



Data – Machine Learning Skewness, Distribution

Diptangshu Banik

Twitter - @dipbanik



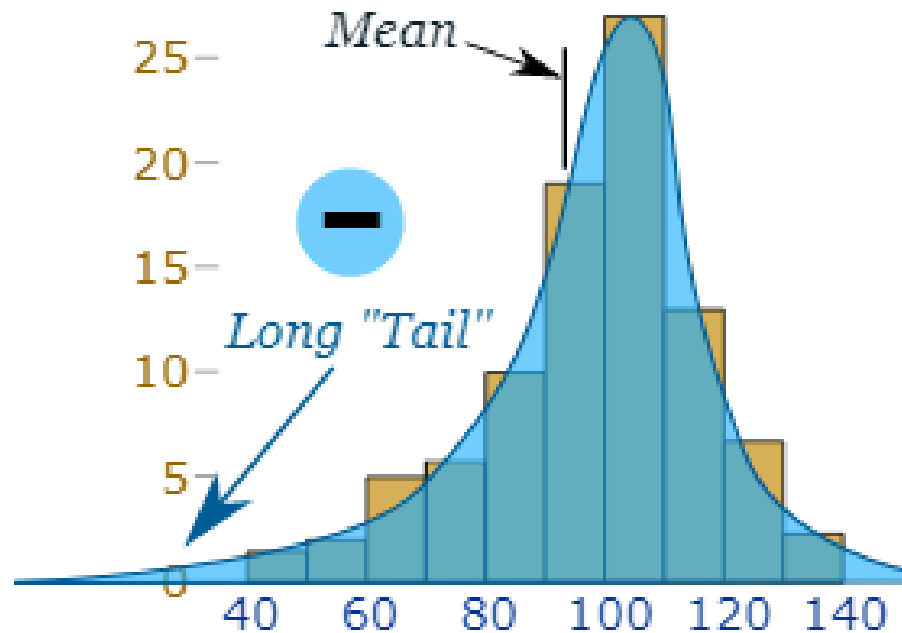
Overview

- Mean – Average of all data points
- Median – Middle value of data points
- Mode – Most common element in the set
- Standard Deviation – Average amount we expect a point to differ or deviate from the mean.



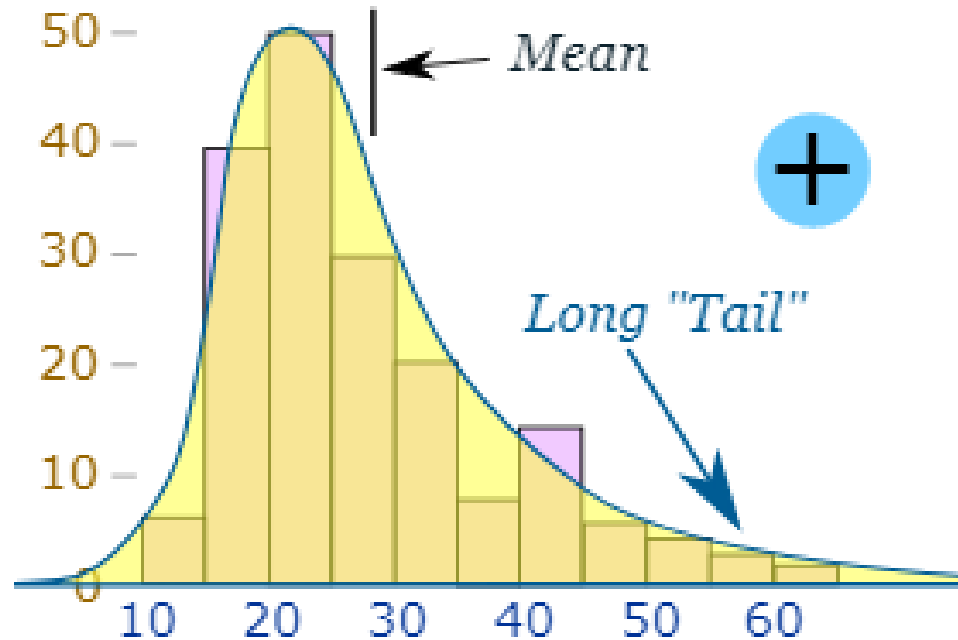
Skewed Data

- Left Skewed Data –



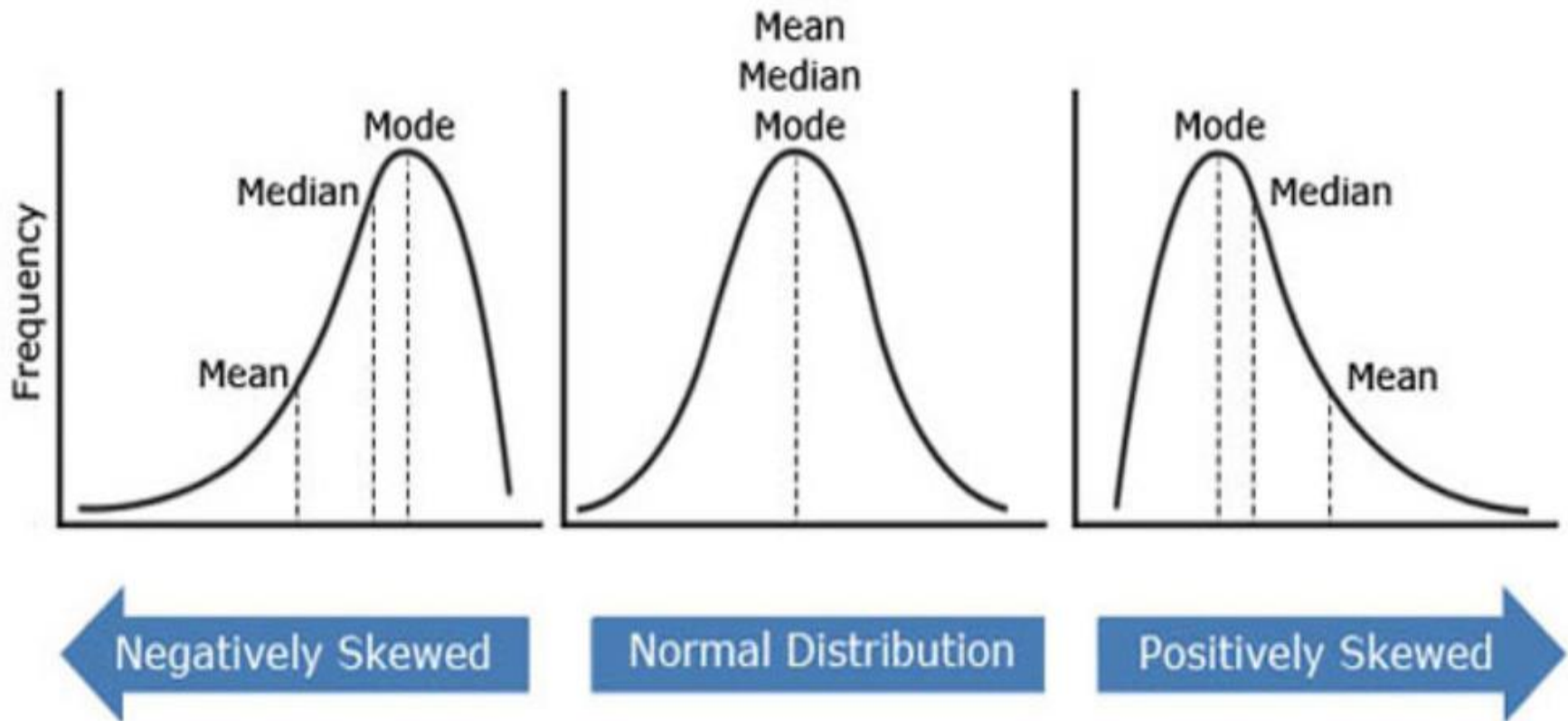


- Right Skewed Data –





Symmetric Data –





Standard Deviation (σ)

- Used to understand the distribution of data.
- Use when distribution of data is approximately normal. Not when there are multiple dunes.
- Low – Data is closely clustered around mean.
- High – Data is dispersed over a wider range of values.



Standard Deviation (σ)

- Used to understand the distribution of data.
- Use when distribution of data is approximately normal. Not when there are multiple dunes.
- Low – Data is closely clustered around mean.
- High – Data is dispersed over a wider range of values.