then

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

is called the conditional probability of A, given B.

Conditional probability II

Lemma

If $0 < P(B) < 1$, *then*

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$

Proof.

$$\begin{aligned} P(A \mid B)P(B) + P(A \mid B^c)P(B^c) &= P(A \cap B) + P(A \cap B^c) \\ &= P((A \cap B) \cup (A \cap B^c)) \\ &= P(A \cap (B \cup B^c)) \\ &= P(A \cap \Omega) = P(A) \end{aligned}$$

$\square$

Conditional probability III: Examples

1. Fair dice. $A = \{2, 4, 6\}$, $B = \{3, 4, 5\}$. Then $A \cap B = \{4\}$, i.e. $P(A \cap B) = 1/6$, $P(B) = 1/2$, thus
$$P(A \mid B) = 1/3.$$
2. Draw two cards without replacement from a standard 52-card deck. Let $A = 1st\ card\ is\ ace$ and $B = 2nd\ card\ is\ ace$. Then $P(A) = 4/52$ and $P(B \mid A) = 3/51$, thus
$$P(A \cap B) = P(B \mid A)P(A) = \frac{4}{52}\frac{3}{51} = \frac{1}{221}.$$

Bayes theorem

Theorem

If $P(A), P(B) > 0$, *then*

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}.$$

Proof.

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

$\square$

Example: Medical diagnosis problem

Let D be a disease such that $P(D) = 0.008, P(\neg D) = 0.992$ and Test T for D with the following properties

$$\begin{array}{ll} P(T\ pos \mid D) = 0.98 & P(T\ neg \mid D) = 0.02 \\ P(T\ pos \mid \neg D) = 0.03 & P(T\ neg \mid \neg D) = 0.97 \end{array}$$

How large is $P(D \mid T\ is\ pos)$?

$$P(D \mid T\ pos) = \frac{P(T\ pos \mid D)P(D)}{P(T\ pos)} = \frac{0.98 \times 0.008}{P(T\ pos)}$$

$$\begin{aligned} P(T\ pos) &= P(T\ pos \mid D)P(D) + P(T\ pos \mid \neg D)P(\neg D) \\ &= 0.98 \times 0.008 + 0.03 \times 0.992 \\ &= 0.00784 + 0.02976 = 0.0376 \end{aligned}$$

Thus

$$P(D \mid T\ pos) = \frac{0.00784}{0.0376} = 0.2085.$$