

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Student's Name:

Mobile No:

Roll Number:

Branch:

1 a.

Table 1 Minimum and maximum attribute values before and after normalization

S. No.	Attribute	Before normalization		After normalization	
		Minimum	Maximum	Minimum	Maximum
1	pregs				
2	plas				
3	pres (in mm Hg)				
4	skin (in mm)				
5	test (in μ U/mL)				
6	BMI (in kg/m^2)				
7	pedi				
8	Age (in years)				

Inferences:

1. Infer upon the need for outlier correction
2. Justify the method employed for outlier correction.
3. Infer upon what happens before and after normalization
4. Inference 4(You may add or delete the number of inferences)

b.

Table 2 Mean and standard deviation before and after standardization

S. No.	Attribute	Before standardization		After standardization	
		Mean	Std. Deviation	Mean	Std. Deviation
1	pregs				
2	plas				
3	pres (in mm Hg)				
4	skin (in mm)				
5	test (in μ U/mL)				
6	BMI (in kg/m^2)				
7	pedi				
8	Age (in years)				

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Inferences:

1. Infer upon what happens before and after standardization
2. Inference 2(You may add or delete the number of inferences)

2 a.

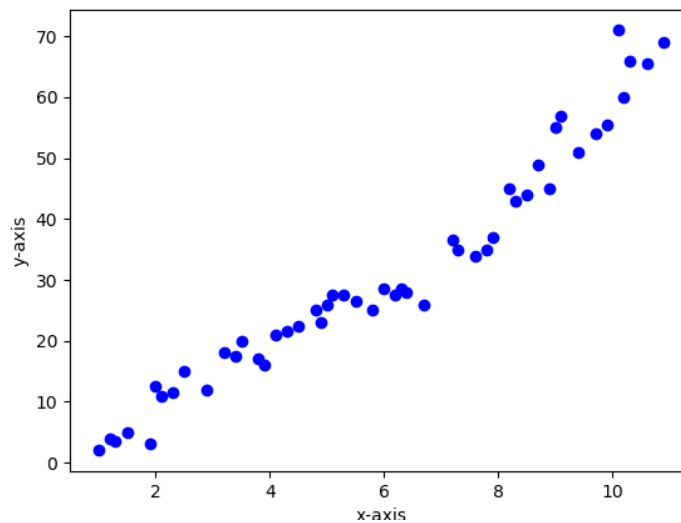


Figure 1 Scatter plot of 2D synthetic data of 1000 samples

Inferences:

1. Infer how attribute 1 is correlated to attribute 2 based upon the spread of the data points
2. Inference based on the density of points
3. Inference 3(You may add or delete the number of inferences)

Note: The scatter plot above is for illustration purposes. Replace it with the scatter plot obtained by you. Rename x-axis legend with x1 and y-axis legend with x2.

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

b.

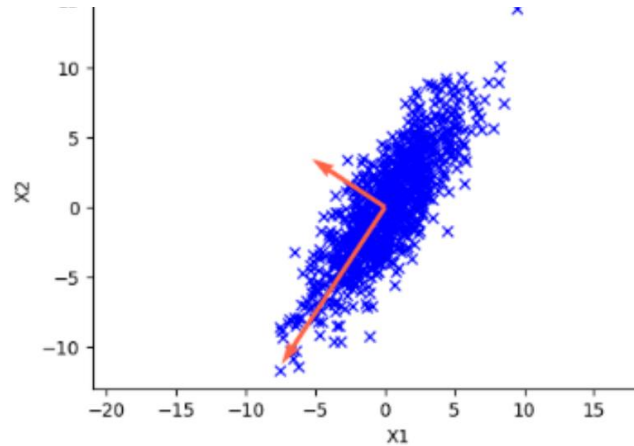


Figure 2 Plot of 2D synthetic data and Eigen directions

Inferences:

1. Infer the spread of data based upon the magnitude of Eigenvalues.
2. Inference based on the density of points near the intersection of Eigen axes and gradually away from it.
3. Inference 3 (You may add or delete the number of inferences)

Note: The scatter plot above is for illustration purposes. Replace it with the scatter plot obtained by you. Rename x-axis legend with x1 and y-axis legend with x2.

c.

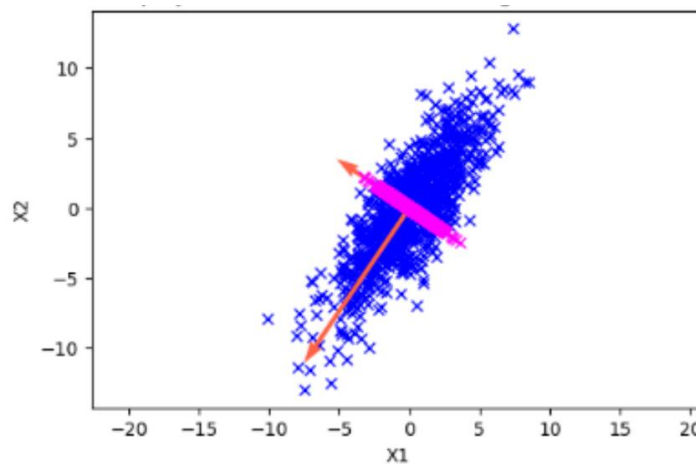


Figure 3 Projected Eigen directions onto the scatter plot with 1st Eigen direction highlighted

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

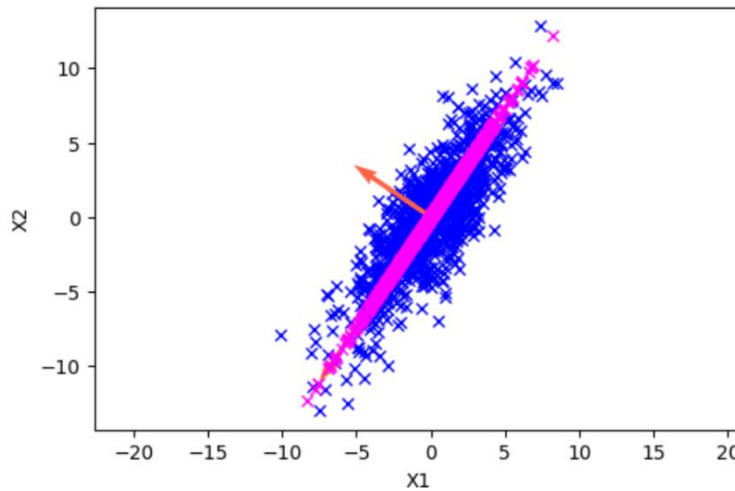


Figure 4 Projected Eigen directions onto the scatter plot with 2nd Eigen direction highlighted

Inferences:

1. Compare and contrast the magnitude of Eigenvalues
2. Infer variance of data along the Eigen axes from spread & density of points and relate it to the magnitude of Eigenvalues.
3. Inference 3(You may add or delete the number of inferences)

Note: The scatter plots above are for illustration purposes. Replace it with the scatter plot obtained by you. Rename x-axis legend with x1 and y-axis legend with x2.

d. Reconstruction error = (report only up to three decimal places)

Inferences:

1. Infer how the magnitude of reconstruction error affects the quality of reconstruction.
2. Inference 2(You may add or delete the number of inferences)

3 a.

Table 3 Variance and Eigenvalues of the projected data along the two directions

Direction	Variance	Eigenvalue
1		
2		

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Inferences:

1. Compare variance of the projected data along the two directions with the Eigenvalues of the two directions of projection
2. Inference 2(You may add or delete the number of inferences)

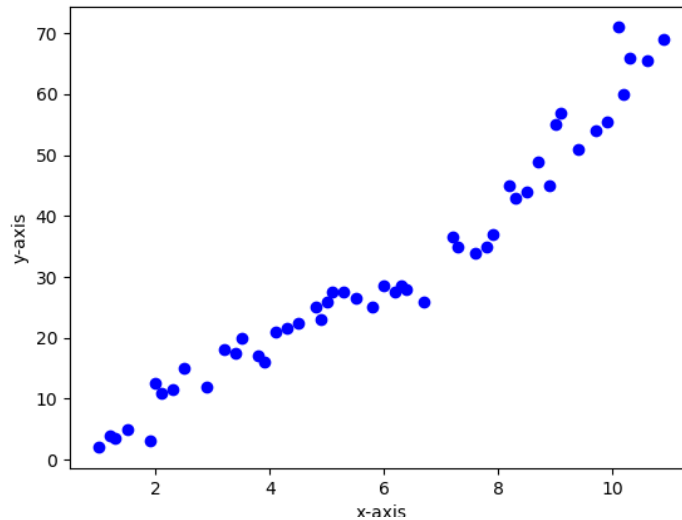


Figure 5 Plot of data after dimensionality reduction

Inferences:

1. Infer the correlation between the two attributes obtained after dimensionality reduction from the spread of data points
2. Inference 2(You may add or delete the number of inferences)

Note: The scatter plots above are for illustration purposes. Replace it with the scatter plot obtained by you. Rename x-axis legend with x1 and y-axis legend with x2.

b.



Figure 6 Plot of Eigenvalues in descending order

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Inferences:

1. Infer whether the subsequent Eigenvalues decrease gradually or rapidly
2. Identify the Eigenvalue from where the rate of decrease changes substantially
3. Inference 3 (You may add or delete the number of inferences)

Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. Rename x-axis legend with Eigenvalues and y-axis legend with magnitude.

c.

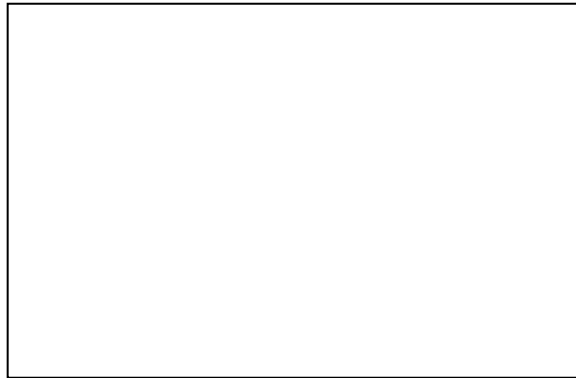


Figure 7 Line plot to demonstrate reconstruction error vs. components

Inferences:

1. Infer how the magnitude of reconstruction error affects the quality of reconstruction.
2. Inference 2 (You may add or delete the number of inferences)

Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. Rename x-axis legend with No. of components and y-axis legend with Reconstruction error.

Table 4 Covariance matrix for dimensionally reduced data (I=2)

	x1	x2
x1		
x2		

Table 5 Covariance matrix for dimensionally reduced data (I=3)

	x1	x2	x3
x1			
x2			
x3			

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Table 6 Covariance matrix for dimensionally reduced data (I=4)

	x1	x2	x3	x4
x1				
x2				
x3				
x4				

Table 7 Covariance matrix for dimensionally reduced data (I=5)

	x1	x2	x3	x4	x5
x1					
x2					
x3					
x4					
x5					

Table 8 Covariance matrix for dimensionally reduced data (I=6)

	x1	x2	x3	x4	x5	x6
x1						
x2						
x3						
x4						
x5						
x6						

Table 9 Covariance matrix for dimensionally reduced data (I=7)

	x1	x2	x3	x4	x5	x6	x7
x1							
x2							
x3							
x4							
x5							
x6							
x7							

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

Table 10 Covariance matrix for dimensionally reduced data (l=8)

	x1	x2	x3	x4	x5	x6	x7	x8
x1								
x2								
x3								
x4								
x5								
x6								
x7								
x8								

Inferences:

1. Observe off-diagonal elements and state the reason for the observed trend.
2. Observe the difference between diagonal and off-diagonal values and justify the reason for your observation.
3. Infer the trend in diagonal values.
4. Justify the reason for the increase/ decrease.
5. From the magnitude of diagonal elements, which component captures data variations the best?
6. From the value of diagonal elements, estimate the number of components that shall give the optimum reconstruction along with dimensionality reduction.
7. Observe the magnitude of the 1st diagonal element (topmost left corner) in each of the obtained covariance matrices. Is the magnitude the same or different? Ponder upon the underlying reason for your observation.
8. Observe the magnitude of the 2nd diagonal element in each of the obtained covariance matrices. Is the magnitude the same or different? Ponder upon the underlying reason for your observation.
9. Compare 3rd, 4th, 5th, 6th, and 7th diagonal elements across covariance matrices. Are they the same or different?
10. Inference 10(You may add or delete the number of inferences)

IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III

Attribute normalization, standardization and dimension reduction of data

d.

Table 11 Covariance matrix for original data

	pregs	plas	pres	skin	test	BMI	pedi	Age
pregs								
plas								
pres (in mm Hg)								
skin (in mm)								
test (in μ U/mL)								
BMI (in kg/m^2)								
pedi								
Age (in years)								

Inferences:

1. Observe the off-diagonal values and compare with the covariance matrix obtained after PCA $l=8$ reduction.
2. Similarly, compare the magnitudes of diagonal values.
3. Is there any trade of increase or decrease in diagonal elements like/ unlike covariance obtained after dimensionality reduction?
4. Inference 4(You may add or delete the number of inferences)

Guidelines for Report (Delete this while you submit the report):

- The plot/graph/figure/table should be centre justified with sequence number and caption.
- Inferences should be written as a numbered list.
- Use specific and technical terms to write inferences.
- Values observed/calculated should be rounded off to three decimal places.
- The quantities which have units should be written with units.