Introduction

Supervised Learning ← · · · · → Data Science Maths

Hyper-parameter ← · · · → Machine Learning

Data Science

Unsupervised Learning ← · · · → Ensemble Learning

Data Cleaning ← · · · → Deep Learning
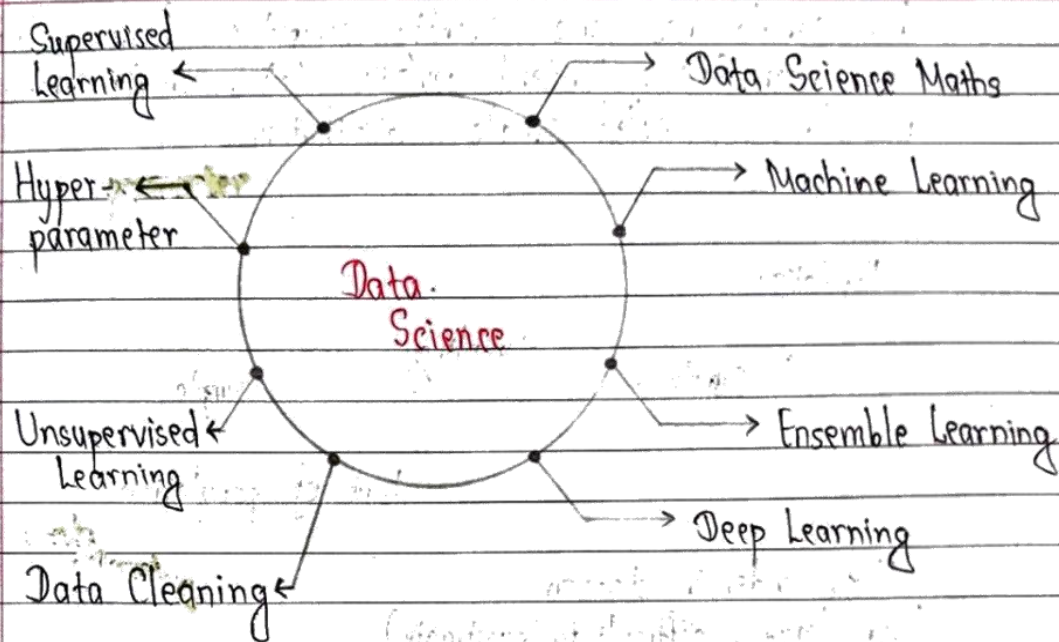
1. Population and Sample Data
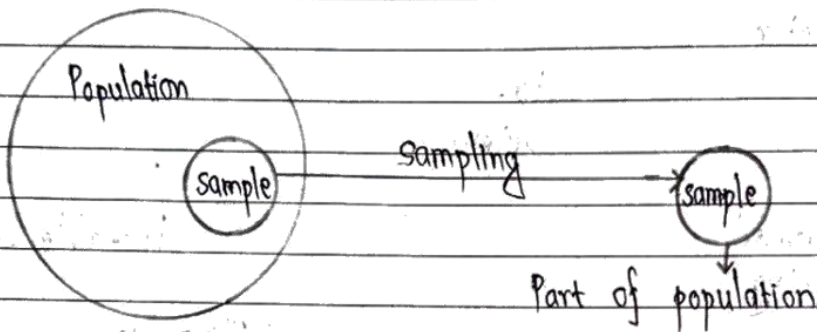
Population (N):

A population includes all of the elements from a set of data. The population is a whole set of values or individuals, you are interested in.

Population characteristics are mean (μ), standerd deviation (σ), proportion (P), median, percentiles, etc. The value of a population characteristics is fixed. These characteristics are called population distribution.

Sample (n):

The sample is subset of population and is the set of values you actually use in your estimation.

This sample has some quantity computed from values
e.g. mean $(\bar{x})$, standerd deviation (s), sample proportion,
etc. This is called sample distribution.

Population

Sample ——— Sampling ——→ sample

Part of population

Data under investigation
(very large, difficult to investigate)