Lecture 1, Jan 8, 2024

Events and Probability

- The *sample space* S is the set of all outcomes for an experiment
  - An outcome s is a member of the sample space: $s \in S$
- An *event* is a set of outcomes that satisfy a certain condition
  - Event A is a subset of S: $A \subseteq S$
- The *complement* of A is the set of all elements in S that are not in A, denoted A^c; note $A \cup A^C = S$ and $A \cap A^C = \varnothing$
- Given two events A and B, the event of either occurring is denoted $A \cup B$; both occurring is $A \cap B$
  - We can break $A \cup B$ into 3 parts: outcomes in A only, outcomes in both A and B, and outcomes in B only; these are mutually exclusive
  - $A \cup B = (A \cap B^C) \cup (A^C \cap B) \cup (A \cap B)$
- $C \subseteq A$ means event $C \implies A$
- Associate with each event a *probability*, satisfying the following:
  1. $P[A] \geq 0 \forall A$
  2. $P[S] = 1$
  3. $A \cap B = \varnothing \iff P[A \cup B] = P[A] + P[B]$
     - Note if A and B are not mutually exclusive then $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

Conditional Probability and Bayes' Rule

- The probability of A *conditioned on* B is defined as $P[A|B] = \frac{P[A \cap B]}{P[B]}$ (assuming $P[B] \neq 0$)
  - $P[A|B]$ is known as the *a posteriori* probability
  - By symmetry, $P[B|A] = \frac{P[A \cap B]}{P[A]}$
- This gives the *product formula*: $P[A \cap B] = P[A|B]P[B] = P[B|A]P[A]$
- Events A and B are *independent* (denoted $A \perp B$) iff $P[A|B] = P[A]$, equivalently $P[A \cap B] = P[A]P[B]$

Partitioning

- A *partition* of S is sets $H_1, \dots, H_n$ such that $S = H_1 \cup H_2 \dots \cup H_n$ and $i \neq j \implies H_i \cap H_j = \varnothing$
- Since the H sets are mutually exclusive, $A \cap H_i$ are also mutually exclusive
  - Then $P[A] = P[A \cap H_1] + \cdots + P[A \cap H_n] = P[A|H_1]P[H_1] + \cdots + P[A|H_n]P[H_n]$
  - This is the *total probability theorem*
- Now we can find $P[H_i|A] = \frac{P[A \cap H_i]}{P[A]} = \frac{P[A|H_i]P[H_i]}{\sum_j P[A|H_j]P[H_j]}$
  - This is the *Bayesian* approach
  - The *frequentist* approach assumes no knowledge about the underlying $P[H_i]$ so we can only maximize $P[A|H_i]$