Probability Theory

Huseyin Onder Bektas

Data Scientist at DenizBank

Content

Probability Theory and Fundamental Concepts

- Results of previously unknown events are called random experiments or sometimes called experiments. For example, rolling a die or number of car accident occurring in Istanbul etc.
- All possible outcome of the experiments are called samples space. I use the notation "S" to represent sample space. Head and Tail are sample space of the rolling a die
- Sample space can be discrete or continuous. For example S_1=0,1 is discrete sample space however, $S_2=\{x: \ 0 \leq x \leq 2\}$ is an example of continuous sample space.

Types of Events

- Events can be classified into several groups. If an event occurs during every repetition, then it is called a certain event.
- The event that does not occur during every repetition of the experiment is called the impossible event
- The term of disjoint events are used to define the events that does not have any common point. In other words the intersection of the events A and B is empty.

$$A \cap B = \varnothing \tag{1}$$

What is probability exactly?

- Basically, probability is just a mathematical function that satisfy 3 axiom.
- First, Probability function p of an event A is greater or equal to 0.

$$0 \leq p(A) \leq 1 \tag{2}$$

- Secondly, Probability function p of an sample space S equal to 1.

$$p(S) = 1 \tag{3}$$

- Lastly, Probability of union of events A and B is represented as

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \tag{4}$$

if A and B are disjoint event

$$p(A \cup B) = p(A) + p(B) \tag{5}$$

What is probability exactly?

- Any function (Linear, nonlinear, continuous or discrete, does not matter!) that satisfy these axioms is a probability function.
- So, why do we need a probability function? What is the beauty of these functions?
- We model the results of unknown events by using probability functions.
- As we can see that we can use the set theory to explain the probability definition. If any set is a events, then It can be used probability axioms to explain the uncertainty.
- Example: Prove that if

$$A \subset B, \quad then \;\; p\,(A) \subseteq p\,(B) \tag{6}$$

Hint : Use Probability Axioms

Conditional Probability

- In the mathematical point of wiev, we see that probability is the function that model the uncertain events.
- Sometimes the events that we are interested in, depends on another event. This situation is called conditional probability. In other words, The probability of occurrence of event A depends on the event B.
- Conditional probability is represented by $p(A|B)$. This notation shows the probability of event A due to occurrence of event B.
- Conditional probability is calculated by

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{7}$$

Conditional Probability

Icebreaker Question

What exactly does equation (6) tell us?

- Conditional probability is depend on the intersection of the events. If the intersection is zero (Remember these are disjoint event!) then conditional probability would be zero. According to the equation (6), the If event B occurs, event A does not occur.
- What if $B \subset A$?
  The intersection of A and B is simply B.

$$A \cap B = B \tag{8}$$

In these case, conditional probability is 1 (It is a certain event!)

Conditional Probability

- One more scenario!, What if the conditional probability of A given B is equal to probability of A?

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = p(A) \tag{9}$$

- As we can see in the equatiın 8, event B has no effect on event A. The probability of intersection is a product of individual probabilities. In this case, the events A and B are called independent event.

Required condition for independent events

$$p(A \cap B) = p(A)\, p(B) \tag{10}$$

Conditional Probability

- Conditional probability satisfy the probability axioms.

$$0 \leq p\left(A|B\right) \leq 1 \tag{11}$$

$$p\left(S|B\right) = 1 \tag{12}$$

$$p\left(A_1 \cup A_2|B\right) = p\left(A_1|B\right) + p\left(A_2|B\right) \quad \textit{for} \quad p\left(A_1 \cap A_2\right) = 0 \tag{13}$$

Check this out.

Bayes Rules

- The key point of the conditional probability is the intersection of the events. The intersection term is the common term for two probability function.

$$p\left(A|B\right)=\frac{p\left(A\cap B\right)}{p\left(B\right)}\qquad p\left(B|A\right)=\frac{p\left(B\cap A\right)}{p\left(A\right)}\tag{14}$$

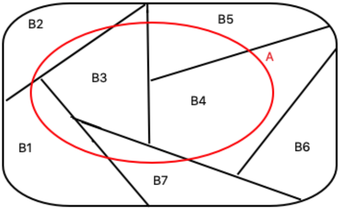
- Just basic math, following Bayes' rule is obtained.

$$p\left(A|B\right)p\left(B\right)=p\left(B|A\right)p\left(A\right)\tag{15}$$

- The importance of Bayes' rule is that it enables us to convert conditional probability just multiplying suitable terms.

Bayes' Rules

- Imagine a set as in the figure below.



- According to the set theory, the probability of A can be written as

$$p(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + .. + P(A \cap B_N) \quad (16)$$

- Here, the intersection set can be written in terms of conditional probabilities. Just a little bit math and manipulation, fallowing equation is obtained.

Bayes' Rules

$$p(A) = p(B_1)P(A|B_1) + p(B_2)P(A|B_2) + .. + p(B_N)P(A|B_N) \tag{17}$$

- People familiar with machine learning understand that this is a classification problem. These are the same configuration. B_1, B_2.. B_N are the classes, A is the data to be classified. Class where conditional probability is highest is the class of A.
- We can calculate the conditional probability for any class.

$$p\left(B_j|A\right) = \frac{p\left(B_j \cap A\right)}{p\left(A\right)} \tag{18}$$

- If we rewrite the the probability of A with respect to B_i i=1,2,..N, fallowing equation is obtained.

$$p\left(B_j|A\right) = \frac{p\left(B_j\right)p\left(A|B_j\right)}{\sum_i^N p\left(B_i\right)p\left(A|B_j\right)} \tag{19}$$

Equation 19 is called Bayes Theorem.

Bayes Rule

Icebreaker Question

As you can see in the equation (16), probability of A is the sum of all possible intersections. However, we assume that intersection of each classes $B_1, B_2, ..B_N$ is zero. For example $B_1 \cap B_2$ or any combination of classes is zero. What if it is not equal to zero?

Questions ?