Introduction to Machine Learning CS 436/580 L

Arti Ramesh
Binghamton University

Review of Probability and Statistics 101

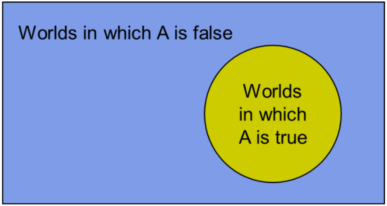
Elements of Probability Theory

- Events, Sample Space and Random Variables
- Axioms of Probability
- Independent Events
- Conditional Probability
- Bayes Theorem
- Joint Probability Distribution
- Expectations and Variance
- Independence and Conditional Independence
- Continuous versus Discrete Distributions
  - Common Continuous and Discrete Distributions

Events, Sample Space and Random Variables

- A sample space is a set of possible outcomes in your domain.
  - All possible entries in a truth table.
  - Can be Infinite. Example: Set of Real numbers
- Random Variable is a function defined over the sample space S
  - A Boolean random variable X: $S \rightarrow \{True, False\}$
  - Stock price of Google G: $S \rightarrow$ Set of Reals
- An Event is a subset of S
  - A subset of S for which $X = True$.
  - Stock price of Google is between 575 and 580.

Events, Sample Space and Random Variables: Picture



$P(A)$ is the area of the oval

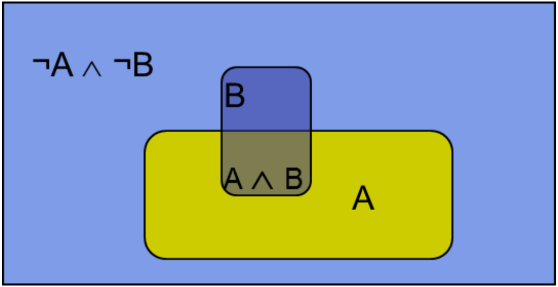
Sample Space: The Rectangle. Random variable: A. Event: A is *True*

Probability: A real function defined over the events in the sample space.

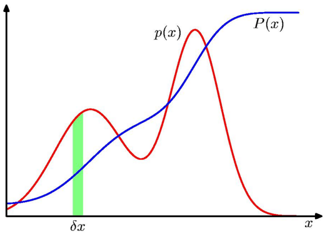
Axioms of Probability

Four Axioms of Probability:

- $0 \leq P(A) \leq 1$
- $P(True) = 1$ (i.e., an event in which all outcomes occur)
- $P(False) = 0$ (i.e., an event in no outcomes occur)
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Probability Densities



- Probability Density:

$$p(x \in (a,b)) = \int_a^b p(x)dx$$

- Cumulative Distribution Function: $P(z) = \int_{-\infty}^{z} p(x)dx$

Such that:

- $p(x) \geq 0$
- $\int_{-\infty}^{\infty} p(x)dx = 1$

Probability Mass Functions

- $A_1, \ldots, A_n$ is a set of mutually exclusive events such that
$$\sum_{i=1}^{n} P(A_i) = 1$$
- P is called a probability mass function or a probability distribution.
- Each A_i can be regarded as specific value in the discretization of a continuous quantity.

Sum Rule

- $0 \leq P(A) \leq 1$
- $P(True) = 1$ (i.e., an event in which all outcomes occur)
- $P(False) = 0$ (i.e., an event in no outcomes occur)
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

To prove that:

1. $P(A) = 1 - P(\neg A)$
2. $P(A) = P(A \wedge B) + P(A \wedge \neg B)$

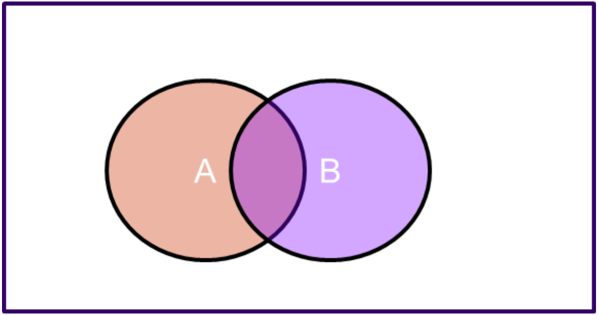
SUM RULE:

$$P(A) = \sum_{i=1}^{n} P(A \wedge B_i)$$

where $\{B_1, \ldots, B_n\}$ is a set of of mutually exclusive and exhaustive events.

Conditional Probability

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Chain Rule

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \wedge B) = P(A|B)P(B)$$

$$P(A \wedge B \wedge C) = P(A|B \wedge C)P(B|C)P(C)$$

$$P(A_1 \wedge A_2 \wedge \ldots \wedge A_n) = \prod_{i=1}^{n} P(A_i|A_1 \wedge \ldots \wedge A_{i-1})$$

Independence and Conditional Independence

Independence:

- Two events are independent if $P(A \wedge B) = P(A)P(B)$
- Implies that: $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- Knowing A tells me nothing about B and vice versa.
- A: Getting a 3 on the face of a die.
- B: New England Patriots win the Superbowl.

Conditional Independence:

- A and C are conditionally independent given B iff $P(A|B \wedge C) = P(A|B)$
- Knowing C tells us nothing about A given B.

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Proof.

$P(A|B) = \frac{P(A \wedge B)}{P(B)}$ – (1)

$P(B|A) = \frac{P(A \wedge B)}{P(A)}$ – (2)

Therefore,

$P(A \wedge B) = P(B|A)P(A)$ – (3)

Substituting $P(A \wedge B)$ in Equation (1), we get Bayes Rule. $\square$

Other Forms of Bayes Rule

Form 1:

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(A \wedge B) + P(\neg A \wedge B)} && (1) \\
&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} && (2)
\end{aligned}
$$

Form 2:

$$
P(A|B \wedge C) = \frac{P(B|A \wedge C)P(A \wedge C)}{P(B \wedge C)}
$$

Applying Bayes Rule: Example

- The probability that a person fails a lie detector test given that he/she is cheats on a test is 0.98. The probability that a person fails the test given that he/she does not cheat on the test is 0.05.
- You are a CS graduate student and the probability that a CS graduate student will cheat on a test is 1 in 10000.
- A person will be expelled from the university if the probability that they cheat is greater than 0.005 (i.e., $> 0.5\%$).

Today, you find out that you have failed the lie detector test. Convince the university that they should not expel you.

Another Interpretation of the Bayes Rule

$$posterior = \frac{likelihood \times prior}{Probability\ of\ evidence}$$

$$P(Cheating = yes | Test = Fail) = \frac{P(Test = Fail | Cheating = yes) \times P(Cheating = yes)}{P(Test = Fail)}$$

- Prior probability of cheating on a test
- Likelihood of failing the test given that a person is cheating
- Test=Fail is the evidence

Expectation and Variance

Expectation:

$$\mathbb{E}[f] = \sum_x p(x)f(x)$$

$$\mathbb{E}[f] = \int p(x)f(x)dx$$

Conditional Expectation:

$$\mathbb{E}[f|y] = \sum_x p(x|y)f(x)$$

Variance:

$$\mathrm{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

Joint Distribution

- Assign a probability value to joint assignments to random variables.
- If all variables are discrete, we consider Cartesian product of their sets of values For Boolean variables, we attach a value to each row of a truth table
- The sum of probabilities should sum to 1.

Outlook	Humidity	Tennis?	Value
Sunny	High	Yes	0.05
Sunny	High	No	0.2
Sunny	Normal	Yes	0.2
Sunny	Normal	No	0.1
Windy	High	Yes	0.2
Windy	High	No	0.05
Windy	Normal	Yes	0.05
Windy	Normal	No	0.15

The Joint Distribution

Represents complete knowledge about the domain
Can be used to answer any question that you might have about the domain

- $P(Event)$ = Sum of Probabilities where the Event is True
- $P(Outlook = Sunny) =$
- $P(Humidity = High \wedge Tennis? = No) =$
- $P(Humidity = High | Tennis? = No) =$

Outlook	Humidity	Tennis?	Value
Sunny	High	Yes	0.05
Sunny	High	No	0.2
Sunny	Normal	Yes	0.2
Sunny	Normal	No	0.1
Windy	High	Yes	0.2
Windy	High	No	0.05
Windy	Normal	Yes	0.05
Windy	Normal	No	0.15