

# PRINCIPAL COMPONENT ANALYSIS



SHIVANI BHATIA

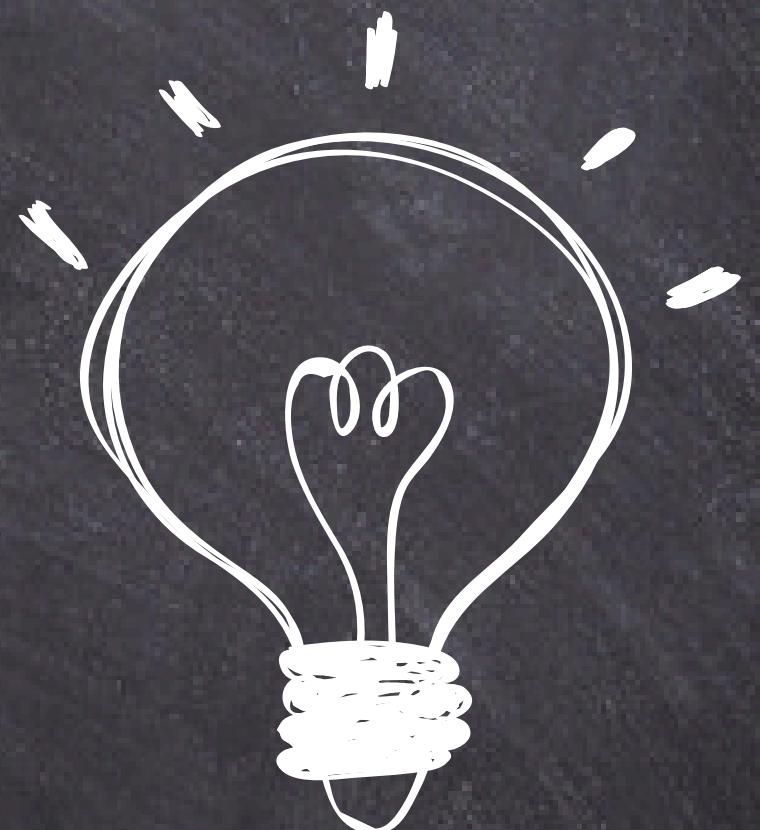
AUGUST 2024

# WHAT IS PCA?

→ PCA IS FUNDAMENTALLY A  
DIMENSIONALITY REDUCTION  
TECHNIQUE

Dimensionality reduction is a way of transforming a dataset with a high number of features into a smaller dataset, dropping unnecessary features.

→ PCA IS A STATISTICAL PROCEDURE  
TO CONVERT OBSERVATIONS OF  
POSSIBLY CORRELATED VARIABLES  
TO 'PRINCIPAL COMPONENTS'



# WHY PCA?

DIMENSIONALITY  
REDUCTION

DATA  
VISUALIZATION

HANDLE  
MULTICOLLINEARITY

THE PREDICTIVE MODEL SETUP: HAVING A LOT OF CORRELATED FEATURES LEADS TO THE MULTICOLLINEARITY PROBLEM. ITERATIVELY REMOVING FEATURES IS TIME-CONSUMING AND ALSO LEADS TO SOME INFORMATION LOSS.

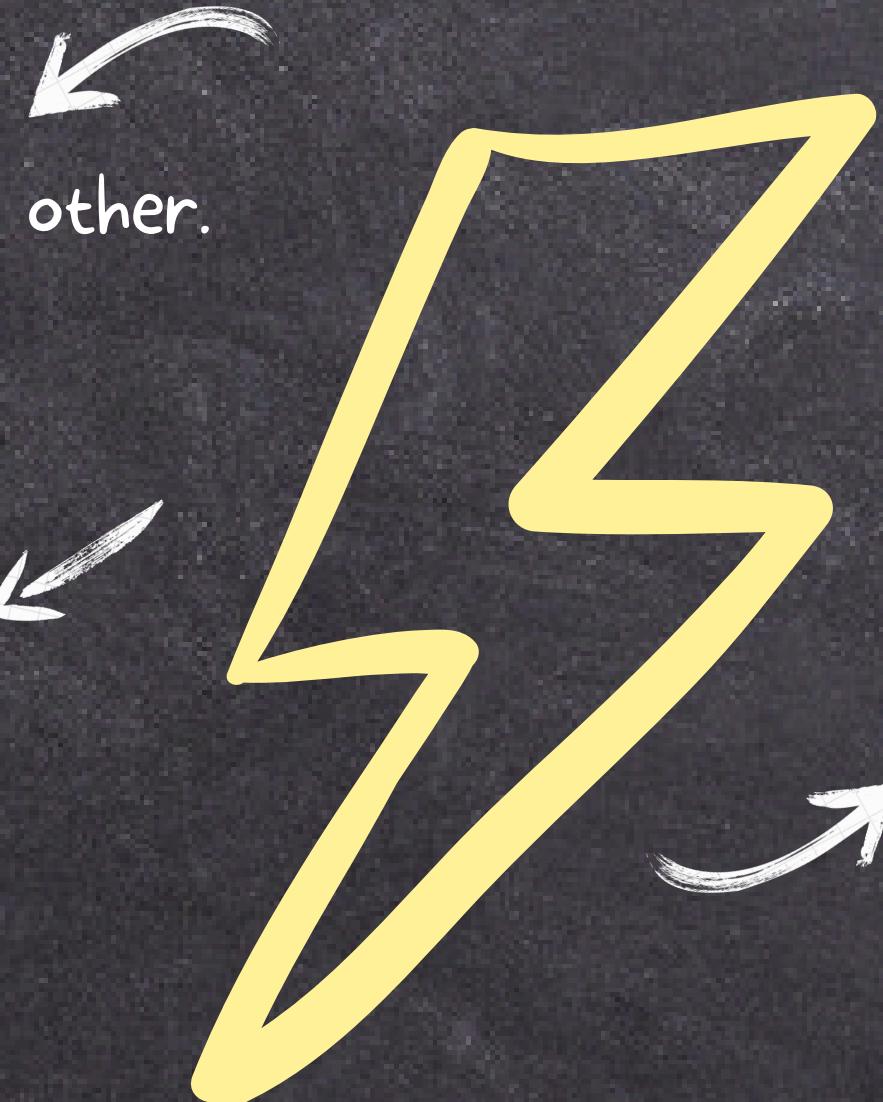
DATA VISUALISATION: IT IS NOT POSSIBLE TO VISUALISE MORE THAN TWO VARIABLES AT THE SAME TIME USING ANY 2-D PLOT. THEREFORE, FINDING RELATIONSHIPS BETWEEN THE OBSERVATIONS IN A DATA SET HAVING SEVERAL VARIABLES THROUGH VISUALISATION IS QUITE DIFFICULT.

1

- They are uncorrelated with each other.

2

- They are linear combinations of the original variables.

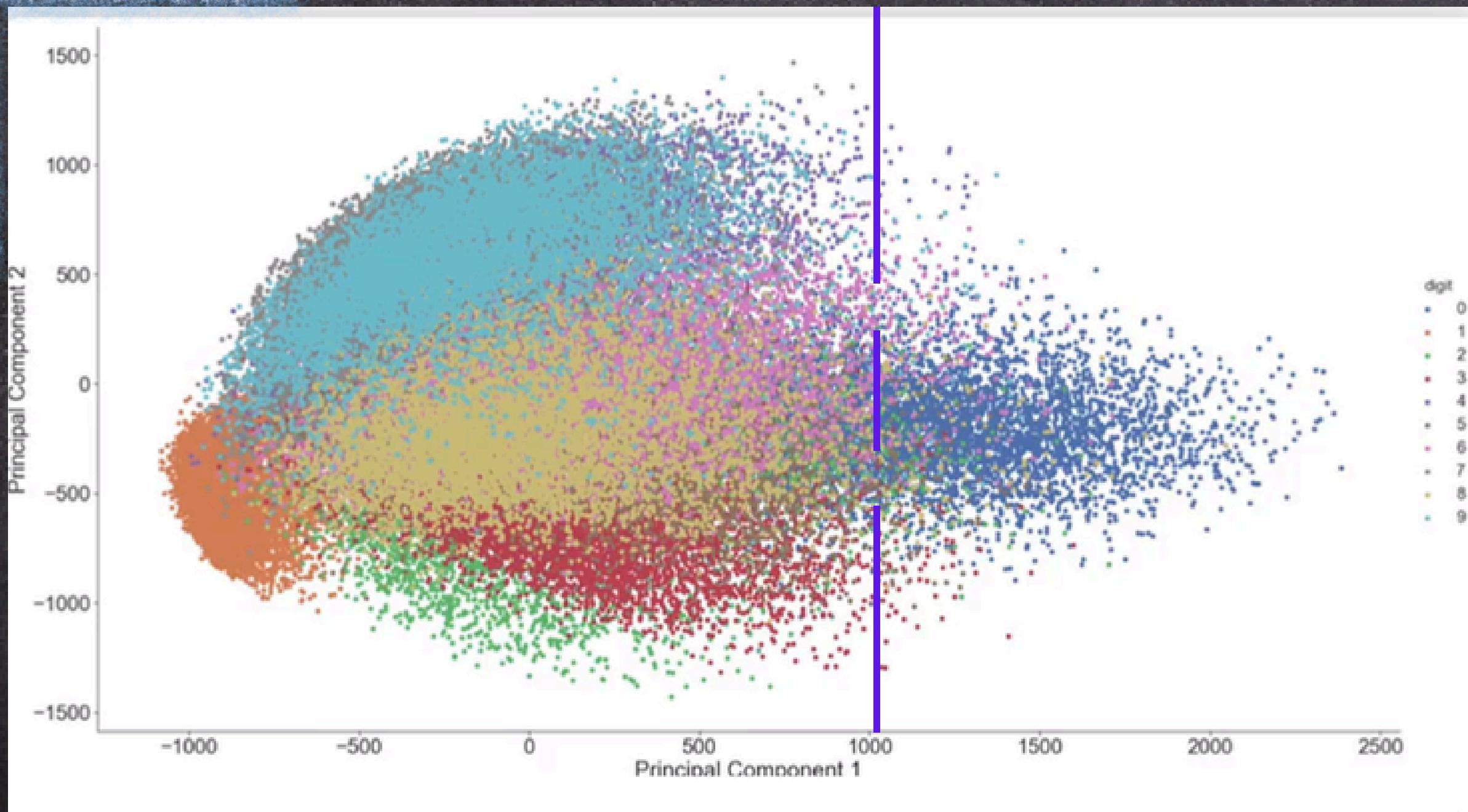


3

- They help in capturing maximum information in the data set.

# PRINCIPAL COMPONENTS

# PRINCIPAL COMPONENT SEPARATOR FOR BLUE POINTS



If we draw a line parallel to Principal Component 2, we can achieve the objective. Hence, we draw the separator at Principal Component 1 = 1000 or  $x = 1000$ .

# CLOSING NOTES



- PCA HELPS IN SOLVING THE MULTICOLLINEARITY PROBLEM BY CREATING NEW UNCORRELATED FEATURES THAT ARE USED AS INPUT FOR THE PREDICTIVE MODEL.
- WITH A DIMENSIONALITY REDUCTION TECHNIQUE LIKE PCA, WE CONVERT A DATASET HAVING N DIMENSIONS TO ANOTHER DATASET HAVING K DIMENSIONS WHERE  $N > K$ .
- PCA IS A TYPE OF UNSUPERVISED LEARNING