Name: Anooshka Bajaj

1 a.

Table 1 Minimum and Maximum Attribute Values Before and After Min-Max Normalization

S. No.	Attribute	Before Min-Max Normalization		After Min-Max Normalization	
		Minimum	Maximum	Minimum	Maximum
1	Temperature (in °C)	10.085110	31.375000	3.0	9.0
2	Humidity (in $g.m^{-3}$)	34.205670	99.720000	3.0	9.0
3	Pressure (in mb)	992.654583	1037.604386	3.0	9.0
4	Rain (in ml)	0.000000	2470.500000	3.0	9.0
5	Lightavgw/o0 (in lux)	0.000000	10565.352300	3.0	9.0
6	Lightmax (in lux)	2259.000000	54612.000000	3.0	9.0
7	Moisture (in %)	0.000000	100.000000	3.0	9.0

Inferences:

1. When the outliers are replaced by median of the respective attributes, then the number of outliers in each attribute in the replaced data is reduced.
2. Before normalization, the data is widely spread for every attribute. One attribute may overshadow the other. After normalization, the values of the attributes are changed to a common scale. The minimum values and the maximum values of each attribute is same.

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	Temperature (in °C)	21.369665	4.125407	0	1
2	Humidity (in $g.m^{-3}$)	83.992117	17.565823	0	1
3	Pressure (in mb)	1014.760524	6.121343	0	1
4	Rain (in ml)	168.400011	399.689066	0	1

5	Lightavgw/o0 (in lux)	2197.392401	2220.820133	0	1
6	Lightmax (in lux)	21788.623280	22064.993089	0	1
7	Moisture (in %)	32.386053	33.653245	0	1

Inferences:

1. Before standardization, the data is spread along wide ranges values for every attribute. One attribute may overshadow other. After standardization, each attribute has unit standard deviation and 0 mean which overcomes this problem.

2 a.

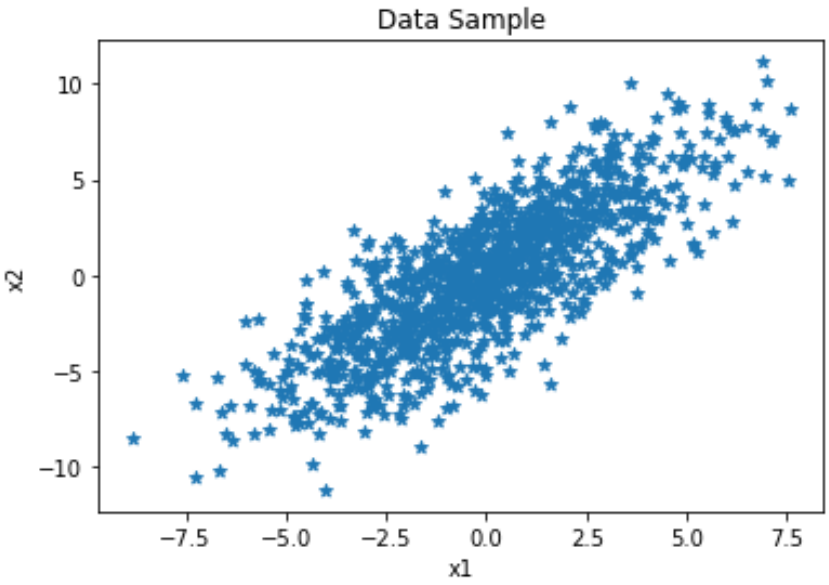


Figure 1 Scatter Plot of 2D Synthetic Data of 1000 samples

Inferences:

1. Attribute 1 and Attribute 2 have high positive correlation.
2. The density of points near origin is more. As we move outwards, density decreases.

b.

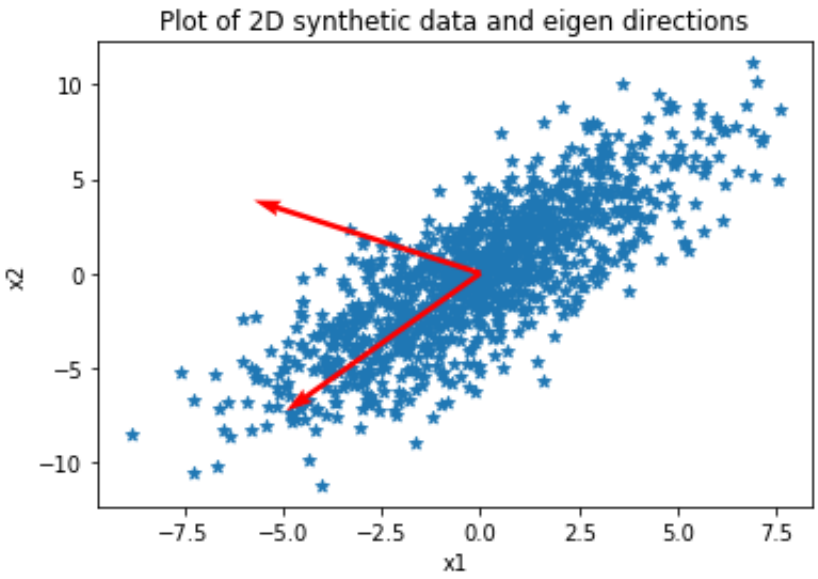


Figure 2 Plot of 2D Synthetic Data and Eigen Directions

Inferences:

1. The spread of the data is more across 2nd eigen vector than the 1st because the magnitude of eigen value 2 is greater than eigen value 1. (18.16910025 and 1.69971065)
2. The density of points at the intersection is more and as we move outwards, density decreases.

c.

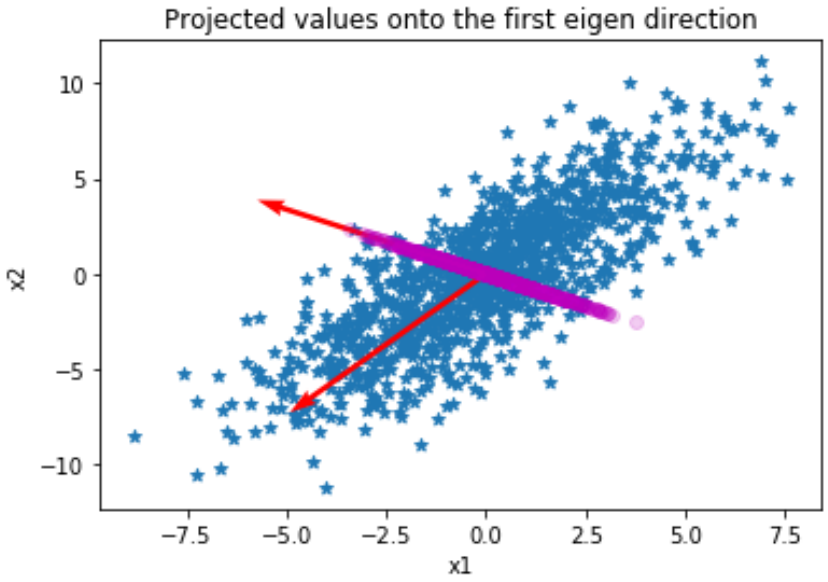


Figure 3 Projected Eigen Directions onto the Scatter Plot with 1st Eigen Direction highlighted

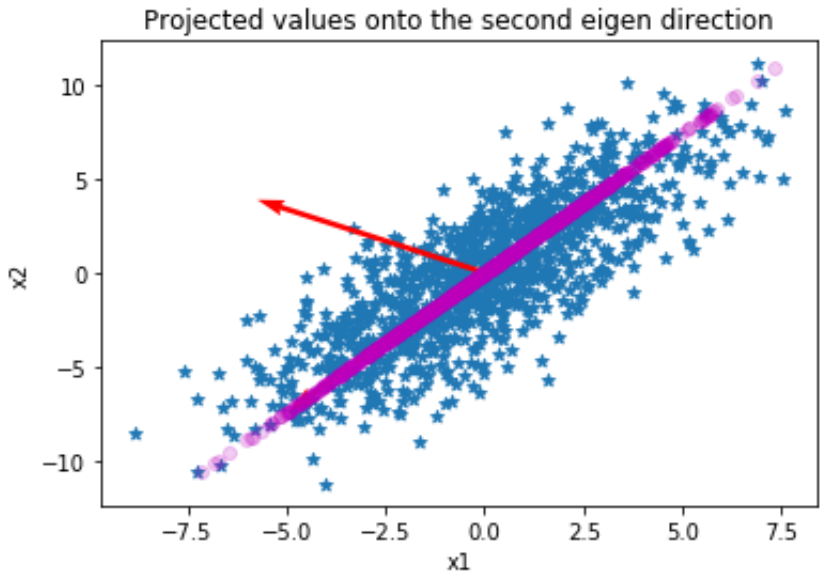


Figure 4 Projected Eigen Directions onto the Scatter Plot with 2nd Eigen Direction highlighted

Inferences:

1. Eigen value is directly proportional to the variance of the data. In this case, eigen value 2 is greater than eigen value 1.
2. Larger eigen value means that the it contains more data information. the

d. Reconstruction Error = 6.02

Inferences:

1. The value of reconstruction error is low. This means that the original data is reconstructed properly i.e. the data is lossless.

3 a.

Table 3 Variance and Eigen Values of the projected data along the two directions

Direction	Variance	Eigen Value
1	2.19996801	2.20229848
2	1.41932231	1.42082583

Inferences:

1. More the variance i.e. spread of the data, greater is the eigen value.

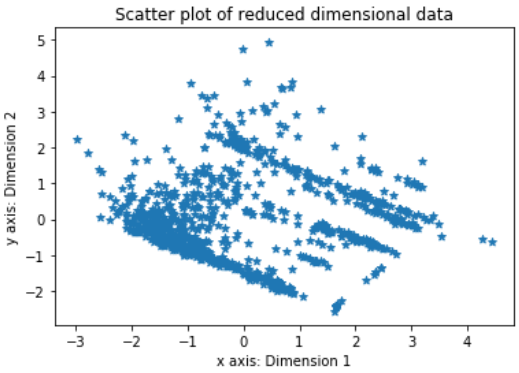
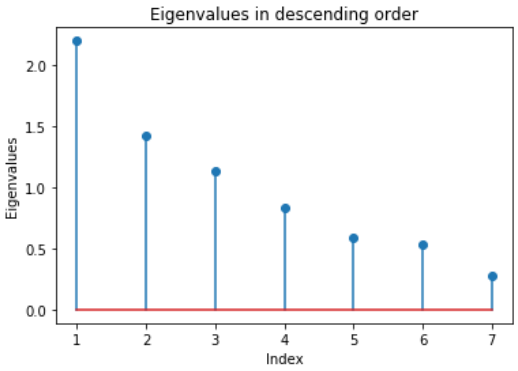


Figure 5 Plot of Landslide Data after dimensionality reduction

Inferences:

1. The density around the median of reduced data is very high and reduces as we move away. The reduced data follows a skewed Gaussian distribution. The variance of each attribute of the reduced data is the eigen value corresponding to it.

b.



c. **Figure 6 Plot of Eigen Values in descending order**

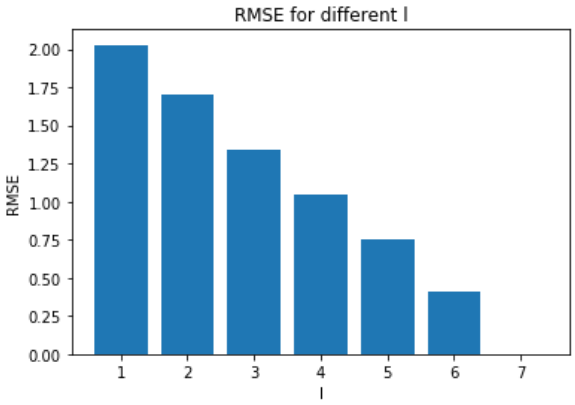


Figure 7 Line Plot to demonstrate Reconstruction Error vs. Components

Inferences:

1. The value of error is inversely proportional to the quality of reconstruction.