Apurva Mandalika
3340035575

Assignment - 1

Question 1. Linear Algebra

$X_{M\times N}$ $Y_{N\times N}$ $a_{M\times 1}$ $b_{N\times 1}$ $s \rightarrow$ scalar.

(i) $XY \Rightarrow X_{M\times N} \; Y_{N\times N}$
$N = N$

$\Rightarrow$ operation is valid and can be computed
The size of the result is $\underline{\underline{M\times N}}$ [no. of ~~columns~~ rows of X $\times$ no. of columns of Y]

(ii) $YX \Rightarrow Y_{N\times N} \; X_{M\times N}$
$N \neq M$

Not a valid operation, because the no. of columns of $Y \neq$ no. of rows of X.

(iii) $YX^T \Rightarrow Y_{N\times N} \; X^T_{N\times M}$
$N = N$

Valid operation, can be computed and the size of the result is $\underline{N\times M}$.

(iv) $aX \Rightarrow a_{M\times 1} \; X_{M\times N}$
$1 \neq M$

Not a valid operation.

(v) $a^T X$ $\quad a^T_{1\times M} \; X_{M \times N}$.

$M = M$

Valid operation and result can be computed. The dimensions of the result are $\underline{1 \times N}$.

(vi) $a X^T$ $\quad a_{M\times 1} \; X^T_{N\times M}$

$1 \neq N$

Not a valid operation.

(vii) $a^T b$ $\quad a^T_{1\times M} \; b_{N\times 1}$

$M \neq N$

Not a valid operation.

(viii) $b^T b$ $\quad b^T_{1\times N} \; b_{N\times 1}$

$N = N$

Valid operation. Result can be calculated and has the dimension of $\underline{1\times 1}$, i.e. scalar.

(ix) $b b^T$ $\quad b_{N\times 1} \; b^T_{1\times N}$

$1 = 1$

Valid operation. Result can be calculated and has the dimension of $\underline{N\times N}$.

(x) $sX + Y$

s scalar, ~~so~~ sX has dimension of X.

$$X_{M \times N} + Y_{N \times N}$$

Since the dimensions do not match, it is not a valid operation.

Question $\boxed{3.}$ $A = [4, 16, 4, 5, 1]^T \qquad B = [8, 3, 4, 8, 7]^T$

(1.) Mean - Sum of all values divided by no. of values.

A: $\text{Mean}_A = \frac{\Sigma A_i}{n} = \frac{4+16+4+5+1}{5} = \frac{30}{5} = \underline{\underline{6}}$

B: $\text{Mean}_B = \frac{\Sigma B_i}{n} = \frac{8+3+4+8+7}{5} = \frac{30}{5} = \underline{\underline{6}}$

(2.) Median - Middle value when data is sorted. If even no. of values then avg. of the 2 middle values.

A: $1, 4, 4, 5, 16$ → Median_A

B: $3, 4, 7, 8, 8$ → Median_B

(3.) Range - Difference between the max and min values.

$\text{Range}_A = 16 - 1 = \underline{\underline{15}} \qquad \text{Range}_B = 8 - 3 = \underline{\underline{5}}$

(4.) Variance - Sum of the squares of difference between each value and mean, divided by no. of values - 1.

$$\text{Variance} = \frac{\Sigma(x_i - \text{Mean})^2}{n-1}$$

A: $$\text{Variance}_A = \frac{(4-6)^2+(16-6)^2+(4-6)^2+(5-6)^2+(1-6)^2}{(5-1)}$$

$$= \frac{(-2)^2+(10)^2+(-2)^2+(-1)^2+(-5)^2}{4}$$

$$= \frac{4+100+4+1+25}{4}$$

$$= \frac{134}{4} = \boxed{33.5}$$

B: $$\text{Variance}_B = \frac{(8-6)^2+(3-6)^2+(4-6)^2+(8-6)^2+(7-6)^2}{(5-1)}$$

$$= \frac{(2)^2+(-3)^2+(-2)^2+(2)^2+(1)^2}{4}$$

$$= \frac{4+9+4+4+1}{4} = \frac{22}{4} = \boxed{5.5}$$

(5.) To know which dataset will be more affected by 18 being an outlier, we can compare how far it is from the mean of each dataset in terms of standard deviation (Z-score).

A: Mean = 6 , Variance = 33.5

$\sigma_A = \sqrt{33.5} \approx 5.79$

$Dis_A = \frac{18-6}{5.79} \approx 2.07$

B: Mean = 6, Variance = 5.5

$\sigma_B = \sqrt{5.5} \approx 2.35$.

$Dis_B = \frac{18-6}{2.35} \approx 5.11$

Since Dis_B is greater than Dis_A, it means 18 will affect dataset B more than dataset A as an outlier.