✡ Mean, Median, Mode and Range for Machine learning

Mean(Avg) Dataset

Observation	Car-Speed
10.AM	80 KM/h
10:15 AM	90 "
10:30AM	75 "
10:45 AM	100 "
10:55 AM	80 "
"	"
"	"

Averege Speed

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{7}(80+90+75+100+80+60+50)$$

$$= \frac{1}{7}(535)$$

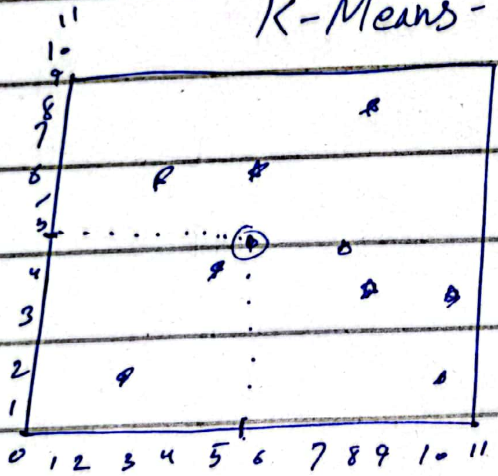
$$= 76.43 \text{ KM/H}$$

Mean Role in ML

Run	Accuracy
1	90%
2	92 %
3	90 %
⋮	⋮
10	90%

Average accuracy = 90.7%

K-Means-Clustering



X_1	X_2
10	4
7	4
10	2
2	6
7	9
4	4
6	5
2	2
4	6
4	6

$Avg(X_1) = 5.6$ | $Avg(X_2) = 4.8$

[dataset]

Median

Observation	Speed
10:00 AM	80 KM/hr
10:15 AM	90 "
10:25 AM	75 "
10:55 AM	100 "
11:15 AM	80 "
11:40 AM	60 "
11:55 AM	50 "

$$\text{Median} = \begin{cases} \dfrac{(N+1)}{2}^{th} \text{ term, when } N \text{ is odd} \\[2ex] \dfrac{\dfrac{N}{2}^{th} \text{ term} + \left(\dfrac{N}{2}+1\right)}{2} \text{ where } N \text{ even} \end{cases}$$

1st Sort speed lowest to highest

Median   odd (len)

$\text{Median} = 50, 60, 75, 80, 80, 90, 100$

$= 80 \text{ KM/hr}$

$\left(\dfrac{7+1}{2}\right) = \dfrac{8}{2}$ 4 ide

if even

$= \dfrac{4^{th} + 5^{th}}{2} = \dfrac{80+80}{2} = 80 \text{ KM/hr}$

Scanned with CamScanner

Median Role in ML.

Run	Accuracy
1	40%
2	42 %
3	90 %
⋮	⋮
10	90 %

Mean Accuracy = 83.4%

Median " = 50, 60, ~~49~~ 90, 90, 92

$= \frac{90+90}{2} = 90$

The Main Advantage of the Median over the mean is that median is less susceptible to outliers.

Dataset,

* MODE

Observation	Speed
10:00AM	80 KM/hr
10:15 AM	90 "
10:25 AM	75 "
10:55 AM	100 "
11:15 AM	80 "
11: 40 AM	60 "
11 : 55 AM	50 "

The Mode value is the value that appears the most number of times.

Mode of speed
= 50, 60, 75, 80, 80, 90, 100
= 80 KM/hr

Dataset

★ RANGE

Observation	Car-Speed
10:00AM	80 KM/hr
"	90 "
"	75 "
"	100 "
"	80 "
	60
"	50 "

The difference b/w lowest and highest values.

range-of-car-speed = 100 - 50
= 50

Role in ML

$$X_{new} = \frac{X_i - \min(X)}{\max(X) - \min(X)}$$

Scaled-speed

(80-50)/(100-50)
(90-50)/ "
(75-50)/ "
100 " "
80 " "
60 " "
50 " "

= 0.6, 0.8, 0.5, 1.0, 0.6, 0.2, 0.0

=

☆ Standard deviation and Variance For ML [Imp]

☆ Standard deviation (σ)

It tells us that, how much data point deviate from their mean. More Specifically, It is square root of the average ~~spread~~ squared deviation from the mean.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$$

mean

☆ Variance

It is the average squared deviation from the mean

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$$

* Calculat Standard deviation and Variance

Systollic Blood Pressure (x)	$x - \bar{x}$	$(x-\bar{x})^2$
98	98-125.71= -28	784
140	140 - " = 14	196
130	130 " = 4	16
120	120 " = -6	36
130	130 " = 4	16
102	102 " = 24	576
160	160 " = 34	+ 1156
		= 2780

Step I = $\bar{x} = \frac{\sum_{i=1}^{N} (N_i)}{2} = 125.72$ (Mean)

Variance = $\sigma^2 = 2780/7 = 397.14$

std = σ = Sqrt(397.14) = 1993

(i) Find the Mean (x)

(ii) Find square of its distance to Find mean

(iii) sum the values from Step2

(iv) Divede the number of data points

(v) Take the quare root if computing standard deviation.

Bessel's Correction

For whole population:

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \quad , \sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n}$$

For subpopulation (samples):

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \quad , \sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$