

# Worksheet -3

## Assignment-3 (Machine Learning)

1. b)
2. c)
3. b)
4. a)
5. a)
6. b)
7. b)
8. d)
9. a)

10. **Bayes' theorem** describes the probability of occurrence of an event related to any condition. It is also considered for the case of [conditional probability](#). Bayes theorem is also known as the formula for the probability of “causes”. For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability. In this article, let us discuss the statement and proof for Bayes theorem, its derivation, formula, and many solved examples.

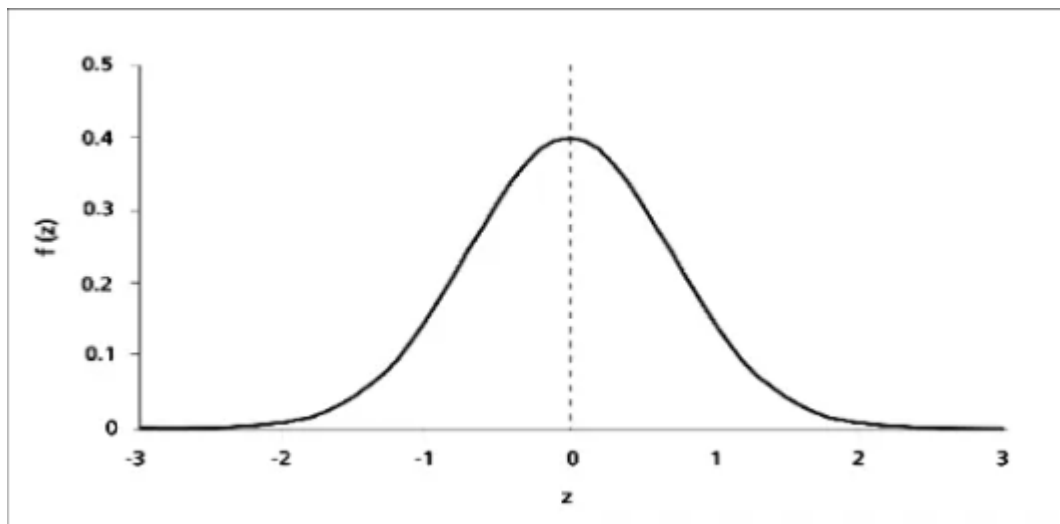
Bayes' theorem is stated mathematically as the following equation:<sup>[3]</sup>

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

where  $A$  and  $B$  are events and  $P(B) \neq 0$ .

- $P(A | B)$  is a [conditional probability](#): the probability of event  $A$  occurring given that  $B$  is true. It is also called the [posterior probability](#) of  $A$  given  $B$ .
- $P(B | A)$  is also a conditional probability: the probability of event  $B$  occurring given that  $A$  is true. It can also be interpreted as the [likelihood](#) of  $A$  given a fixed  $B$  because  $P(B | A) = L(A | B)$ .
- $P(A)$  and  $P(B)$  are the probabilities of observing  $A$  and  $B$  respectively without any given conditions; they are known as the [prior probability](#) and [marginal probability](#).

11. A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The z-score is positive if the value lies above the mean, and negative if it lies below the mean. It is also known as a standard score, because it allows comparison of scores on different kinds of variables by standardizing the distribution. A standard normal distribution (SND) is a normally shaped distribution with a mean of 0 and a standard deviation (SD) of 1.



The Z Score Formula: One Sample

The basic z score formula for a sample is:

$$z = (x - \mu) / \sigma$$

12. A t test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

13. A percentile (or a centile) is a measure used in statistics indicating the value *below which* a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.

The term percentile and the related term *percentile rank* are often used in the reporting of scores from norm-referenced tests. For example, if a score

is at the 86th percentile, where 86 is the percentile rank, it is equal to the value below which 86% of the observations may be found. In contrast, if it is in the 86th percentile, the score is at or below the value of which 86% of the observations may be found. *Every score is in the 100th percentile.* The 25th percentile is also known as the first quartile (Q1), the 50th percentile as the median or second quartile (Q2), and the 75th percentile as the third quartile (Q3). In general, percentiles and quartiles are specific types of quantiles.

14. Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

**The Formula for ANOVA is:**

$$F = \frac{MST}{MSE}$$

**where:**

F = ANOVA coefficient

MST = Mean sum of squares due to treatment

MSE = Mean sum of squares due to error

15. The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as the first example: age, sex, income). When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behaviour.

