AN N-DIMENSIONAL SPACE
IS REPRESENTED BY

AN N-DIMENSIONAL HYPERCUBE

THIS IS A FANCY WORD,
BUT DON'T BE INTIMIDATED -
IT JUST MEANS EACH POINT
IN THIS HYPERCUBE
IS A LIST OF N VALUES

ANYTHING - ANYTHING - CAN BE REPRESENTED AS A POINT IN A HYPERCUBE

BASICALLY - ANY INSTANCE CAN BE REPRESENTED USING A FEATURE VECTOR - A LIST OF NUMBERS THAT DESCRIBE THAT INSTANCE

WE CAN THEN DO ALL KINDS OF COOL THINGS TO THAT INSTANCE - FINDING OTHER VECTORS THAT ARE "LIKE" THIS ONE, FINDING ITS DISTANCE FROM OTHER INSTANCES, AND SO ON

NOW, THE THING IS, THAT AS AN INSTANCE GETS COMPLICATED, ITS FEATURE VECTOR STARTS TO GET REALLY LONG

IN THE EXAMPLE ABOVE, WE WERE SUGGESTING A FEATURE VECTOR FOR AN EMAIL WHERE EVERY POSSIBLE WORD WAS REPRESENTED WITH A 1 OR A 0

THIS FEATURE VECTOR WOULD BE INFINITELY LONG, AND IMPOSSIBLE TO DO ANYTHING WITH!

THIS GETS TO THE HEART OF
SOMETHING CALLED

THE CURSE OF DIMENSIONALITY

ON THE ONE HAND

ANY RICH REPRESENTATION OF A COMPLEX INSTANCE REQUIRES A LOT OF FEATURES

ON THE OTHER HAND

WE ARE NOT SET UP TO EITHER VISUALIZE OR EFFICIENTLY PROCESS DATA OF VERY HIGH DIMENSIONALITY

THE SOLUTION?

DIMENSIONALITY REDUCTION TECHNIQUES

WHICH EFFECTIVELY REDUCE THE NUMBER OF DIMENSIONS THAT WE NEED TO EXPRESS OUR DATA IN

FEATURE EXTRACTION TECHNIQUES

PERFORM DIMENSIONALITY REDUCTION
BY RE-EXPRESSING THE DATA
IN A LOWER DIMENSIONALITY FORM

THE MOST FAMOUS FEATURE EXTRACTION
TECHNIQUE IS

PRINCIPAL COMPONENTS ANALYSIS (PCA)

GIVEN A LARGE NUMBER OF CORRELATED
TIME SERIES, PCA WILL FIND 2-3
UNDERLYING CAUSES THAT EXPLAIN
MOST OF THE MOVEMENTS

SAY YOU CONDUCTED A SURVEY FOR
A MARKET STUDY

WHAT DO PEOPLE LOOK FOR WHEN THEY BUY A CAR?

YOU ASKED 10000 PEOPLE
TO FILL OUT A FORM

THE FORM HAS 50 QUESTIONS

WHILE ALL THESE QUESTIONS MIGHT
BE RELEVANT, IN REALITY THERE MIGHT BE
JUST 2 OR 3 THINGS THAT TRIGGER THE
PURCHASE

YOU HAVE 10000 PROBLEM INSTANCES
AND **50** DIMENSIONS

THAT'S A LOT OF COMPLEX DATA!
AND IT IS VERY DIFFICULT TO VISUALIZE
OR MAKE SENSE OF

PRINCIPAL COMPONENTS ANALYSIS TO THE RESCUE!

PRINCIPAL COMPONENTS ARE DIRECTIONS IN WHICH THE DATA IS MOST SPREAD OUT

AN ORTHOGONAL TRANSFORMATION IS LIKE A CHANGE IN PERSPECTIVE

(IMAGINE YOU ARE ROTATING THE DATA OR LOOKING AT IT'S REFLECTION ALONG SOME AXIS – OR BOTH)

WE FIND THE PRINCIPAL COMPONENTS BY PERFORMING AN **ORTHOGONAL TRANSFORMATION**

THE NUMBER OF PRINCIPAL COMPONENTS CAN BE LESS THAN OR EQUAL TO THE NUMBER OF ORIGINAL DIMENSIONS

THE TRANSFORMATION IS DONE SUCH THAT FIRST PRINCIPAL COMPONENT HAS THE LARGEST POSSIBLE VARIANCE

IE ACCOUNTS FOR AS MUCH OF THE VARIABILITY IN THE DATA AS POSSIBLE

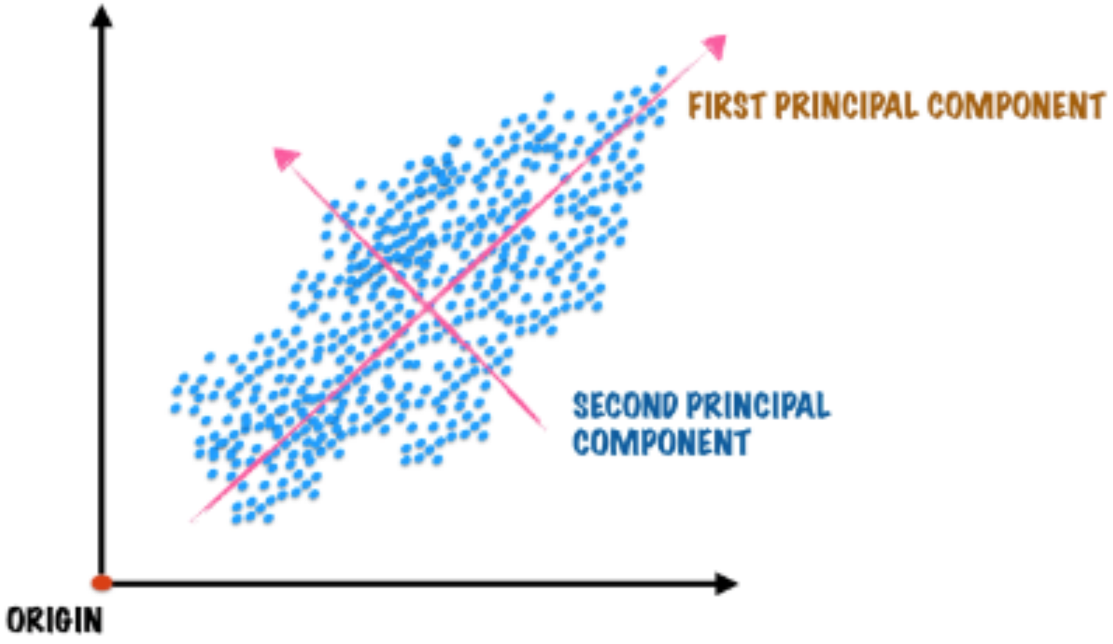
ONCE WE TRANSFORM THE DATA, ITS EXPRESSED IN TERMS OF THE PRINCIPAL COMPONENTS RATHER THAN THE ORIGINAL VARIABLES

THE SECOND PRINCIPAL COMPONENT ACCOUNTS FOR THE NEXT HIGHEST VARIANCE, WHILE REMAINING ORTHOGONAL TO THE FIRST **AND SO ON**

LET'S SEE ANOTHER EXAMPLE

IN GENERAL, THERE CAN BE AS MANY PRINCIPAL COMPONENTS AS THE ORIGINAL NUMBER OF DIMENSIONS IN THE DATA

HOW DOES KNOWING THE PRINCIPAL COMPONENTS HELP IN DIMENSIONALITY REDUCTION



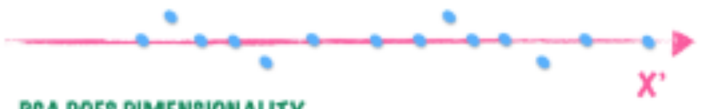
IMAGINE WE HAVE SOME DATA LIKE THIS IN A 2D SPACE

DIMENSIONALITY REDUCTION FROM 2D TO 1D

THE Y DIMENSION IS USELESS FOR THIS DATA. IT CAN ALL BE EXPRESSED USING ONLY ONE DIMENSION

BUT WHAT IF THERE WAS SOME VARIATION IN THE Y DIRECTION?

THE NEW AXES ARE NOW THE PRINICIPAL COMPONENTS IE. THE CO-ORDINATES ARE EXPRESSED IN TERMS OF THE PRINCIPAL COMPONENTS

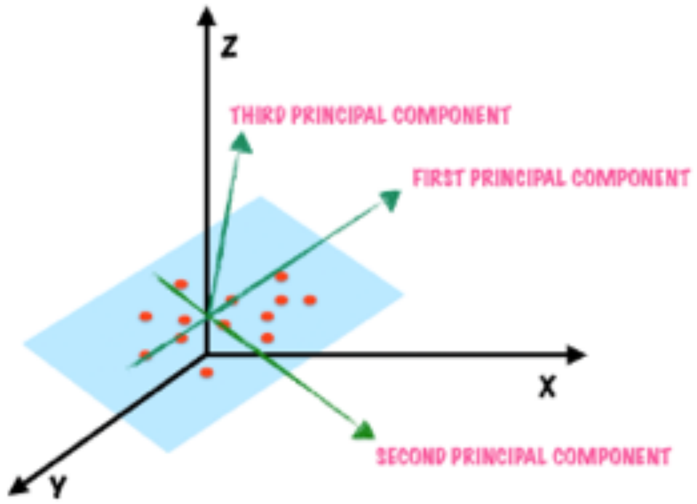


ONCE PCA FINDS THE PRINCIPAL COMPONENTS, IT DOES AN

PCA DOES DIMENSIONALITY REDUCTION BY USING ONLY THE

IT IGNORES THE PRINCIPAL

LET'S SEE ANOTHER EXAMPLE - THIS TIME IN 3D SPACE


Z
THIRD PRINCIPAL COMPONENT
FIRST PRINCIPAL COMPONENT
X
SECOND PRINCIPAL COMPONENT
Y

WE FIND THE PRINCIPAL COMPONENTS BY PERFORMING AN **ORTHOGONAL TRANSFORMATION**

THE NUMBER OF PRINCIPAL COMPONENTS CAN BE LESS THAN OR EQUAL TO THE NUMBER OF ORIGINAL DIMENSIONS

THE TRANSFORMATION IS DONE SUCH THAT FIRST PRINCIPAL COMPONENT HAS THE LARGEST POSSIBLE VARIANCE

IE ACCOUNTS FOR AS MUCH OF THE VARIABILITY IN THE DATA AS POSSIBLE

ONCE WE TRANSFORM THE DATA, ITS EXPRESSED IN TERMS OF THE PRINCIPAL COMPONENTS RATHER THAN THE ORIGINAL VARIABLES

THE SECOND PRINCIPAL COMPONENT ACCOUNTS FOR THE NEXT HIGHEST VARIANCE, WHILE REMAINING ORTHOGONAL TO THE FIRST **AND SO ON**

PCA TAKES A LARGE NUMBER OF CORRELATED VARIABLES AND TRANSFORMS THEM INTO UNCORRELATED VARIABLES EXPRESSED IN THE ORDER OF IMPORTANCE

THEIR VARIANCE IS INSIGNIFICANT - THESE MAY WELL BE NOISE

THE LESS IMPORTANT VARIABLES

THERE ARE STANDARD ALGORITHMS TO PERFORM PCA, ONE POPULAR METHOD INVOLVES SOMETHING KNOWN AS **SINGULAR VALUE DECOMPOSITION (SVD)**

WE WON'T GO INTO THE DETAILS, BUT SVD CAN BE IMPLEMENTED IN MANY PROGRAMMING LANGUAGES WITH 1 LINE OF CODE