Siddhardhan
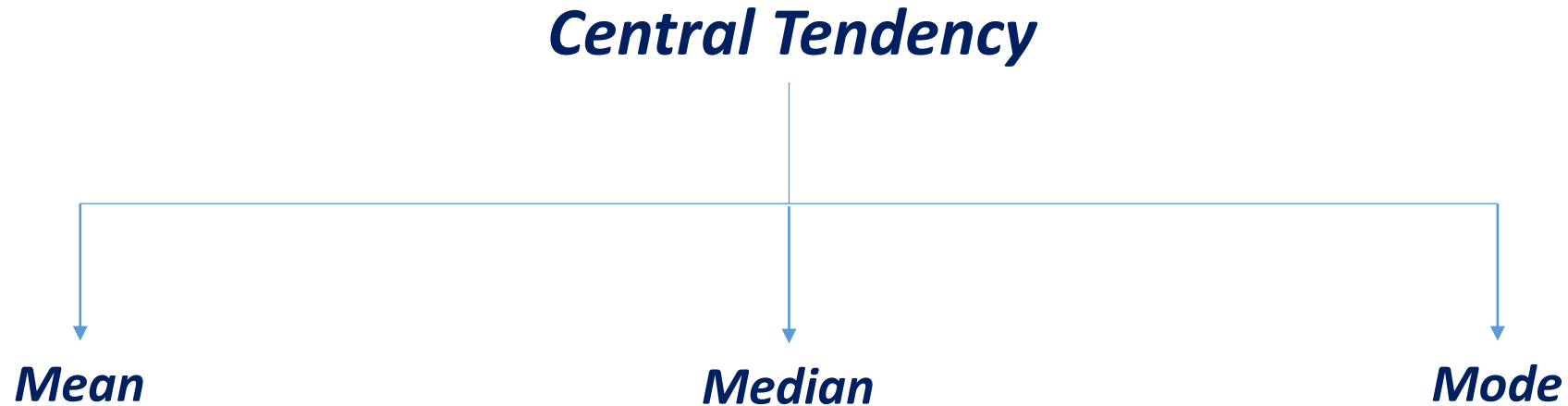
Measure of Central Tendencies: Mean, Median & Mode

Math for Machine Learning

Central Tendency

Central Tendency:

*A measure of **central tendency** is a value that represents the center point or typical value of a dataset. It is a value that summarizes the data.*

Central Tendency

Mean   **Median**   **Mode**

Central Tendencies

Mean

Mean or arithmetic mean is the sum of values divided by the number of values.

$$M = \frac{\Sigma x}{N}$$

Heights

160
172
165
168
174

$$\frac{160+172+165+168+174}{5}$$

Mean = 167.8

Median

The **median** is the **middle** value in the list of numbers. To find the median, the numbers have to be listed in numerical order from smallest to largest.

160 165 168 172 174

160 165 168 172 174 176

$$\frac{168+172}{2} = 170$$

Median = 170

Mode

The **mode** is the value that occurs most often. If no number in the list is repeated, then there is no mode for the list.

Heights

160
172
160
168
174

Mode = 160

Central Tendencies in Data Pre-Processing

Central Tendencies are very useful in **handling the missing values** in a dataset

Mean : Missing values in a dataset can be replaced with **mean** value, if the data is uniformly distributed.

Right Skewed Distribution

Median : Missing values in a dataset can be replaced with **median** value, if the data is skewed.

Mode : Missing values in a dataset can be replaced with **mode** value, if the data is skewed. Missing categorical values can also be replaced with **mode** value.