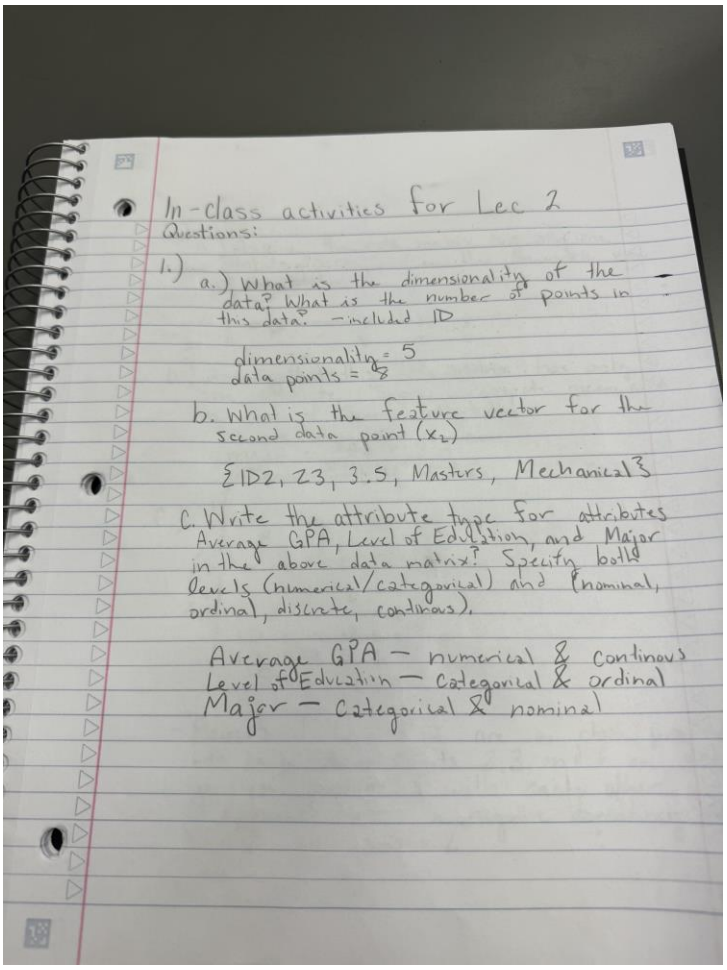


In-class activities for Lec 2

Questions:

1.)

a.) What is the dimensionality of the data? What is the number of points in this data? - included ID

dimensionality = 5
data points = 8

b. What is the feature vector for the second data point (x_2)

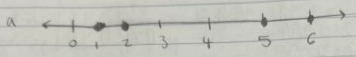
{ID2, 23, 3.5, Masters, Mechanical}

C. Write the attribute type for attributes Average GPA, Level of Education, and Major in the above data matrix? Specify both levels (numerical/categorical) and (nominal, ordinal, discrete, continous).

Average GPA – numerical & continous
Level of Education – categorical & ordinal
Major – categorical & nominal

2.)

a.) Assume the data matrix has only one attribute 'score,' draw the Geometric view of the data.

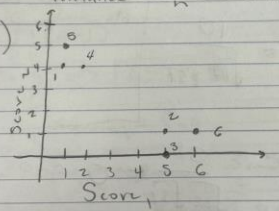


b.) Assume that this data matrix has only one attribute, 'Score,' compute mean, and variance of 'Score' attribute

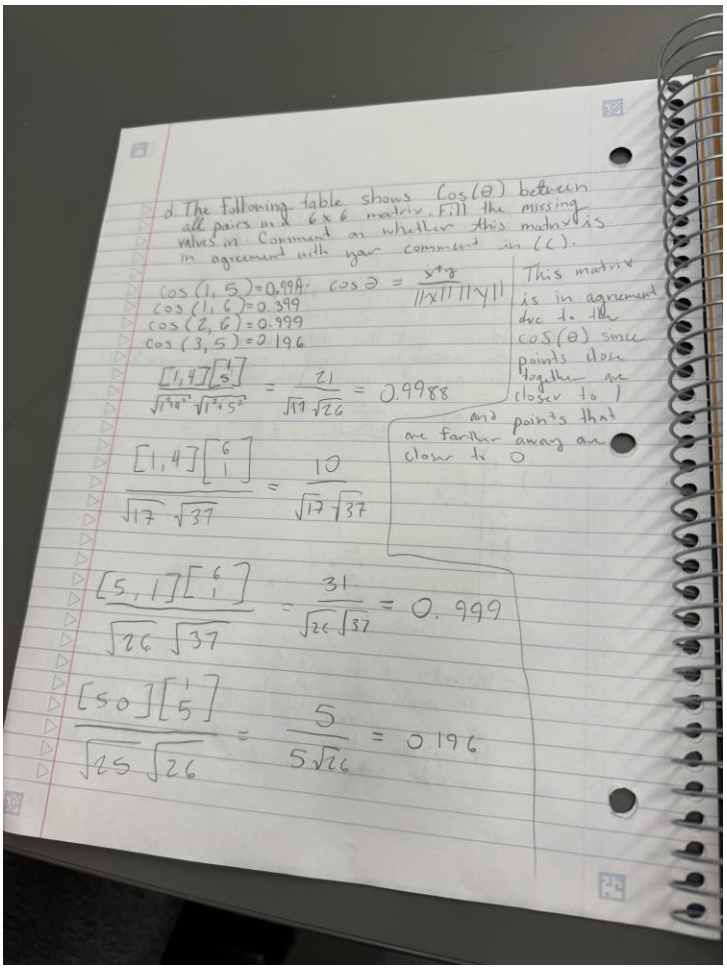
$$\text{mean} = \frac{\sum x_i}{n} = 3.\overline{3}$$

$$\text{variance} = \frac{\sum (x_i - \mu_x)^2}{n} = 4.\overline{2}$$

C.)



- Points 1, 4, and 5 are in close proximity with each other. Points 2, 3, and 6 are also in close proximity with each other. It appears there is a negative dependency with score 2 and score 2



d. The following table shows $\cos(\theta)$ between all pairs in a 6×6 matrix. Fill the missing values in. Comment on whether this matrix is in agreement with your comment in (c).

$\cos(1,5) = 0.998$ $\quad \cos\theta = \frac{x \cdot y}{\|x\| \|y\|}$

$\cos(1,6) = 0.399$

$\cos(2,6) = 0.999$

$\cos(3,5) = 0.196$

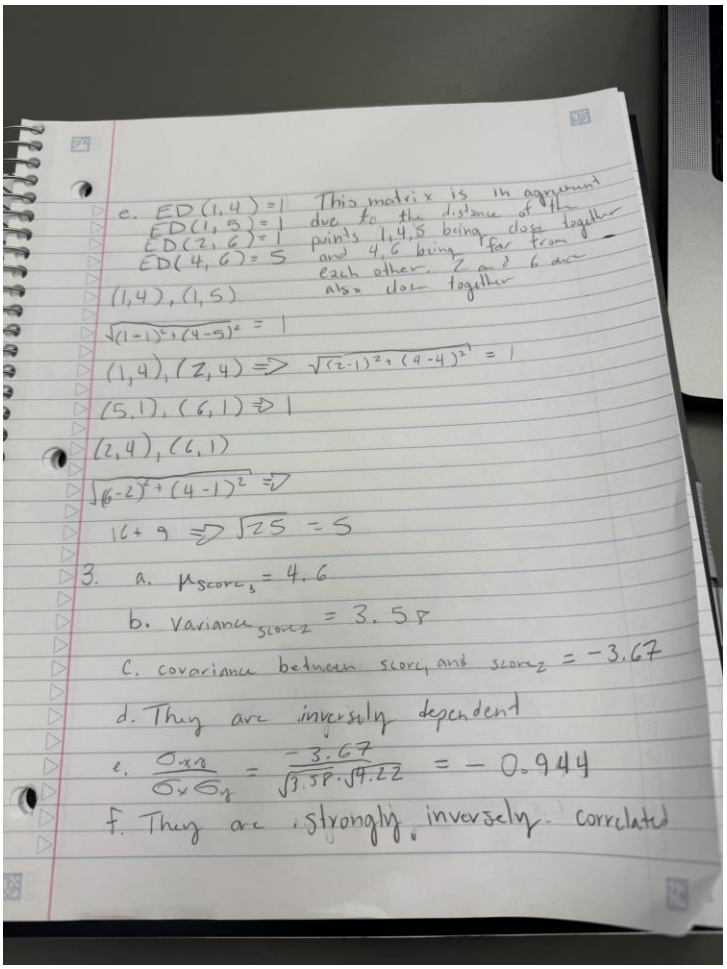
$$\frac{[1,4]\begin{bmatrix}1\\5\end{bmatrix}}{\sqrt{1^2+4^2}\sqrt{1^2+5^2}} = \frac{21}{\sqrt{17}\sqrt{26}} = 0.9988$$

$$\frac{[1,4]\begin{bmatrix}6\\1\end{bmatrix}}{\sqrt{17}\sqrt{37}} = \frac{10}{\sqrt{17}\sqrt{37}}$$

$$\frac{[5,1]\begin{bmatrix}6\\1\end{bmatrix}}{\sqrt{26}\sqrt{37}} = \frac{31}{\sqrt{26}\sqrt{37}} = 0.999$$

$$\frac{[50]\begin{bmatrix}1\\5\end{bmatrix}}{\sqrt{25}\sqrt{26}} = \frac{5}{5\sqrt{26}} = 0.196$$

This matrix is in agreement due to the $\cos(\theta)$ since points close together are closer to 1 and points that are farther away are closer to 0



e. $ED(1,4) = 1$
$ED(1,5) = 1$
$ED(2,6) = 1$
$ED(4,6) = 5$

This matrix is in agreement due to the distance of the points 1, 4, 5 being close together and 4, 6 being far from each other. 2 and 6 are also close together

$(1,4), (1,5)$

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

$(1,4), (2,4) \Rightarrow \sqrt{(2-1)^2 + (4-4)^2} = 1$

$(5,1), (6,1) \Rightarrow 1$

$(2,4), (6,1)$

$\sqrt{(6-2)^2 + (4-1)^2} \Rightarrow$

$16 + 9 \Rightarrow \sqrt{25} = 5$

3. a. $\mu_{score_3} = 4.6$

b. $Variance_{score_2} = 3.58$

c. covariance between $score_1$ and $score_2 = -3.67$

d. They are inversely dependent

e. $\frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-3.67}{\sqrt{3.58} \cdot \sqrt{4.22}} = -0.944$

f. They are strongly inversely correlated