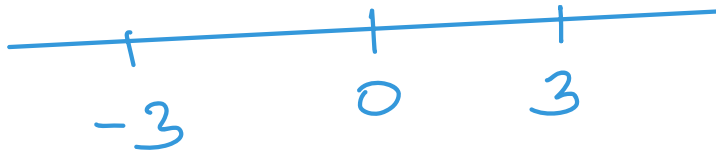


STATISTICS

① Mean (Avg) :

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



$$\mu = \frac{-3 + 0 + 3}{3} = 0$$

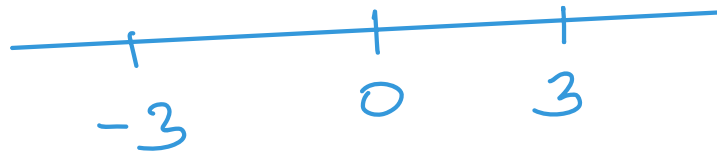
② Variance : How far are datapoints from the mean on an average ?

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

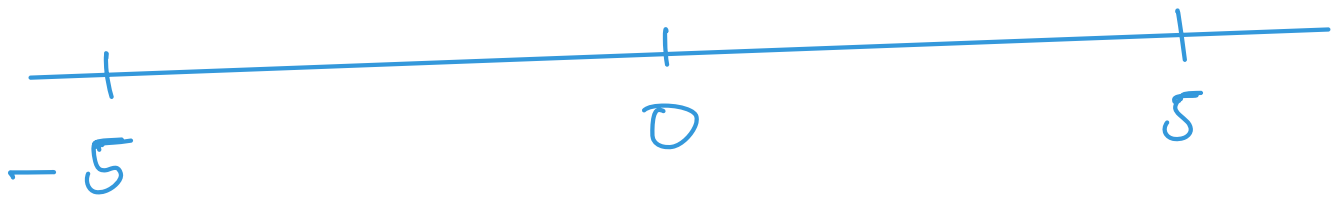
$$\text{Std. dev, } s = \sqrt{V}$$

$$\mu = 0$$

$$\frac{(-3)^2 + 3^2}{3}$$



$$V = 6$$

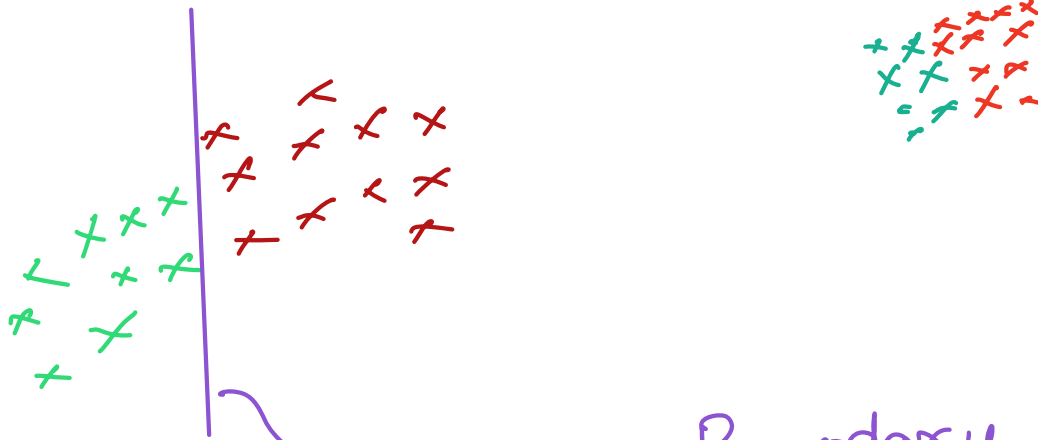


$$\mu = 0$$

$$V = \frac{(-5)^2 + 0^2 + 5^2}{3}$$

$$\Rightarrow V = \frac{50}{3} \sim 16.67$$

Why is high variance desirable?



Decision Boundary
Easier
to draw
when variance
is more !

③ Median

- Middle value in a sorted list
- While mean is sensitive to outliers, median isn't

Eg: Income in Mumbai

④ Mean Absolute Deviation (MAD)

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

△
Less preferred than S.D
because modulus isn't
differentiable

⑤ Median Absolute Deviation

$$\text{Median}(|x_1 - m|, |x_2 - m|, \dots, |x_n - m|)$$



Median

→ Robust to outliers
while capturing variability

Covariance (Matrix)

— Relationship b/w 2 variables/features

$$S_{x,z} = \frac{\sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z})}{n-1}$$

Δ
samples / rows of the dataset

→ +ve \Rightarrow Same direction
(both values are high/low)

→ -ve \Rightarrow Opp direction

Covariance Matrix

$$\Sigma = \begin{bmatrix} S_x^2 & S_{x,z} \\ S_{z,x} & S_z^2 \end{bmatrix}$$

Variance

Covariance

The diagram shows a 2x2 covariance matrix Σ enclosed in blue square brackets. The elements are S_x^2 (top-left), $S_{x,z}$ (top-right), $S_{z,x}$ (bottom-left), and S_z^2 (bottom-right). A green parallelogram is drawn around the diagonal elements S_x^2 and S_z^2 , with a green arrow pointing to the word 'Variance'. Two purple curved arrows point from the off-diagonal elements $S_{x,z}$ and $S_{z,x}$ to the word 'Covariance'.