Turning Data into Probabilities

Machine Learning is an interdisciplinary field that uses statistics, probability, algorithms to learn from data and provide insights which can be used to build intelligent applications.

In probability theory, an **event** is a set of outcomes of an experiment to which a probability is assigned.

If E represents an event, then P(E) represents the probability that E will occur. A situation where E might happen (*success*) or might not happen (*failure*) is called a ***trial***.

This event can be anything like *tossing a coin, rolling a die* or *pulling a colored ball out of a bag.* In these examples the outcome of the event is random, so the variable that represents the outcome of these events is called a **random variable.**

Theoretical probability on the other hand is given by the number of ways the particular event can occur divided by the total number of possible outcomes. So a head can occur once and possible outcomes are two (head, tail). The true (theoretical) probability of a head is 1/2.

Joint Probability

Probability of events A and B denoted by **P(A and B) or P(A ∩ B)** is the probability that events A and B both occur.

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

This only applies if A and B are independent, which means that if A occurred, that doesn't change the probability of B, and vice versa.

Conditional Probability

Let us consider A and B are not independent, because if A occurred, the probability of B is higher. When A and B are not independent, it is often useful to compute the conditional probability, P (A|B), which is the probability of A given that B occurred:

$$P(A|B) = P(A \cap B)/\ P(B)$$

Bayes’ Theorem

Bayes’s theorem is a relationship between the conditional probabilities of two events.

For example, if we want to find the probability of selling ice cream on a hot and sunny day, Bayes’ theorem gives us the tools to use prior knowledge about the likelihood of selling ice cream on any other type of day (rainy, windy, snowy etc.).

Prior Probability

Likelihood of the evidence 'E' if the Hypothesis 'H' is true

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

Posterior Probability of 'H' given the evidence

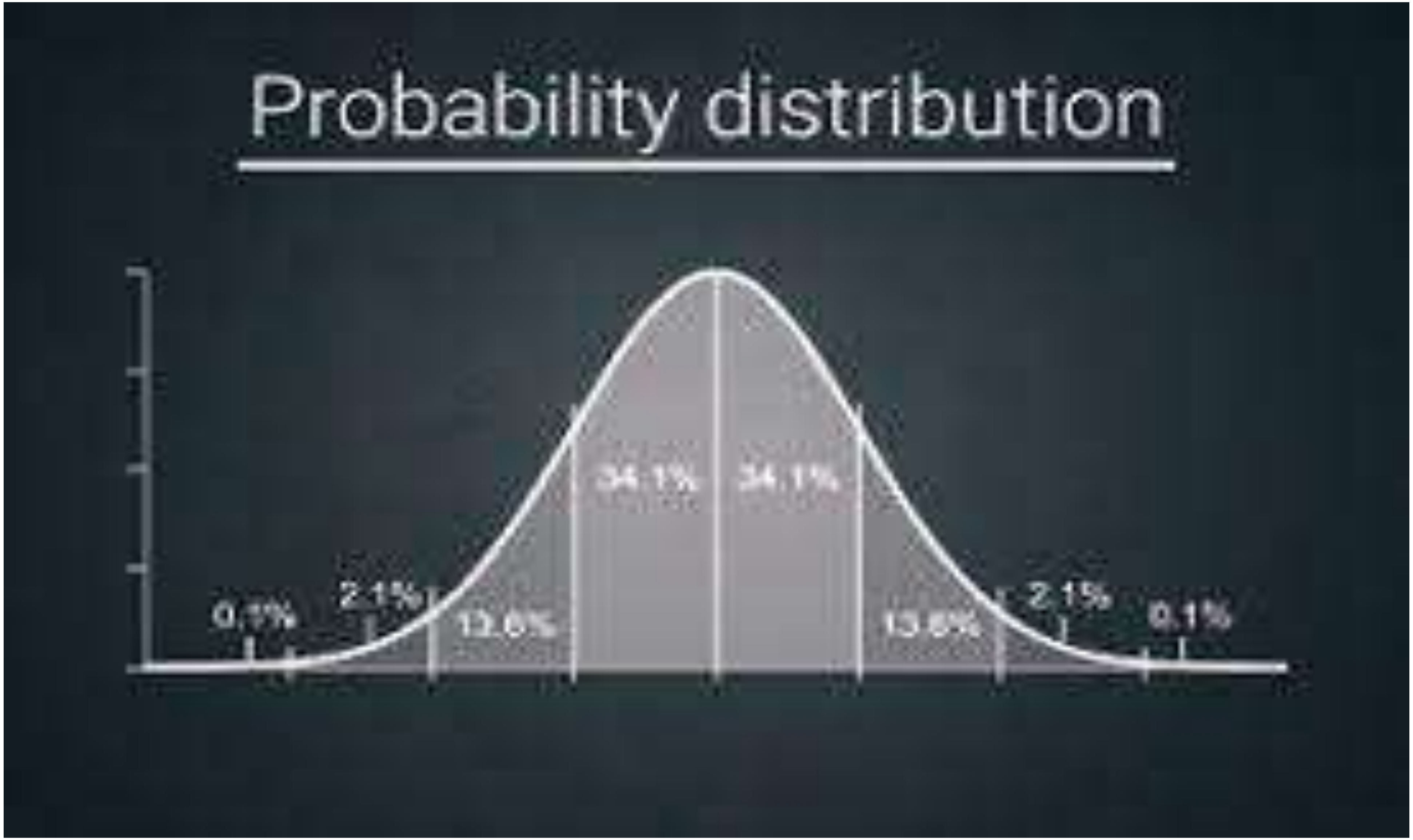
Priori probability that the evidence itself is true

Where H and E are events, *P(H|E)* is the conditional probability that event H occurs given that event E has already occurred.

The probability P(H) in the equation is basically frequency analysis; given our ***prior*** *data* what is the probability of the event occurring.

The P(E\H) in the equation is basically frequency analysis; given our ***prior*** *data* what is the probability of the event occurring.

The P(E) is the probability that the actual **evidence** is true.


Probability distribution
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
0.1%
0.1%