Machine Learning Course | Arabic

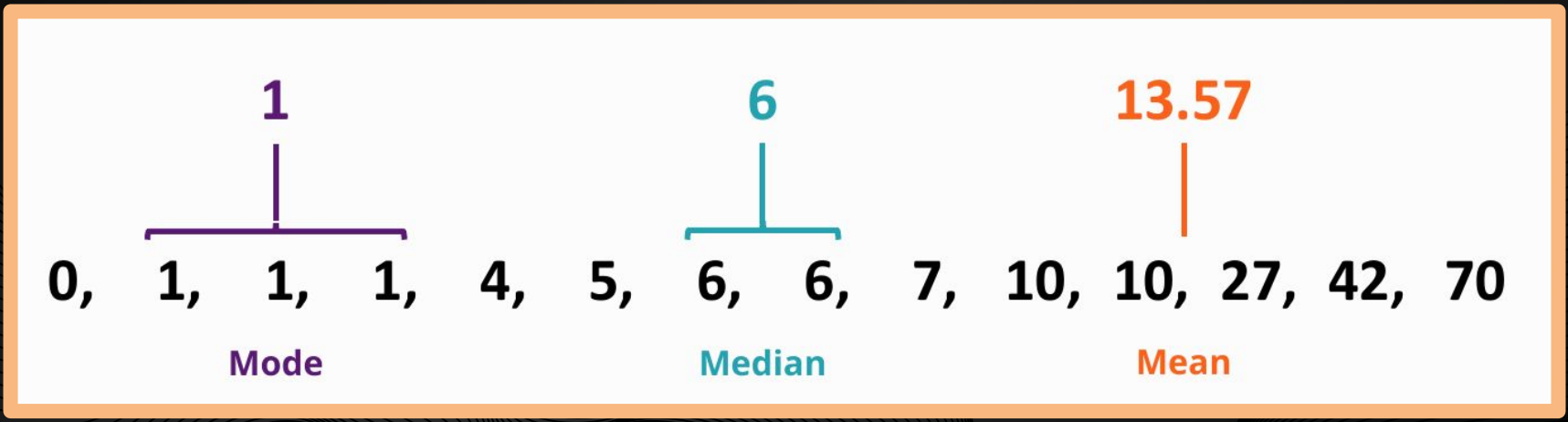
Data Preprocessing

Level - 01

Link to Lecture on Youtube

01 MEAN, MEDIAN, MODE

Introduction



Arithmetic Mean

Is the "**average**" you're used to, where you add up all the numbers and then divide by the total number of elements.

$$\text{Arithmetic mean} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Ex.1: [1,2,3,4,5,6,7,8,9,10]

Ex.2: [1,2,3,4,5,6,7,8,9,10,100]

Solution:

Mean = (1+2+3+4+5+6+7+8+9+10) / 10 = 5.5

Mean = (1+2+3+4+5+6+7+8+9+10+100) / 11 = 14.01

Geometric Mean

Is an "n_{th}" root of the product of all numbers in a list.

$$\text{Geometric mean} = \sqrt[n]{X_1 \times X_2 \times \ldots \times X_n}$$

Ex.1: [1,2,3,4,5,6,7,8,9,10]

Ex.2: [1,2,3,4,5,6,7,8,9,10,100]

Solution:

Mean = (1*2*3*4*5*6*7*8*9*10) ^ 1 / 10 = 4.5

Mean = (1*2*3*4*5*6*7*8*9*10*100) ^ 1 / 11 = 6

Mode

Is the “**most repeated**” value in the list. If no one is repeated, then any number is the mode.

Solution:

Median = (5+6)/2 = 5.5

Median = 6

Median

Is the "**middle**" value in the list of “**sorted**” numbers.

Ex.1: [1,2,3,4,5,6,7,8,9,10]

Ex.2: [1,2,3,4,5,6,7,8,9,10,100]

Mode = any number

Mode = any number

Use Cases

Mean:

- Company performance.

Median:

- Diminishes the effect of outliers.
- Employee salaries.

Mode:

- Call-center busy hours.
- Restaurant busy hours.

Ex.

Mohamed wants to find shares to invest. He is a big fan of Apple Inc. He knows that the company has strong financial resources. But, to make sure that this investment will bring him a great return, he decided to check the stock's performance in the past.

He collected monthly revenue data for the past five months to verify as follows

Ex.

	Stock Price	Return (%)	Return (decimal)
December	167.90	N/A	
January	166.11	-1.07%	0.99
February	176.72	6.39%	1.06
March	167.14	-5.42%	0.95
April	164.63	-1.50%	0.98
May	186.15	13.07%	1.13
June	185.11	-0.56%	0.99

Calculate the Mean.

Ex.

1- Arithmetic mean:

$$\text{Arithmetic mean} = \frac{-1.07\% + 6.39\% - 5.42\% - 1.50\% + 13.07\% - 0.56\%}{6} = 1.82\%$$

2- Geometric mean:

$$\text{Geometric mean} = \sqrt[6]{0.99 \times 1.06 \times 0.95 \times 0.98 \times 1.13 \times 0.99} = 1.0164 \text{ or } 1.64\%$$

Conclusion:

we can see that the **Geometric Mean** could deal with the **volatility** in the data samples and got more accurate result.

Thank You!

Do you have any questions?

Write them in the comments

hozaifazaki99@gmail.com