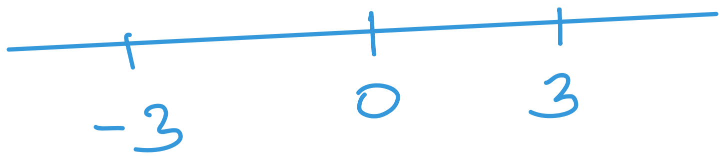
STATISTICS

① <u>Mean</u> (Avg) : $\boxed{\mu = \frac{1}{n}\sum_{i=1}^{n} x_i}$



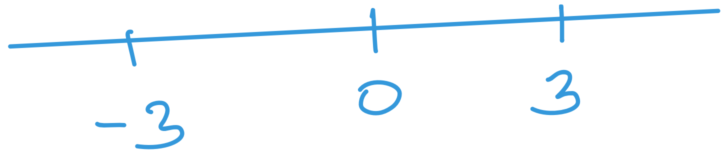
$\mu = \frac{-3+0+3}{3}$

$= 0$

② <u>Variance</u> : How far are datapoints from the mean on an average?

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

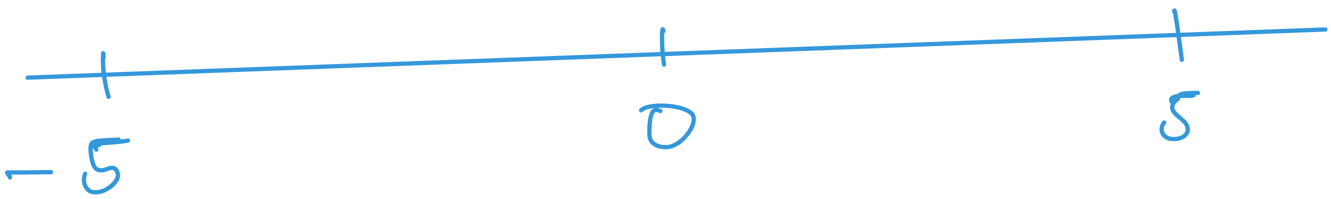
Std. dev , $s = \sqrt{V}$



$$\boxed{\mu = 0}$$

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

$$\boxed{V = 6}$$

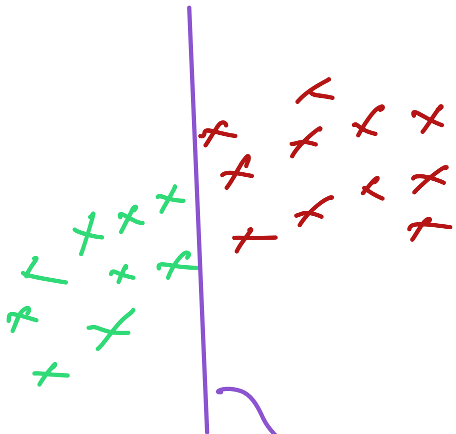


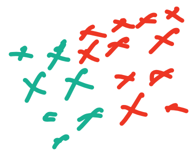
$$\boxed{\mu = 0}$$

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

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

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 oultiers, 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 preffered than S.D because modulus isn't differentiable

⑤ Median Absolute Deviation

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

Median

→ Robust to outliers while capturing variability

Covariance (Matrix)

- Relationship b/w 2 variables/features

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

samples/rows of the dataset

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

→ -ve ⇒ Opp direction

Covariance Matrix

Variance

Co-variance

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