PDF & CDF IN MACHINE LEARNING

BY – C. RACHITA

VARIABLES

- Discrete

Discrete variables can only take up countable values or simply put, a discrete variable is a variable whose value is obtained by counting. Examples include number of children, number of petals in a flower etc.
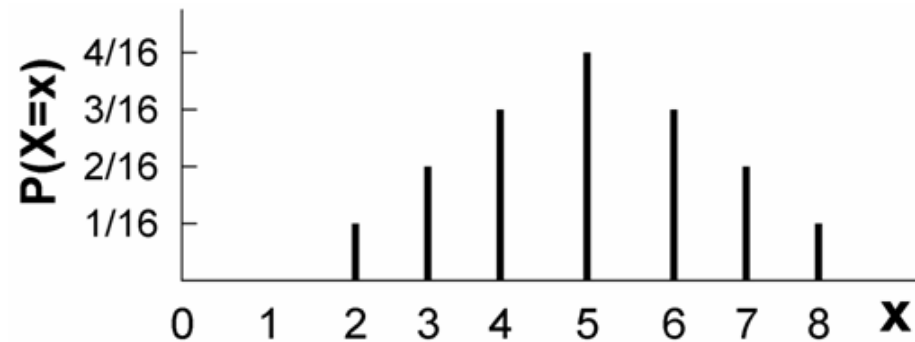
- Continuous

Continuous data is data that can take any value. It is not something that can be counted, for example area of a room, speed of wind etc.

PROBABILITY MASS FUNCTION

- a probability mass function is a function that gives the probability that a discrete random variable is exactly equal to some value.

- Graphically,

PDF – PROBABILITY DENSITY FUNCTION

- a probability density function (PDF) is used to define the random variable's probability coming within a distinct range of values, as opposed to taking on any one value.

- However, unlike a discrete random variable the probability of continuous variable cannot be specified directly; instead calculated as integral (area under the curve) of a small range.

- Mathematically,
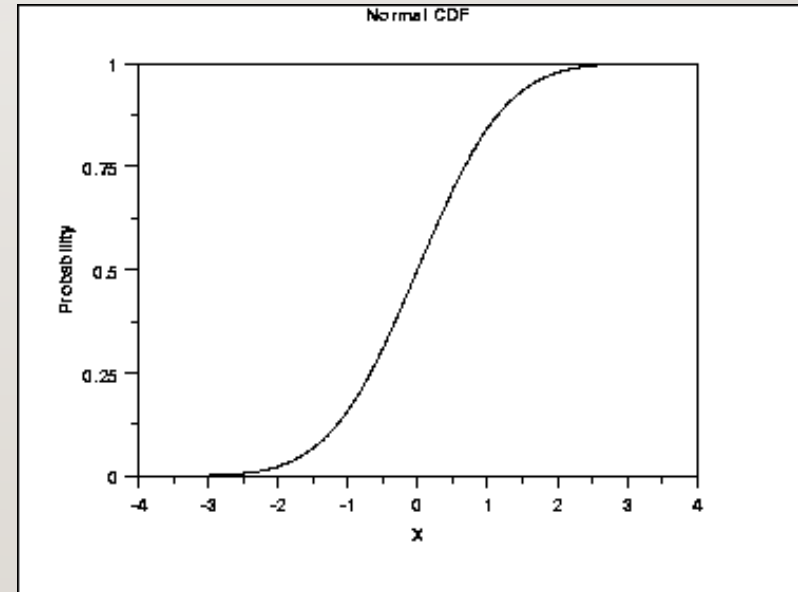
Probability Density Function

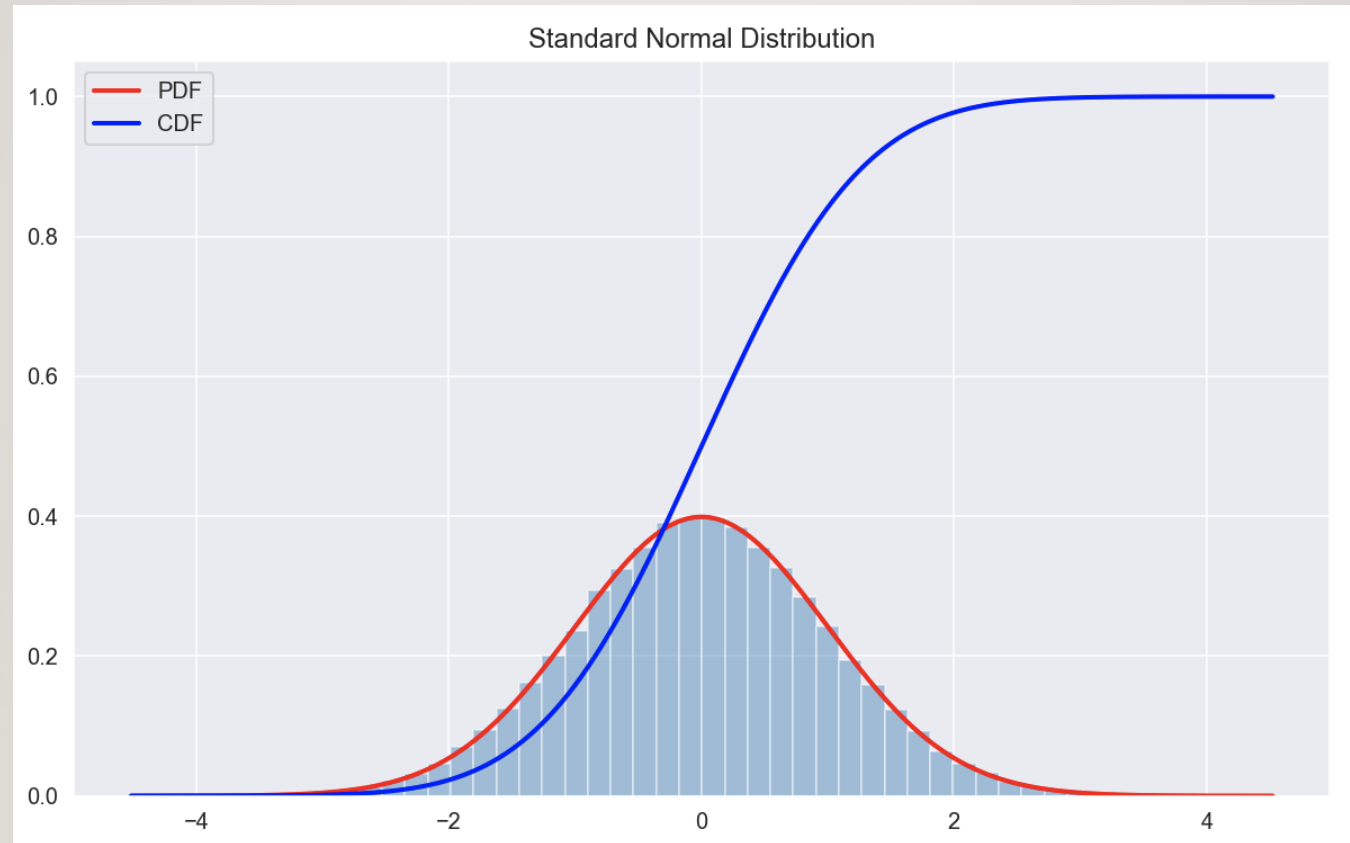$$F(x) = P(a \leq x \leq b) = \int_a^b f(x)dx \geq 0$$

GRAPH
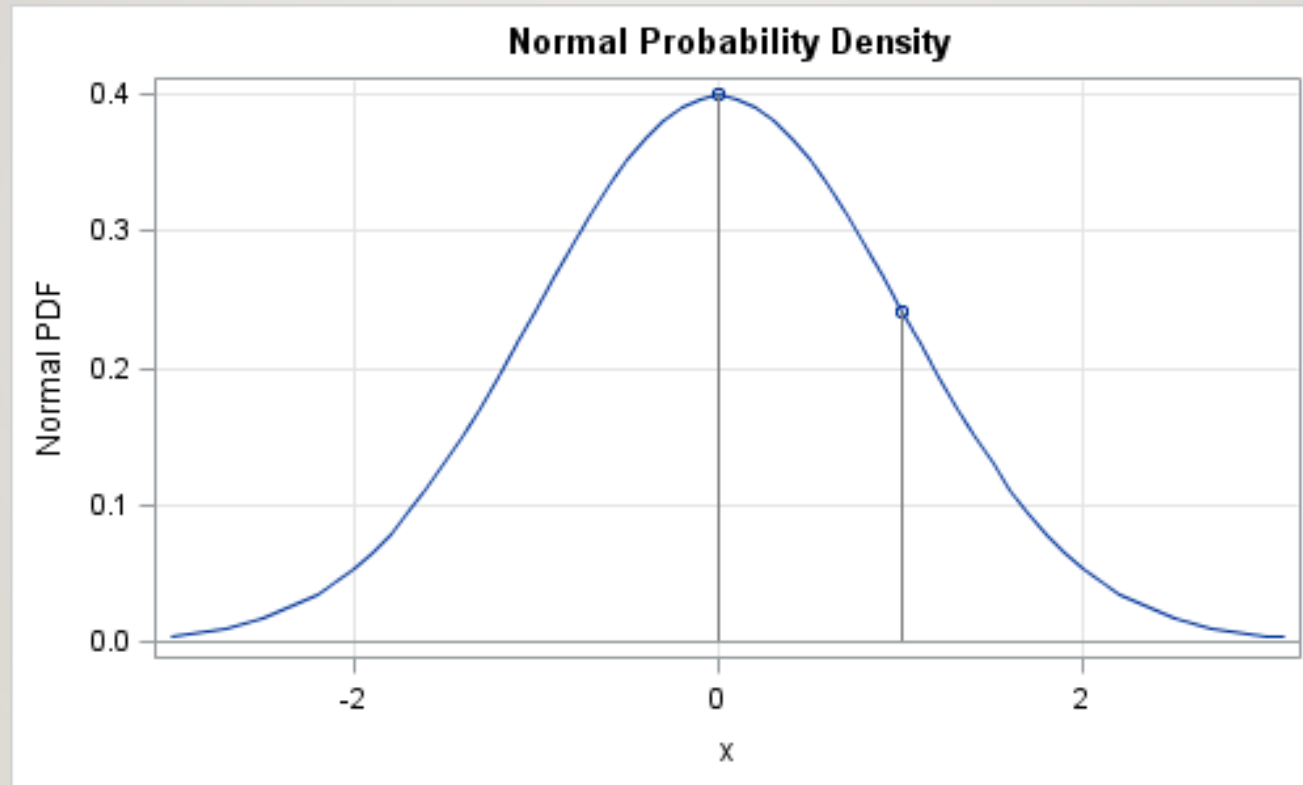
CDF – CUMULATIVE DISTRIBUTION FUNCTION

- What a cumulative distribution function tells is the probability that X will take a value less than or equal to x.

- Mathematically, given as P(X <= x)

COMPARISON

Standard Normal Distribution

DATA ANALYSIS

Q. How can this information be used in data analysis?

Answer: The amount of data points less or greater than a range or within a range gets clearer.

- For example, we need information about how many students can surpass the criteria of height more than 150cm. CFD will specify P(X<150)=0.6. This means that about 60% students fall short and the rest 40% have surpassed the criteria.

- Similarly, information about amount of students falling in the category of heights 140 to 160 can be found out using PDF. If it specifies that P(140<x<160)=0.3, we know that about 30% of students fall in our required category.

This is how it can used in real world problems.

IMPLEMENTATION

- Head over to the .ipynb file for codes and explanation.