# UNIT 3

# DATA PRE-PROCESSING & Dimensionality Reduction

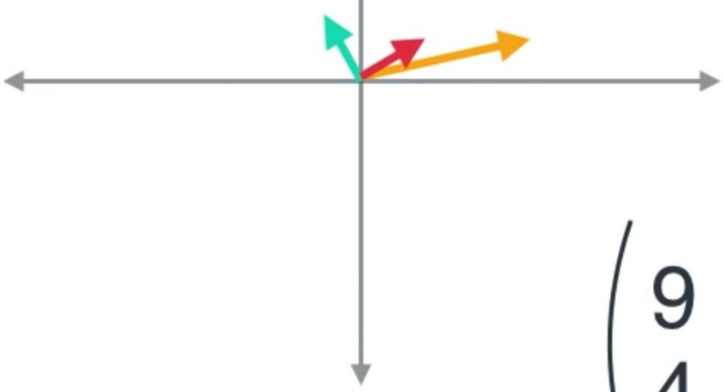
## METHOD 1: PCA (Principal Component Analysis)

# Concept of Eigen Values and Eigen Vectors -

## (Prelude to PCA and LDA )

# Eigenstuff

Eigen Vectors remain in the same direction as the original feature vector space, even after linear transformation. The Eigen Value changes in magnitude(“stretches”),direction is unchanged. **Used** as a “reference point” to know where the original points are after transformation - here - dimensionality reduction.

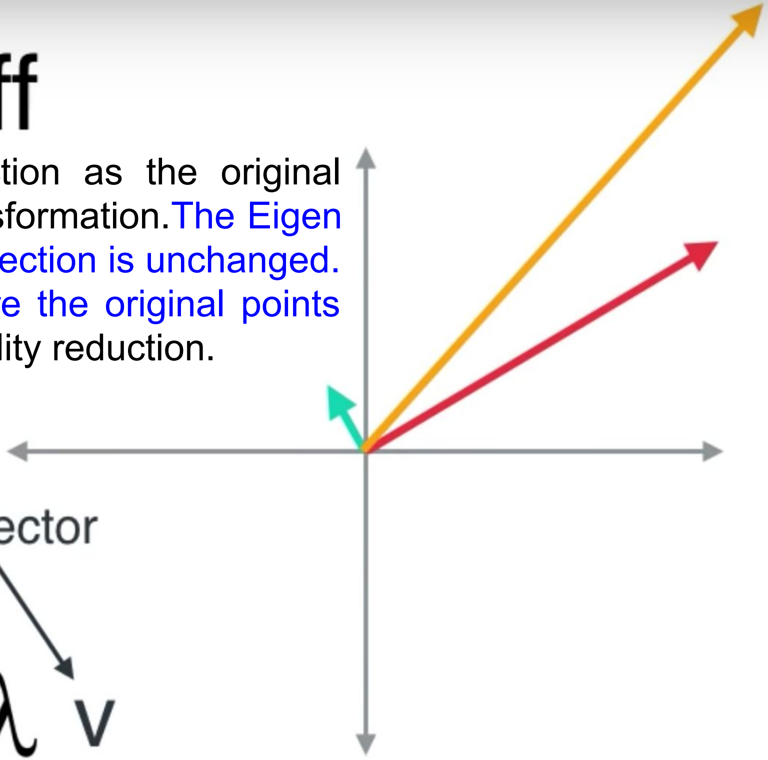


Original Feature Vector Space.

Eigenvector

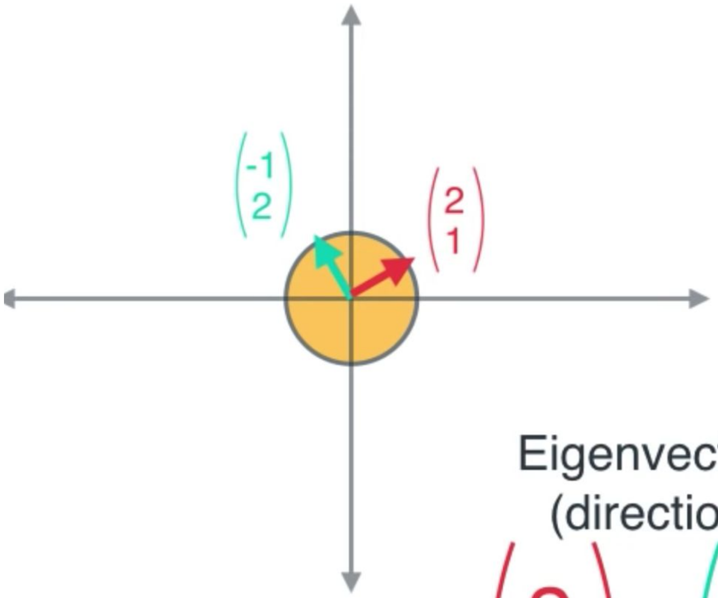
$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} v = \lambda v$$

Eigenvalue



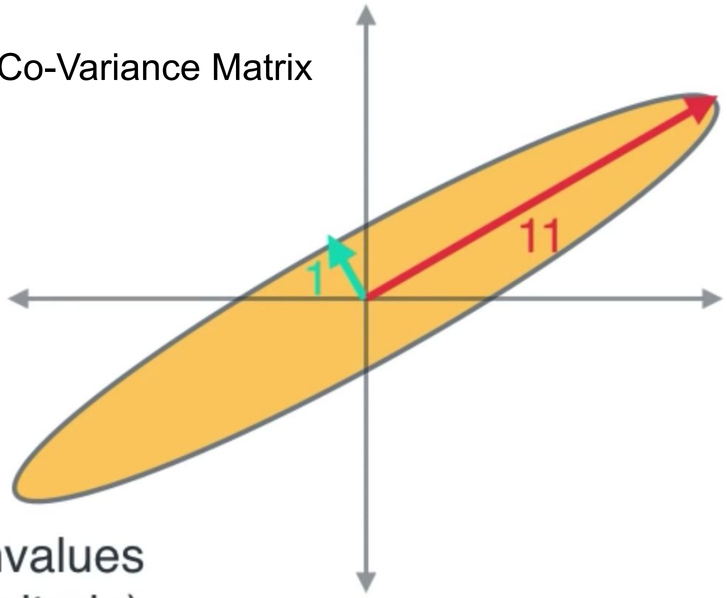
After Linear Transformation. “Yellow Vector”, representing every data point other than Eigenvectors changes direction and magnitude.

# Linear Transformations




$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} \longrightarrow$$

Co-Variance Matrix



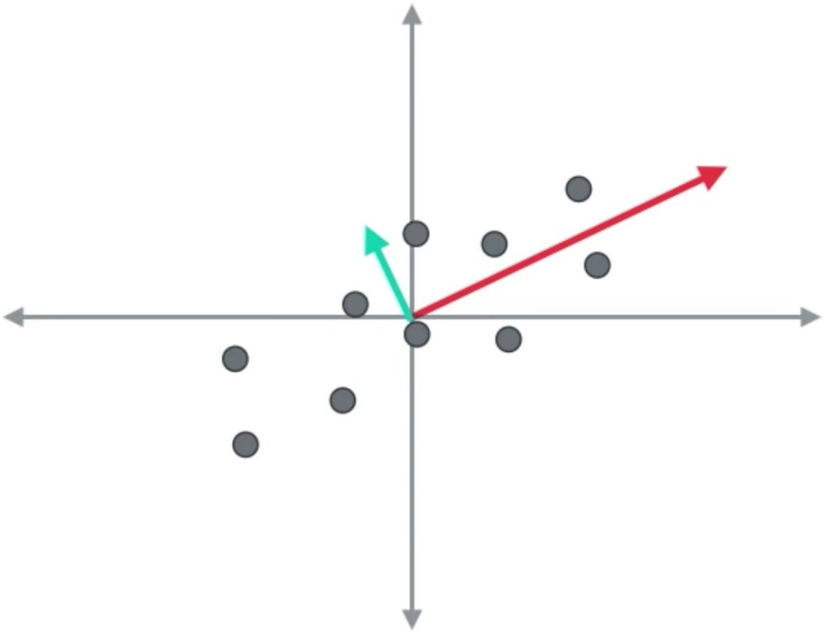

Eigenvectors
(direction)

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Eigenvalues
(magnitude)

11 1

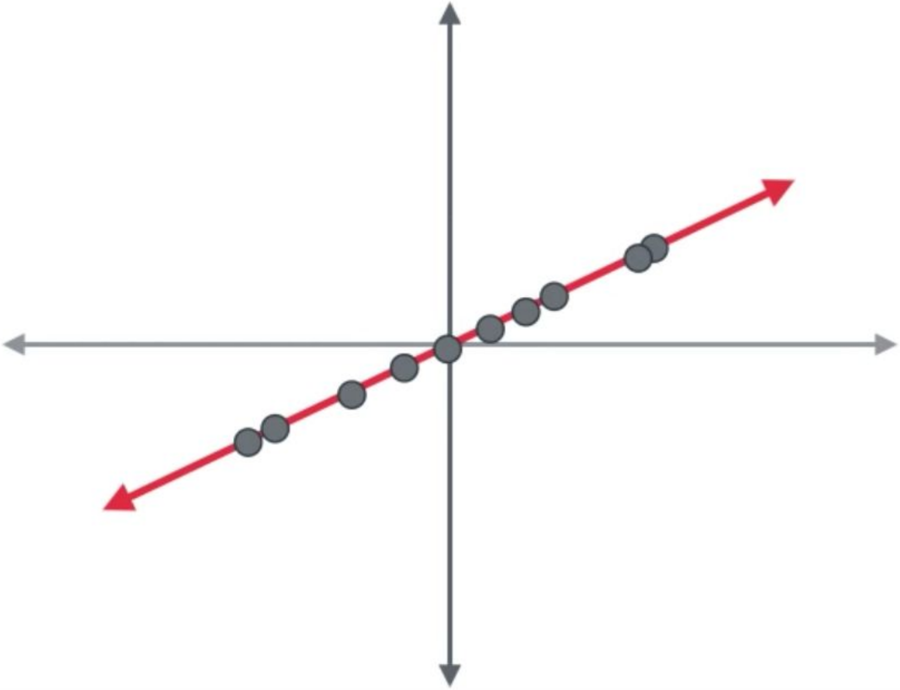
# Principal Component Analysis (PCA)



$$\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

| | | |
|---|---|---|
| $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ | $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ | Eigenvectors (direction) |
| 11 | 1 | Eigenvalues (magnitude) |

# Principal Component Analysis (PCA)



$$\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

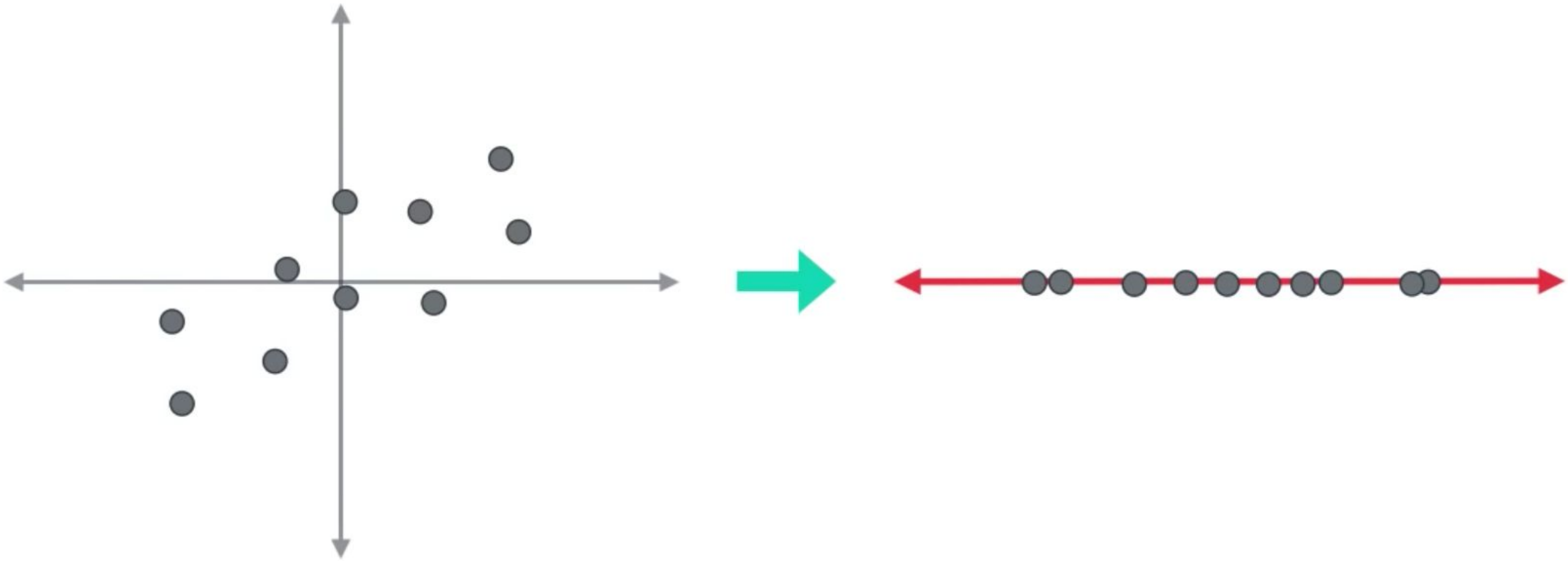
$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Eigenvectors (direction)

$11$

Eigenvalues (magnitude)

# Principal Component Analysis (PCA)



# Practical Uses OF
# PCA (Principal Component Analysis)

## E.g. Eigen Faces - Dimensionality Reduction and Image Reconstruction

# Eigenfaces example

Training images
$x_1, \ldots, x_M$

n x d = 9000 x 18,000
9000 samples
18,000 features/dimensions

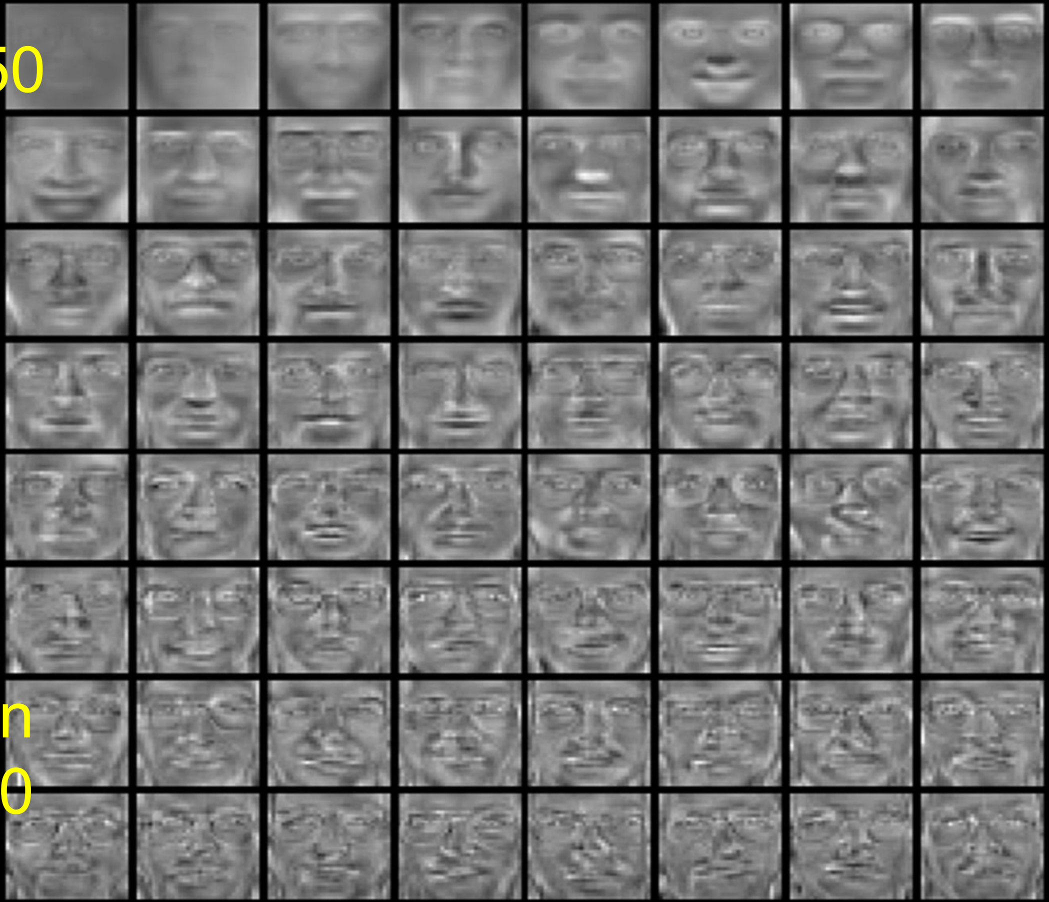
n x d is reduced to m x k where k <<< d.

Here it is m x k is 8 x 150

n x d = 9000 x 18,000

Mean: μ

Top eigenvectors: $u_1,\ldots u_k$



Dimensionality reduction from 18,000 to 150 using PCA

# Eigenfaces: Key idea *(Turk and Pentland, 1991)*

- Assume that most face images lie on a low-dimensional subspace determined by the first $k$ ($k <<< d$) directions of maximum variance
- Use PCA to determine the vectors or "eigenfaces" $u_1, u_2, \ldots u_k$ that span that subspace
- Represent all face images in the dataset as linear combinations of eigenfaces. Find the coefficients by dot product.

$$\mathbf{x} \rightarrow [\mathbf{u}_1^T(\mathbf{x}-\mu), \ldots, \mathbf{u}_k^T(\mathbf{x}-\mu)]$$

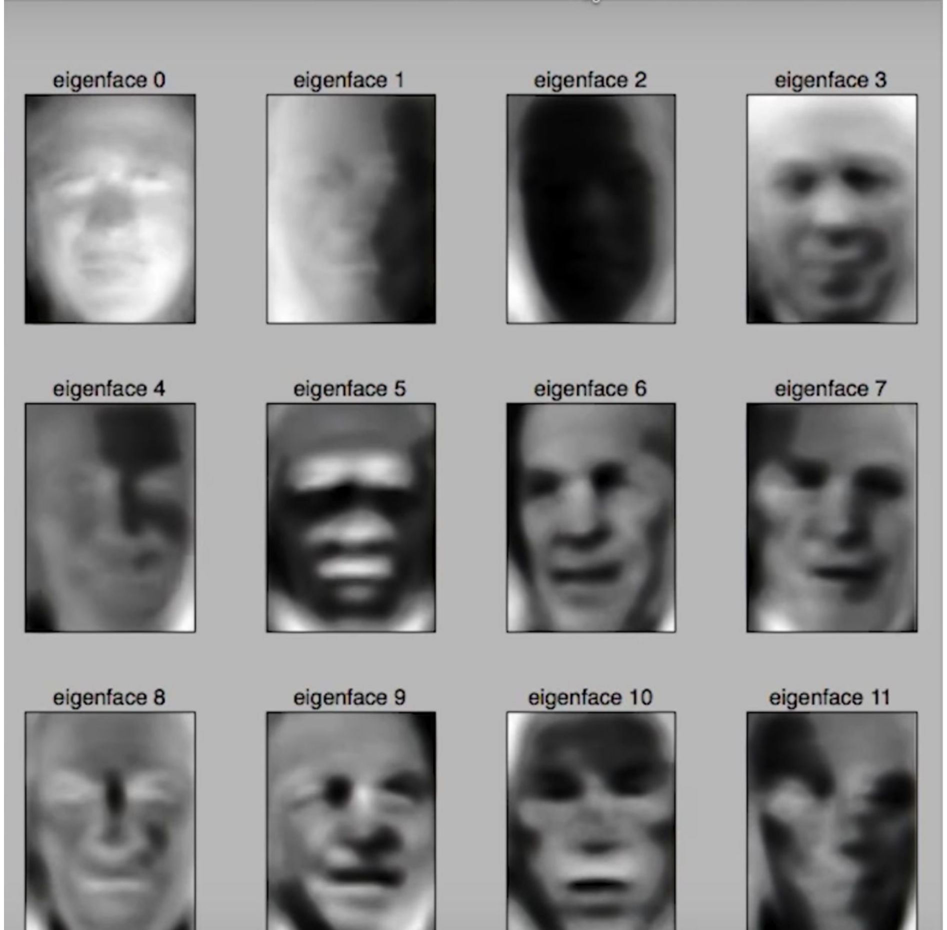
$$= [w_1, \ldots w_k]$$

*This vector is the representation of the face.*

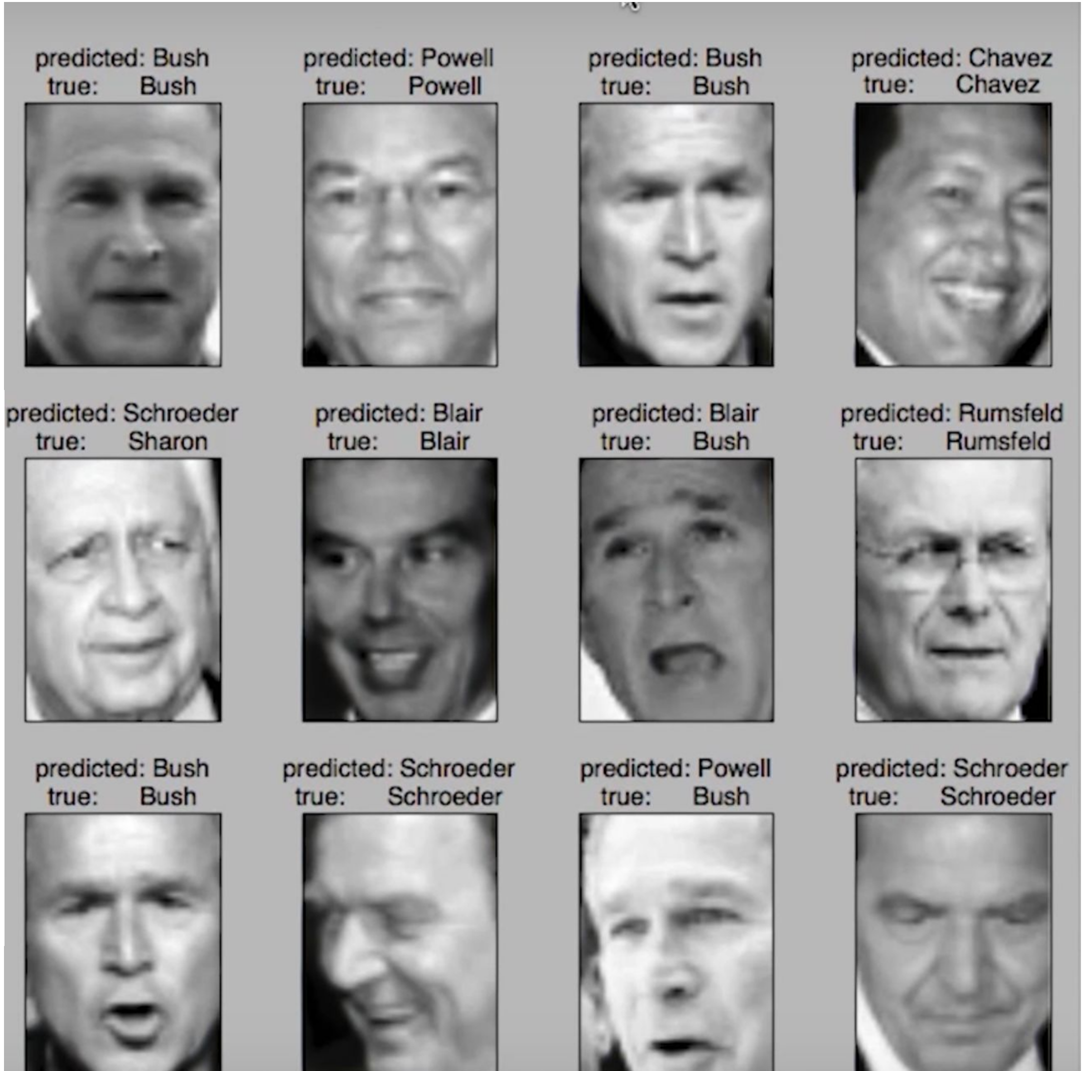
- Reconstruction:

$$\hat{\mathrm{x}} = \mu + w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + w_3\mathbf{u}_3 + w_4\mathbf{u}_4 + \ldots$$

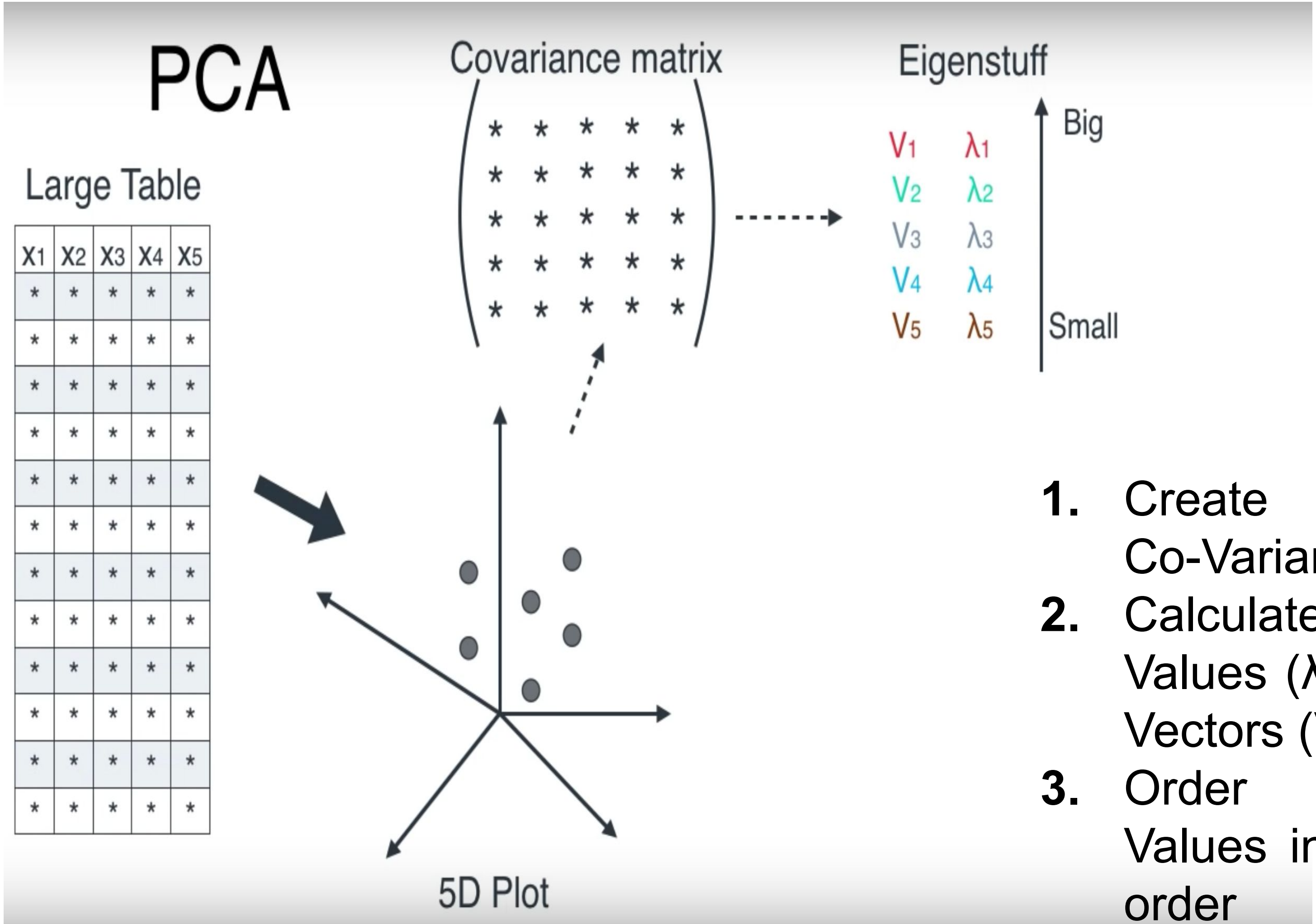
# Eigenfaces: Key idea *(Turk and Pentland, 1991)*



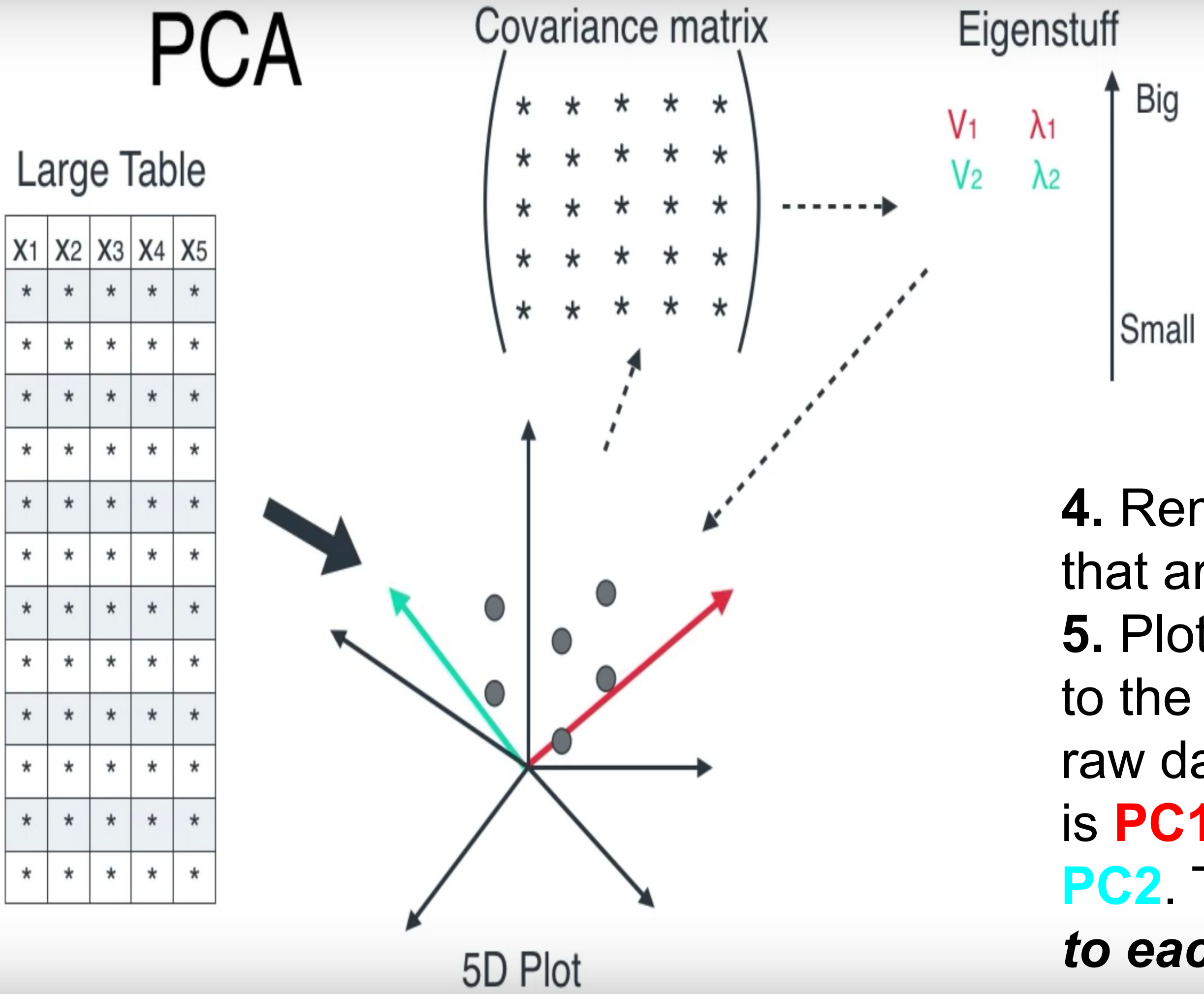
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Ariel Sharon | 0.69 | 0.60 | 0.64 | 15 |
| Colin Powell | 0.80 | 0.84 | 0.82 | 61 |
| Donald Rumsfeld | 0.86 | 0.80 | 0.83 | 30 |
| George W Bush | 0.85 | 0.88 | 0.86 | 138 |
| Gerhard Schroeder | 0.86 | 0.79 | 0.83 | 24 |
| Hugo Chavez | 0.93 | 0.64 | 0.76 | 22 |
| Tony Blair | 0.68 | 0.78 | 0.72 | 32 |
| avg / total | 0.82 | 0.82 | 0.82 | 322 |



# Summary OF
# PCA (Principal Component Analysis)




1. Create the Co-Variance Matrix
2. Calculate the Eigen Values (λ) )and Eigen Vectors (V)
3. Order the Eigen Values in descending order



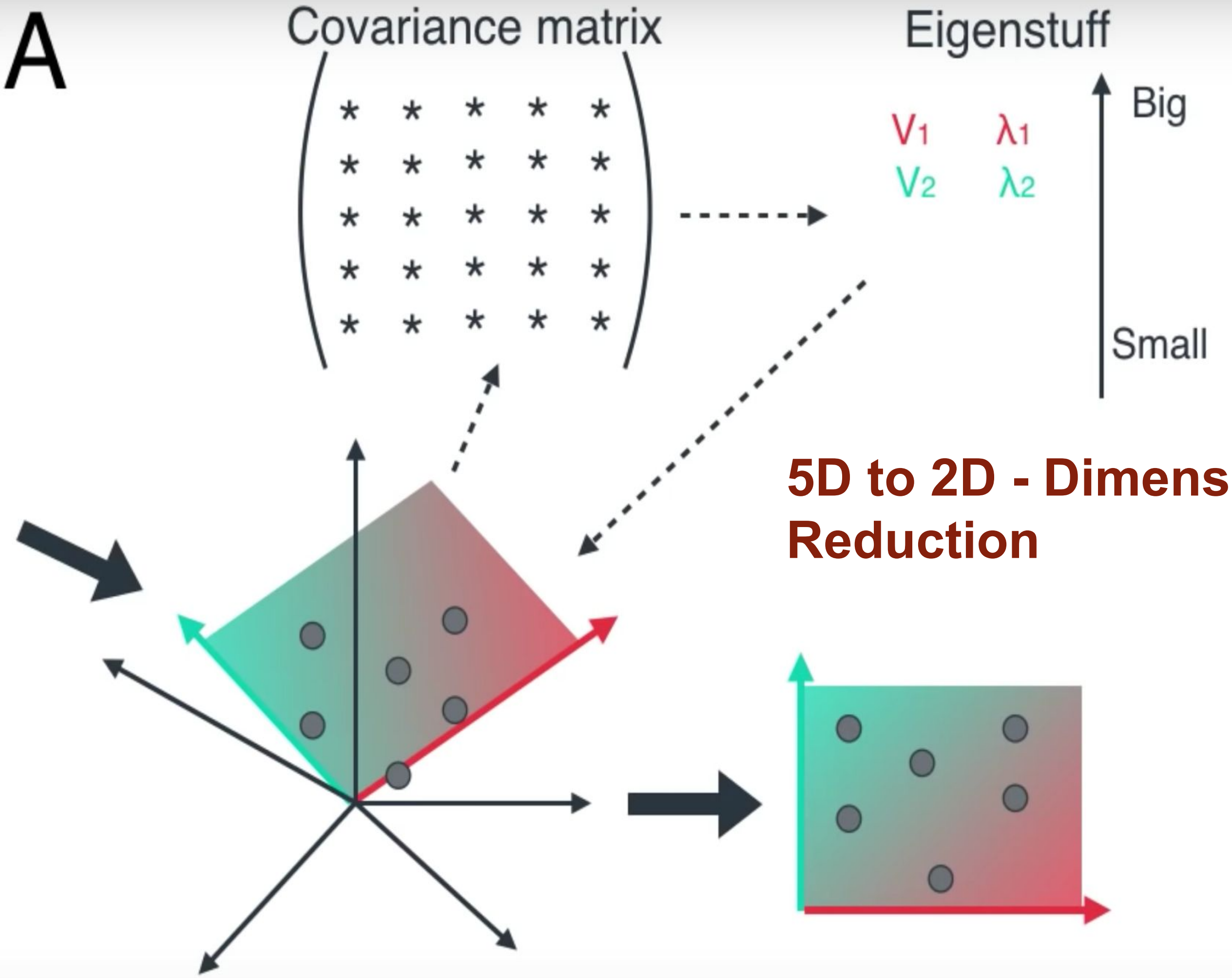

**4.** Remove the Eigen Values that are relatively very small

**5.** Plot the Eigen Vectors on to the original scatter plot of raw data. Here, the **red line** is **PC1** and the **cyan line** is **PC2**. They are ***orthogonal to each other.***

# PCA

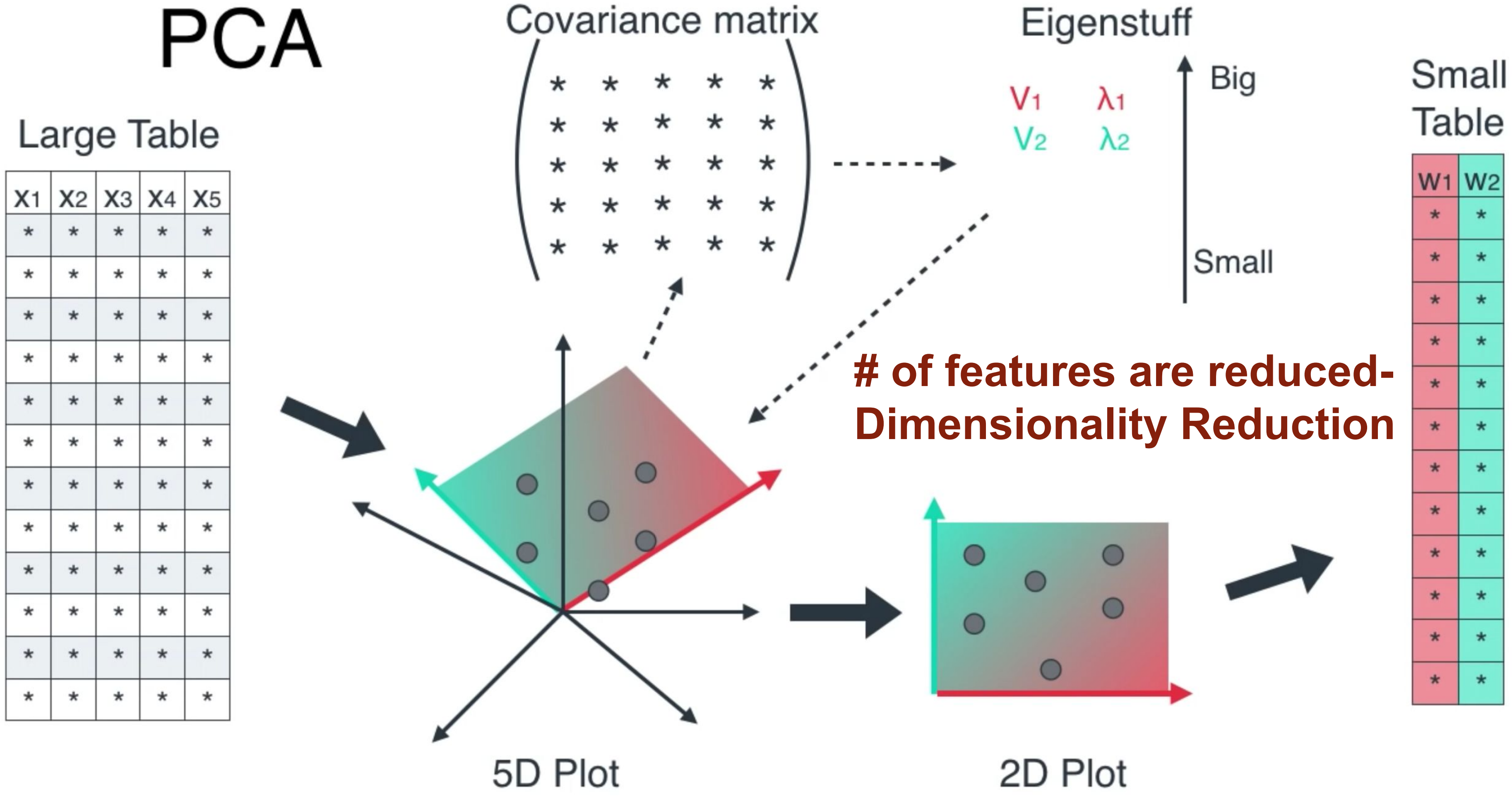
Large Table

| X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |
| * | * | * | * | * |




**5D to 2D - Dimensionality Reduction**

# PCA


Large Table
X1 X2 X3 X4 X5
Covariance matrix
Eigenstuff
V1 λ1
V2 λ2
Big
Small
Small Table
W1 W2
# of features are reduced-
Dimensionality Reduction
5D Plot
2D Plot

# Housing Data

## 5D to 2D - Dimensionality Reduction Example

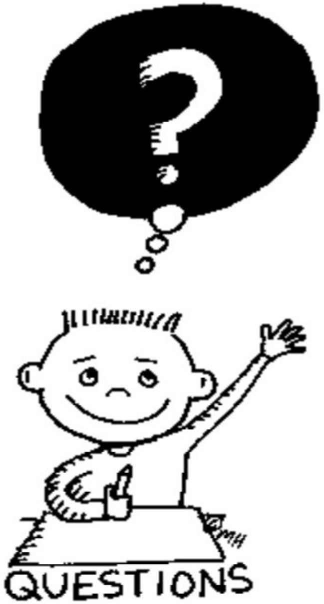
Size
Number of rooms
Number of bathrooms

→ Size feature

Schools around
Crime rate

→ Location feature

# Questions




- How we describe 'most important' features using math?
  - Variance
- How do we represent our data so that the most important features can be extracted easily?
  - Change of basis

**$PX = Y$**

**X** and **Y** be $m \times n$ matrices related by a linear transformation **P**.

**P** is a matrix that transforms **X** into **Y**.

**P**: is the Eigen Vector

# MATH BEHIND
# And
# Numerical solving using
# PCA (Principal Component Analysis)

# Principal component analysis (learning)

The Math

- Given sample $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, \ \ x_i \in \mathcal{R}^d$

Step 1
- compute sample mean: $\hat{\mu} = \frac{1}{n}\sum_i(\mathbf{x}_i)$

Step 2
- compute sample covariance: $\hat{\Sigma} = \frac{1}{n}\sum_i(\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T$

Step 3
- compute eigenvalues and eigenvectors of $\hat{\Sigma}$

$$\hat{\Sigma} = \Phi\Lambda\Phi^T, \ \ \Lambda = diag(\sigma_1^2, \ldots, \sigma_n^2) \ \ \Phi^T\Phi = I$$

Step 4
- order eigenvalues $\sigma_1^2 > \ldots > \sigma_n^2$

- if, for a certain $k$, $\sigma_k << \sigma_1$ eliminate the eigenvalues and eigenvectors above $k$.

Step 5
- Derive the new data set by taking **Y = XV**.

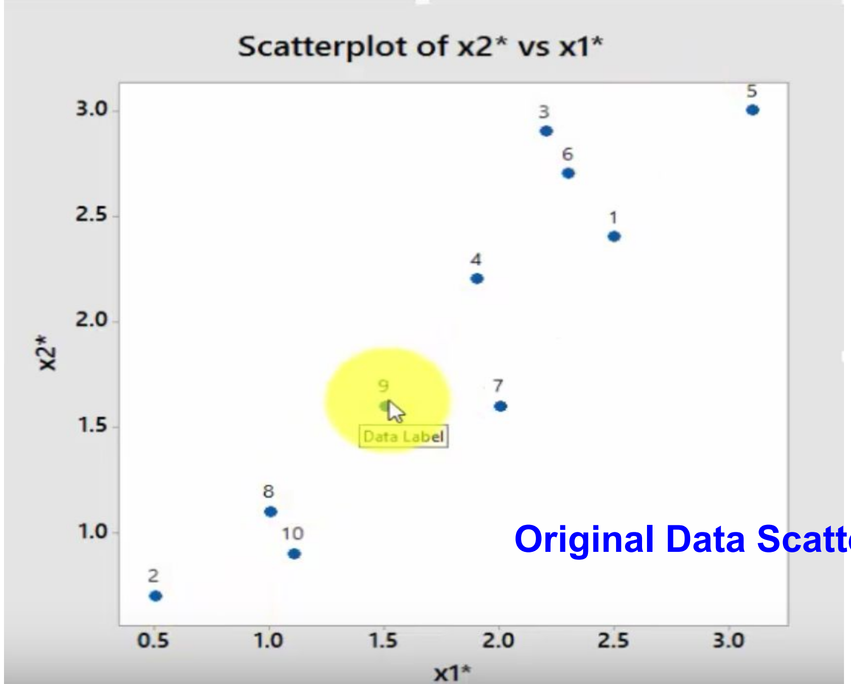
Where ‘V’ is the Eigenvector
‘X’ is the original feature vector

# Numerical solving using PCA (Principal Component Analysis)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_1^*$ | 2.5 | 0.5 | 2.2 | 1.9 | 3.1 | 2.3 | 2.0 | 1.0 | 1.5 | 1.1 |
| $x_2^*$ | 2.4 | 0.7 | 2.9 | 2.2 | 3.0 | 2.7 | 1.6 | 1.1 | 1.6 | 0.9 |

Feature Vector

- # of samples, n=10
- # of features, p = 2
- Bi-Variant Data
- **<u>Goal</u>**: To reduce to univariant data
- That is: **2D to 1D**




Original Data Scatter Plot

**Calculate the mean.**

$$\begin{pmatrix} 1.81 \\ 1.91 \end{pmatrix}$$

Step 1 • compute sample mean: $\widehat{\mu} = \frac{1}{n}\sum_i(\mathbf{x}_i)$

The Math

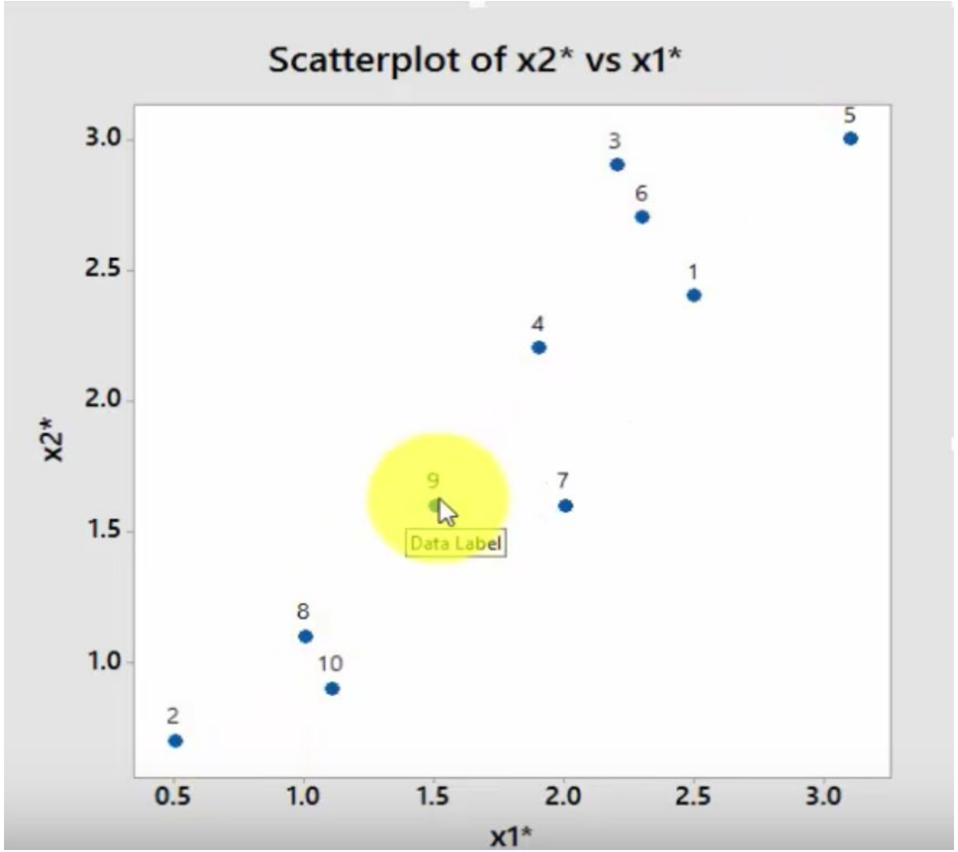
Step 1

Subtract the means from the corresponding data component to re-centre the data set.
Re-construct the scatterplot to view.
Write the "adjusted" data as a matrix **X**.
Note that the "adjusted" data set will have means zero.

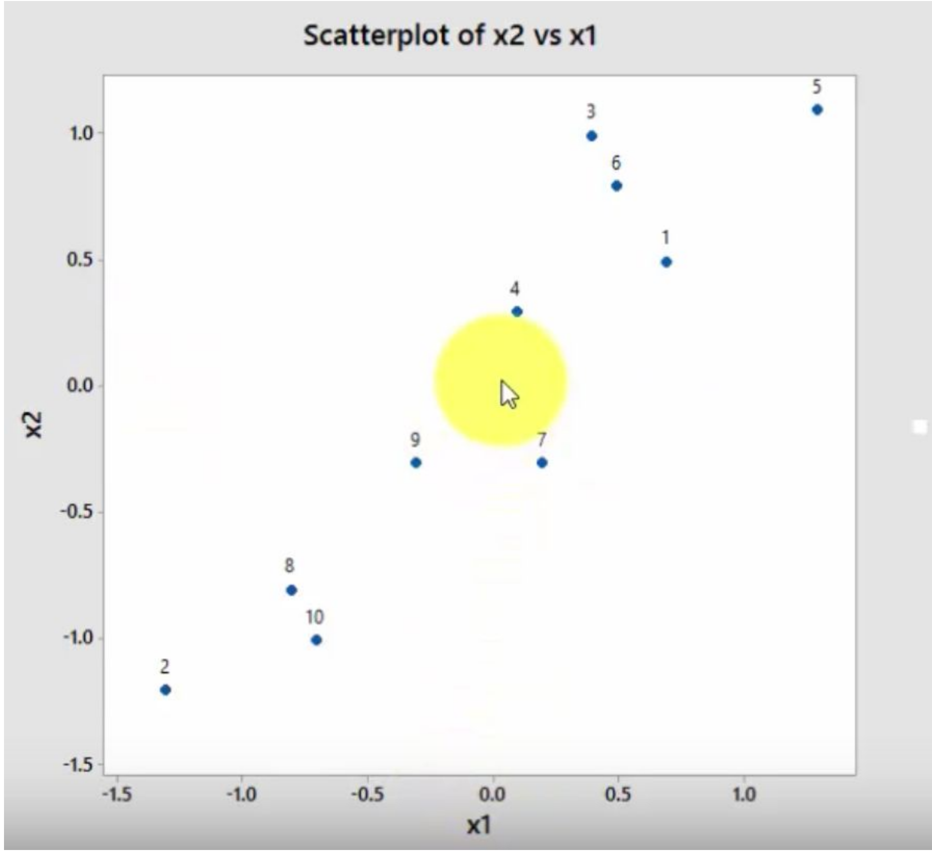
The Numerical

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 0.69 | -1.31 | 0.39 | 0.09 | 1.29 | 0.49 | 0.19 | -0.81 | -0.31 | -0.71 |
| $x_2$ | 0.49 | -1.21 | 0.99 | 0.29 | 1.09 | 0.79 | -0.31 | -0.81 | -0.31 | -1.01 |

"centre"
= mean vector at origin




**Original Data Scatter Plot**




**Original Data - Mean Scatter Plot: Notice Origin is shifted**

**Step 2** Compute the sample variance-covariance matrix C.

Step 2 • compute sample covariance: $\hat{\Sigma} = \frac{1}{n}\sum_i(\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T$

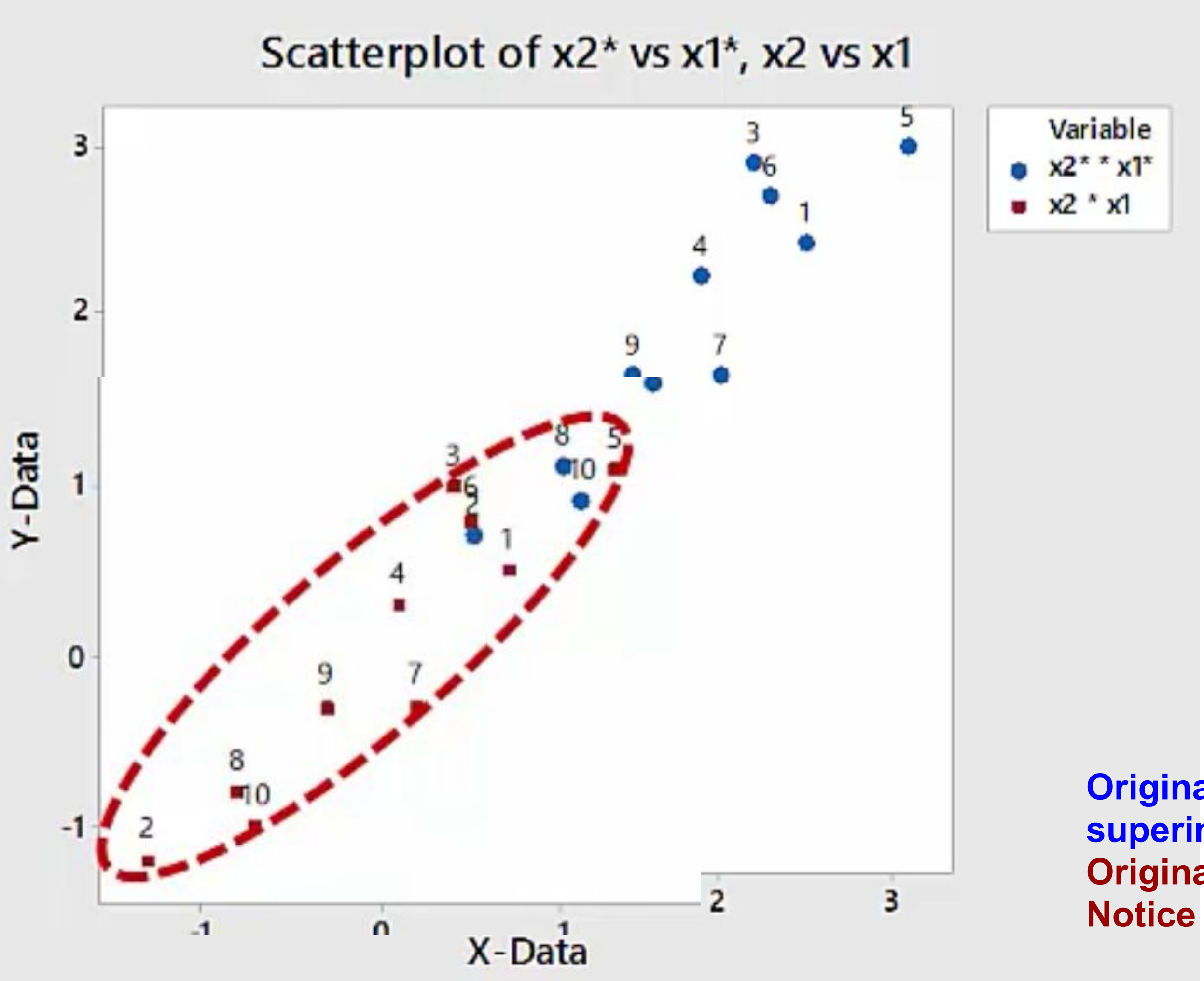
**The Math**

$$C = \frac{1}{10-1} \underbrace{\begin{pmatrix} 0.69 & -1.31 & \cdots \\ 0.49 & -1.21 & \cdots \end{pmatrix}}_{2\times 10} \underbrace{\begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ \vdots & \vdots \end{pmatrix}}_{10\times 2}$$

$$X = \underbrace{\begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ \vdots & \vdots \\ -0.71 & -1.01 \end{pmatrix}}_{10\times 2}$$

$$= \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

**The Numerical**




Find the Axis of Maximum Variance. This will be the Principal Components we want to keep.

**Original Data Scatter Plot superimposed on Original Data - Mean Scatter Plot: Notice Origin is shifted**

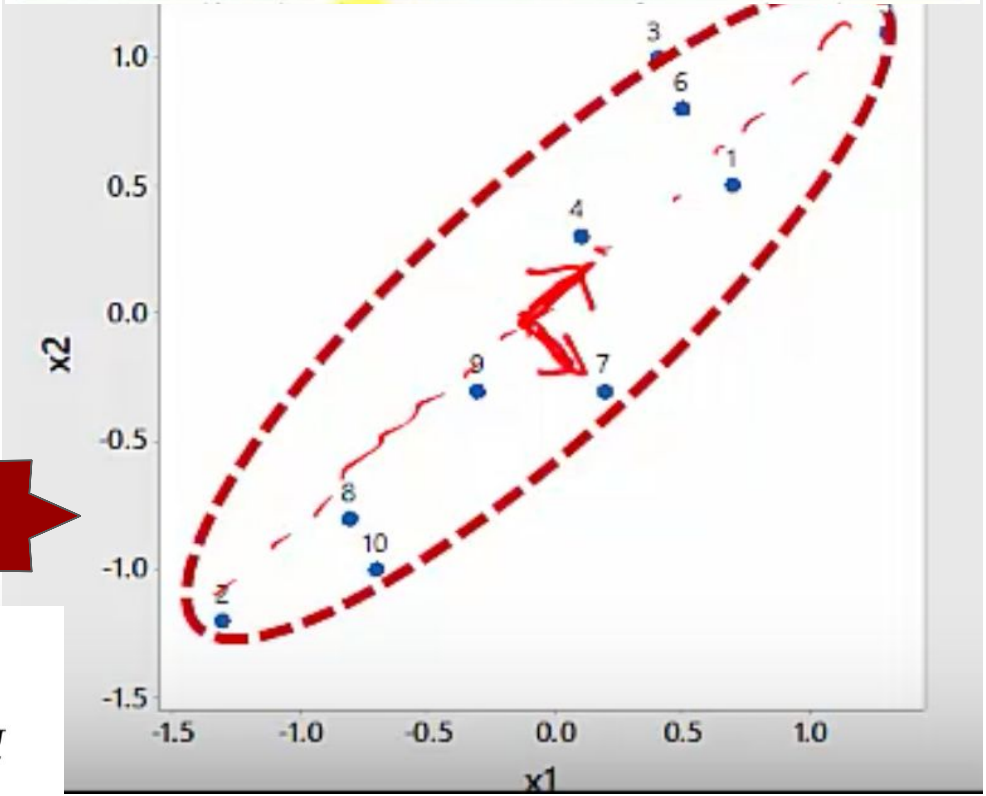
Step 3

Compute the eigenvalues $\lambda_i$, and (unit or normalised) eigenvectors $e_i$ of **C**. order the corresponding pairs from the highest to the lowest eigenvalues.

**normalised** eigenvectors are of length 1.

variance-covariance matrix $C = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$

| Variable | Eigenvector 1 | Eigenvector 2 |
|---|---|---|
| $x_1$ | $v_1 = $ 0.678 | $v_2 = $ 0.735 |
| $x_2$ | 0.735 | −0.678 |
| Eigenvalues | $\lambda_1$=1.2840 | $\lambda_2$=0.0490 |




The Numerical

Step 3

- compute eigenvalues and eigenvectors of $\hat{\Sigma}$

$$\hat{\Sigma} = \Phi \Lambda \Phi^T, \ \Lambda = diag(\sigma_1^2, \ldots, \sigma_n^2) \ \Phi^T \Phi = I$$

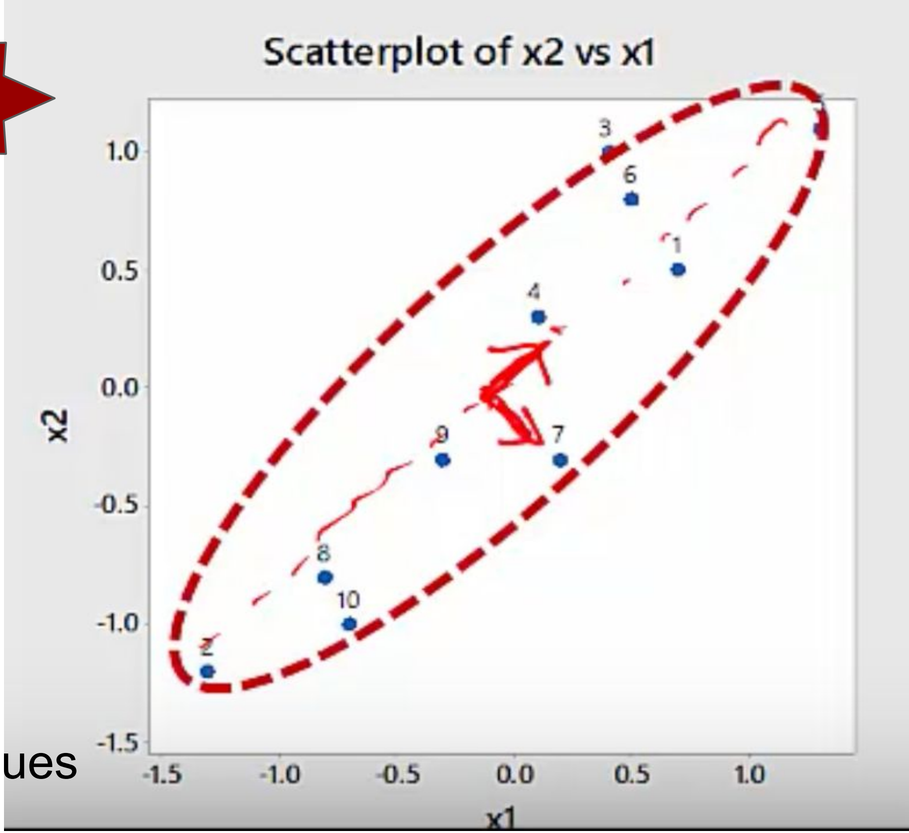
The Math

**Step 3**

Compute the eigenvalues $\lambda_i$, and (unit or normalised) eigenvectors $e_i$ of $\mathbf{C}$, order the corresponding pairs from the highest to the lowest eigenvalues.

**The Numerical**

| Variable | **PC1** Eigenvector 1 | **PC2** Eigenvector 2 |
|---|---|---|
| $x_1$ | $y_1 =$ 0.678 | $y_2 =$ 0.735 |
| $x_2$ | 0.735 | –0.678 |
| Eigenvalues | $\lambda_1$=1.2840 | $\lambda_2$ = 0.0490 |
| % of total variance | 1.28/1.33 = 0.9624 | 0.04/1.33 = 0.037 |




Total Sample Variance = Sum of the Eigen Values

**= 1.2840 + 0.0490 = 1.333**

## Step 4

Choose the components and form the eigenvector matrix **V**.

By ordering the eigenvectors according to the eigenvalues, this gives the components in order of their significance. Hence, the eigenvector with the highest eigenvalue is the principal component. The components of lesser significance can be ignored, so as to reduce the dimensions of the data set.

| Variable | Eigenvector 1 | Eigenvector 2 |
|---|---|---|
| $x_1$ | 0.678 | 0.735 |
| $x_2$ | 0.735 | –0.678 |
| Eigenvalues | 1.2840 | 0.0490 |
| % of total variance | 96.3% | 3.7% |

The Numerical

- Select both components, then **V** =

| Eigenvector 1 | Eigenvector 2 |
|---|---|
| 0.678 | 0.735 |
| 0.735 | –0.678 |

✓

- Or, discard the less significant component, then **V** =

| Eigenvector 1 |
|---|
| 0.678 |
| 0.735 |

## Step 4 transformation matrix **V** based on selection of PCs.

- Select both components, then **V** =

| Eigenvector 1 | Eigenvector 2 |
|---|---|
| 0.678 | 0.735 |
| 0.735 | –0.678 |

- Or, discard the less significant component, then **V** =

| Eigenvector 1 |
|---|
| 0.678 |
| 0.735 |

**The Numerical**

✓

Step 4

- order eigenvalues $\sigma_1^2 > \ldots > \sigma_n^2$

**The Math**

- if, for a certain $k$, $\sigma_k << \sigma_1$ eliminate the eigenvalues and eigenvectors above $k$.

## Step 5

**The Math**

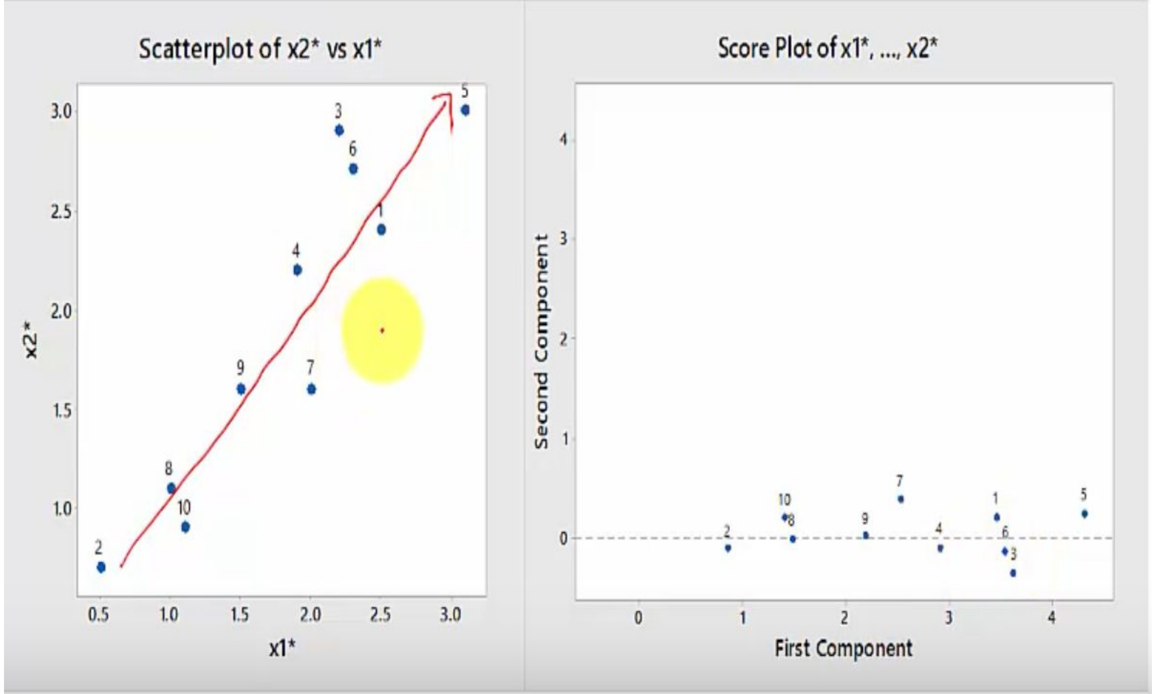
Derive the new data set by taking **Y = XV**.
Basically we have transformed our data so that it is expressed in terms of the patterns between them, where the patterns are the lines that most closely describe the relationships between the data.

## Step 5

- Derive the new data set by taking **Y = XV**.

**The Numerical**

Where 'V' is the Eigenvector
'X' is the original feature vector




$$Y = \begin{bmatrix} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ \vdots & \vdots \\ 1.1 & 0.9 \end{bmatrix} \begin{bmatrix} 0.678 & 0.735 \\ 0.735 & -0.678 \end{bmatrix}$$

$$\therefore Y = \begin{bmatrix} 3.459 & 0.211 \\ 0.854 & -0.107 \\ 3.623 & -0.348 \\ \vdots & \vdots \\ 1.407 & 0.199 \end{bmatrix}$$

$\underset{\sim}{y_2} = 0.735\, \underset{\sim}{x_1^*} - 0.678\, \underset{\sim}{x_2^*}$

$\underset{\sim}{y_1} = 0.678\, \underset{\sim}{x_1^*} + 0.735\, \underset{\sim}{x_2^*}$

## Step 5

Derive the new data set by taking **Y = XV**.

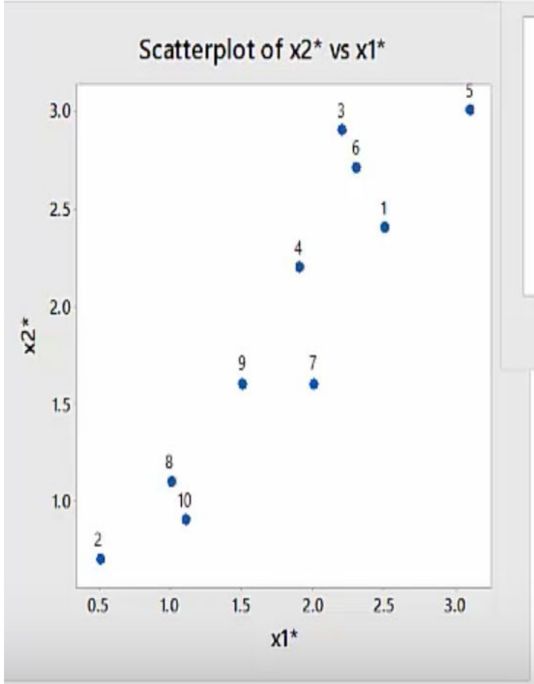
Basically we have transformed our data so that it is expressed in terms of the patterns between them, where the patterns are the lines that most closely describe the relationships between the data.

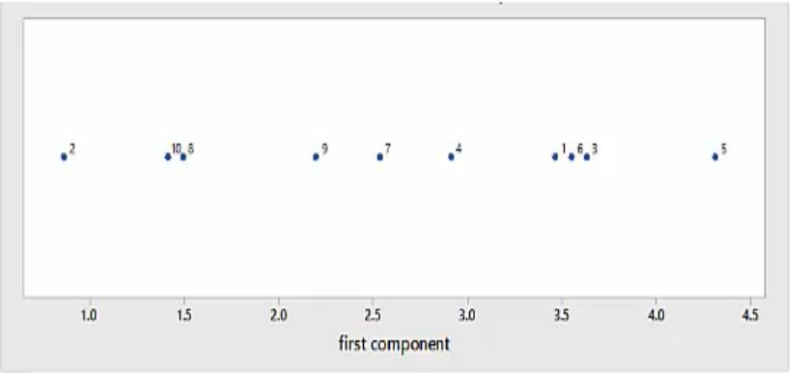
**The Math**

Step 5

- Derive the new data set by taking **Y = XV.**

Where ‘V’ is the Eigenvector
‘X’ is the original feature vector




$$\begin{array}{c} \begin{matrix} x_1^* & x_2^* \end{matrix} \\ Y = \begin{bmatrix} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ \vdots & \vdots \\ 1.1 & 0.9 \end{bmatrix} \begin{bmatrix} 0.678 \\ 0.735 \end{bmatrix} \end{array}$$

$$\therefore Y = \begin{bmatrix} 3.459 \\ 0.854 \\ 3.623 \\ \vdots \\ 1.407 \end{bmatrix}$$

$0.678 x_1^* + 0.735 x_2^*$

**The Numerical**

**Solved Question**:

1. **Do PCA for the following feature vector**

$$\begin{pmatrix} X1 & X2 \\ 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{pmatrix}$$

**a) Compute the Co-Variance Matrix**
**b) Eigen Value**
**c) Eigen Vector**
**d) Scatter Plot of Original Data**
**e) Scatter Plot of the Linear Transformation(2D to 1D)**

(1) Obtain the PCA for feature Vector:

$$\begin{array}{c} X_1 \\ X_2 \end{array} \begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix}$$

Step 1

Ans: • Calculate the Mean $\mu = \begin{bmatrix} 0.5 \\ 2.125 \end{bmatrix}$

• Calculate the Covariance Matrix Vector

$$Covar(X_1, X_2) = \sum_{i=1}^{n} \frac{(X_i - \mu)(X_i - \mu)^T}{n-1}$$

(Note: $n-1$ because its 'Sample' covar not 'population' covar where '$n$' is used)

$X - \mu$

$$\begin{bmatrix} 1.5 & 0.5 & -0.5 & -1.5 \\ 1.875 & 0.875 & -1.125 & -1.625 \end{bmatrix}$$

$(X-\mu)(X-\mu)^T$ → Four Values for four 'X' values

$$\begin{bmatrix} 1.5 \\ 1.875 \end{bmatrix} \begin{bmatrix} 1.5 & 1.875 \end{bmatrix} = \begin{bmatrix} 2.25 & 2.8125 \\ 2.8125 & 3.516 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 \\ 0.875 \end{bmatrix} \begin{bmatrix} 0.5 & 0.875 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.4375 \\ 0.4375 & 0.766 \end{bmatrix}$$

$$\begin{bmatrix} -0.5 \\ -1.125 \end{bmatrix} \begin{bmatrix} -0.5 & -1.125 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.5625 \\ 0.5625 & 1.2656 \end{bmatrix}$$

$$\begin{bmatrix} -1.5 \\ -1.625 \end{bmatrix} \begin{bmatrix} -1.5 & -1.625 \end{bmatrix} = \begin{bmatrix} 2.25 & 2.44 \\ 2.44 & 2.641 \end{bmatrix}$$

Step 2

Summing up

$$= \frac{1}{4-1} \left( \begin{bmatrix} 2.25 & 2.8125 \\ 2.8125 & 3.516 \end{bmatrix} + \begin{bmatrix} 0.25 & 0.4375 \\ 0.4375 & 0.766 \end{bmatrix} + \begin{bmatrix} 0.25 & 0.5625 \\ 0.5625 & 1.2656 \end{bmatrix} + \begin{bmatrix} 2.25 & 2.44 \\ 2.44 & 2.641 \end{bmatrix} \right)$$

$$= \frac{1}{3} \begin{bmatrix} 5 & 6.25 \\ 6.25 & 8.183 \end{bmatrix}$$

$$= \begin{bmatrix} 1.667 & 2.08 \\ 2.08 & 2.727 \end{bmatrix}$$

Variance of X (Diagonal Element)

Variance of Y

Covariance Matrix Vector

Covar of X & Y
Equal Values along reverse diagonal.

- Calculating Eigen Values

$$A \cdot v = \lambda \cdot v$$

$$A \cdot v - \lambda \cdot I \cdot v = 0$$

$$I \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$ Identity Vector

$$(A - \lambda \cdot I) \cdot v = 0$$

ie

$$\det\left[\begin{pmatrix} 1.667 & 2.08 \\ 2.08 & 2.727 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = 0$$

$$\det\left[\begin{pmatrix} 1.667 & 2.08 \\ 2.08 & 2.727 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right] = 0$$

$$\det\left|\begin{bmatrix} 1.667-\lambda & 2.08 \\ 2.08 & 2.727-\lambda \end{bmatrix}\right| = 0$$

$$((1.667-\lambda)\times(2.727-\lambda)) - (2.08)^2 = 0$$

Taking determinants

$$\lambda^2 - 4.4\lambda + 0.2202 = 0$$

$$\boxed{\begin{matrix} \lambda_1 = 4.3494 \\ \lambda_2 = 0.0506 \end{matrix}}$$ Eigen Value.

Step 3

- Calculating Eigen Vectors

$$\begin{bmatrix} 1.667-\lambda & 2.08 \\ 2.08 & 2.727-\lambda \end{bmatrix}\begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

Substituting $\lambda = 4.3494$

$$\begin{bmatrix} 1.667-4.3494 & 2.08 \\ 2.08 & 2.727-4.3494 \end{bmatrix}\begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

$$\begin{bmatrix} -2.6824 & 2.08 \\ 2.08 & -1.6224 \end{bmatrix}\begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

$$-2.6824 a_{11} + 2.08 a_{12} = 0 \to (1)$$

$$2.08 a_{11} - 1.6224 a_{12} = 0 \to (2)$$

given: $a_{11}^2 + a_{12}^2 = 1 \to (3)$

Substitute (1) in (3)

$$(1) \to a_{12} = \frac{2.6824}{2.08} a_{11} = \underline{1.2896 a_{11}}$$

Substitute in (3)

$$a_{11}^2 + (1.2896 a_{11})^2 = 1$$

Pg2

$$a_{11}^2\left(1+(1.2896)^2\right) = 1$$

$$a_{11}^2 = \frac{1}{2.66307} = 0.375$$

$$a_{11} = \sqrt{0.375} = \underline{\underline{0.6123}}$$

|||ly $a_2 = \underline{0.79}$

Substituting $\lambda = 0.0506$

$$\begin{bmatrix} 1.667 - 0.0506 & 2.08 \\ 2.08 & 2.727 - 0.0506 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = 0$$

|||ly Solving $a_{21} = \underline{\underline{0.79}}$

$a_{22} = \underline{\underline{0.61}}$

- Writing the Principal Components of the Feature Vector

$$Z_1 = a_{11} X_1 + a_{12} X_2$$

$$Z_2 = a_{21} X_1 + a_{22} X_2$$

Substituting $a_{11}, a_{12}, a_{21}$ & $a_{22}$

Step 3

$$0.6123 X_1 + 0.79 X_2 = Z_1$$

$$0.79 X_2 + 0.6123 X_1 = Z_2$$

Note: $\left.\begin{matrix} a_{11}, a_{22} \\ a_{12}, a_{21} \end{matrix}\right\}$ are same

This is because the Principal Components $Z_1$ & $Z_2$ are Orthogonal to each other (at Right Angle)

- Getting the Eigen Vectors

Step 3

| Variable | Eigen Vector 1 | Eigen Vector 2 |
|---|---|---|
| $X_1$ | 0.6123 | 0.79 |
| $X_2$ | 0.79 | 0.6123 |
| Eigen Values | $4.3494 = \lambda_1$ | $0.0506 = \lambda_2$ |
| Total Eigen Value | $\lambda = \lambda_1 + \lambda_2 =$ | $4.35 + 0.05 = 4.4$ |
| % of Total Variance | $\frac{4.35}{4.4} = \underline{98\%}$ | $\frac{0.05}{4.4} = 1\%$ |

Pg3

- Deciding which amoung the Principal components have the max effect

Here, it is the Eigen Vector associated with max Eigen Value

Step 4

| Eigen Vector | Eigen Value |
|---|---|
| $\begin{bmatrix} 0.6123 \\ 0.79 \end{bmatrix}$ | $\lambda_1 = 4.3494$ ✓ |
| $\begin{bmatrix} 0.79 \\ 0.6123 \end{bmatrix}$ | $\lambda_2 = 0.05$ |

- Derive new Transformation dataset by taking $Y = XV$

  ↳ The Eigen Vector of the Principal Component chosen, here PC1

Step 5

$$\underbrace{\begin{array}{cc} X_1 & X_2 \\ \end{array}}_{}$$

$$\underbrace{\begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix}}_{X} \cdot \underbrace{\begin{bmatrix} 0.6123 \\ 0.79 \end{bmatrix}}_{V}$$

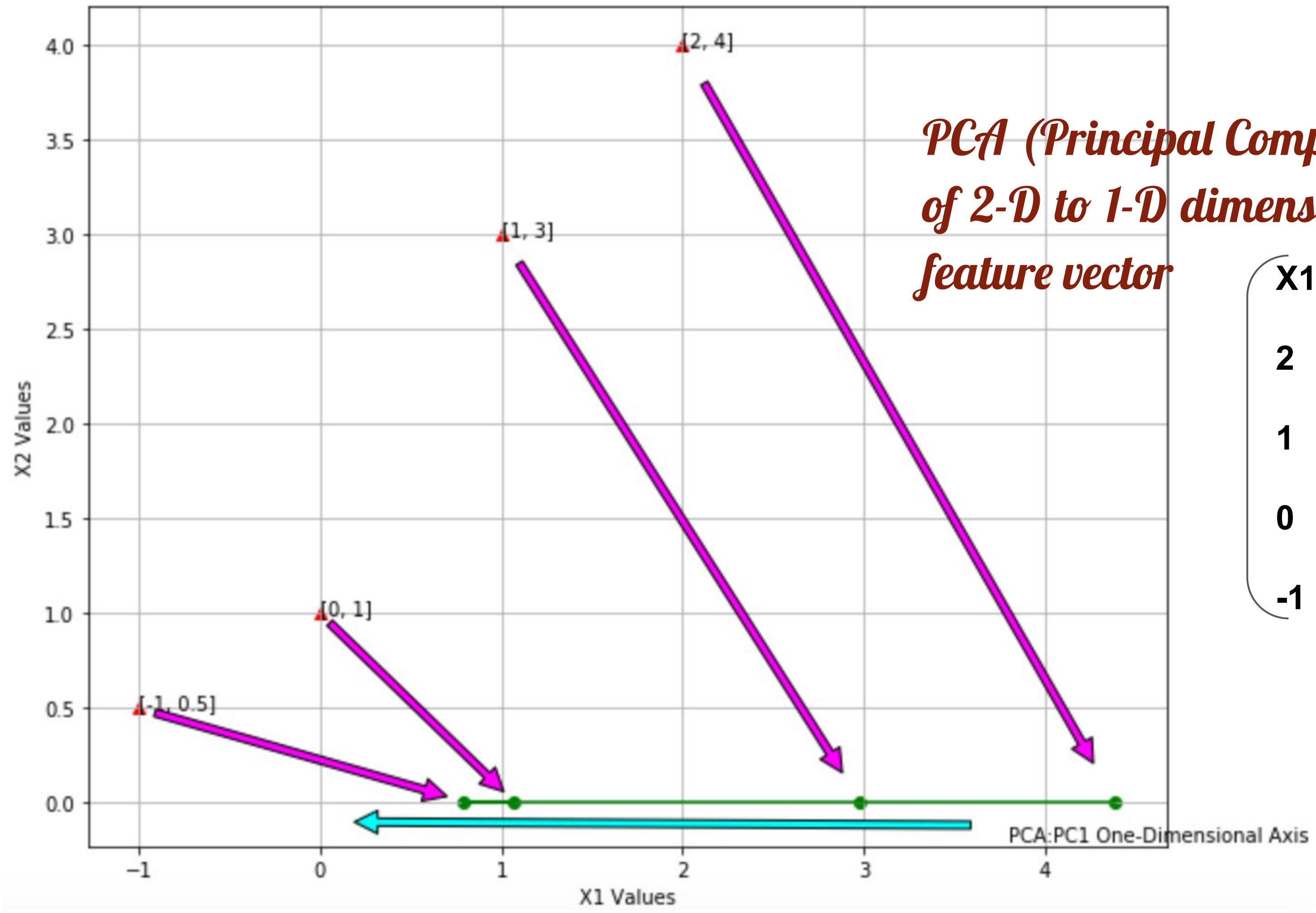
$$= \begin{bmatrix} 4.385 \\ 2.98 \\ 0.79 \\ 1.0673 \end{bmatrix}$$

Step 5

- Plot the Original Feature Vector and the Transformed Feature Vector as a Scatter Plot

Here, it is a 2D to 1D dimension reduction.

Pg 4

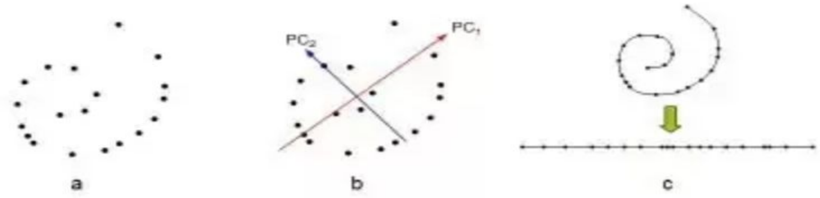



***PCA (Principal Component Analysis)plot of 2-D to 1-D dimensionality reduction of feature vector***

$$\begin{pmatrix} \mathbf{X1} & \mathbf{X2} \\ 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{pmatrix}$$

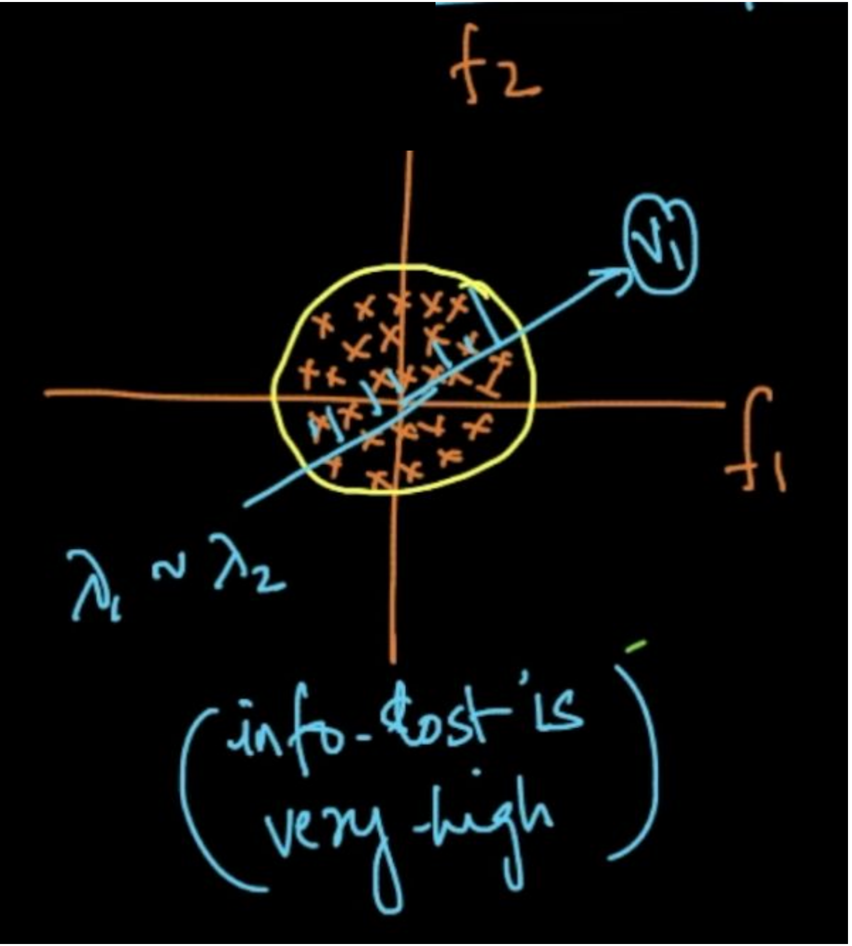
# Limitations of PCA

- If the data is not linearly correlated (f.e. in spiral, where x=t*cos(t) and y =t*sin(t) ), PCA is not enough




- PCA relies on mean and co-variance. This means it assumes that the data follows a Gaussian distribution. This may not always be true - in which case PCA fails.

# Limitations of PCA

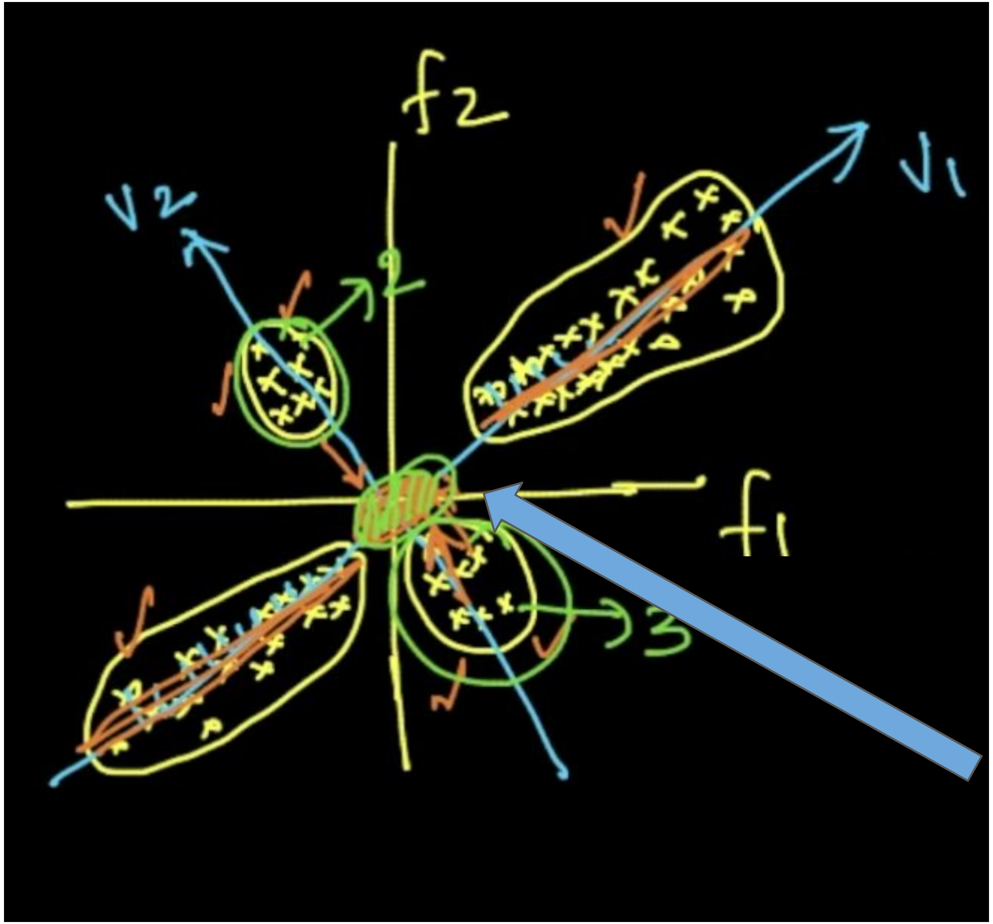



When the Eigen values are the same, that is, $\lambda_1 = \lambda_2$ it means the he data is not linearly separable, e.g

- Circle in 2D space
- Sphere in 3D space
- Hyper-Sphere in high dimensions

When PCA is applied and the data is projected on to the PC Axis (here V1), the loss of information is very high.

# Limitations of PCA

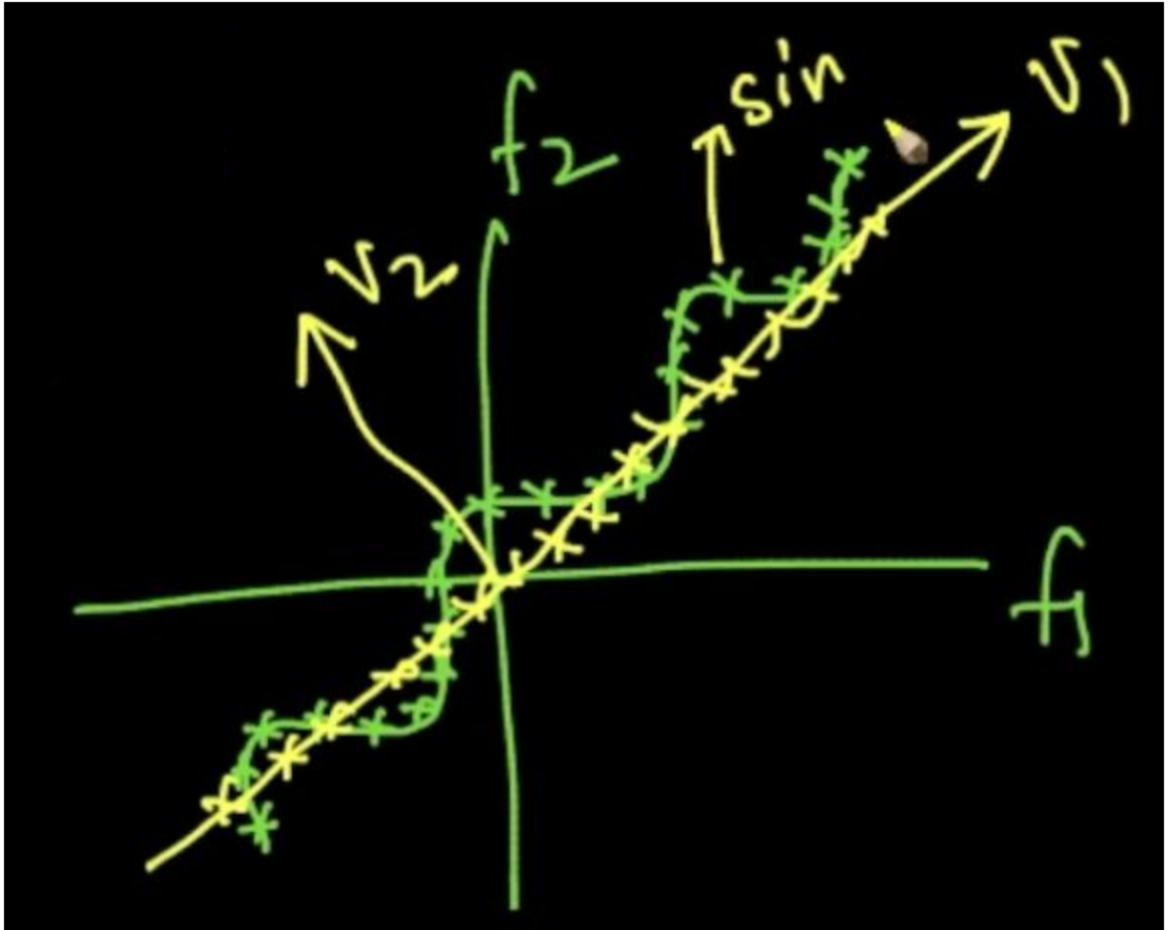



When the original dataset when plotted form distinct clusters as shown in the figure, it distinctly separates the data in the original feature space.

But after applying PCA, the data is projected on to the Principal Component Axis , here $V_1$ .

This means the dataset marked ‘2’ and ‘3’ along $V_2$, would be projected in the middle of $V_1$. These two datasets which is clearly distinguishable in the original feature space, becomes cluttered and **not** clearly separable in the projected space over $PC_1$.

# Limitations of PCA




The shape of the original dataset is lost, when projected on to the PC axes. E.g. Here the original data is a sine-wave, where as the projected data after PCA, is a set of points along $V_1$.

## <u>Unsolved Question</u>:

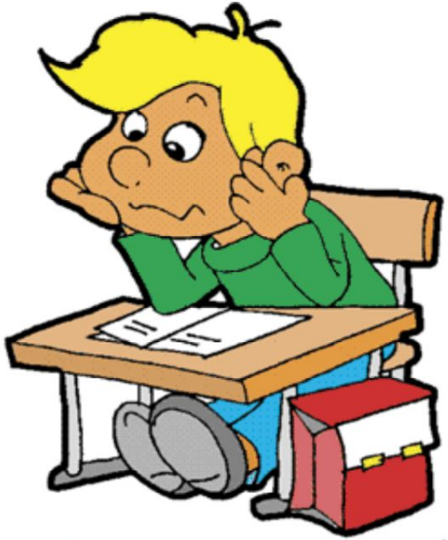
**2. Do PCA for the following feature vector**

$$\begin{array}{cc} \begin{matrix} X1 \\ X2 \end{matrix} & \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 5 & 3 & 6 & 7 & 8 \end{pmatrix} \end{array}$$

**a) Compute the Co-Variance Matrix**
**b) Eigen Value**
**c) Eigen Vector**
**d) Scatter Plot of Original Data**
**e) Scatter Plot of the Linear Transformation(2D to 1D)**

# Example of a problem

- We collected $m$ parameters about 100 students:
  - Height
  - Weight
  - Hair color
  - Average grade
  - …



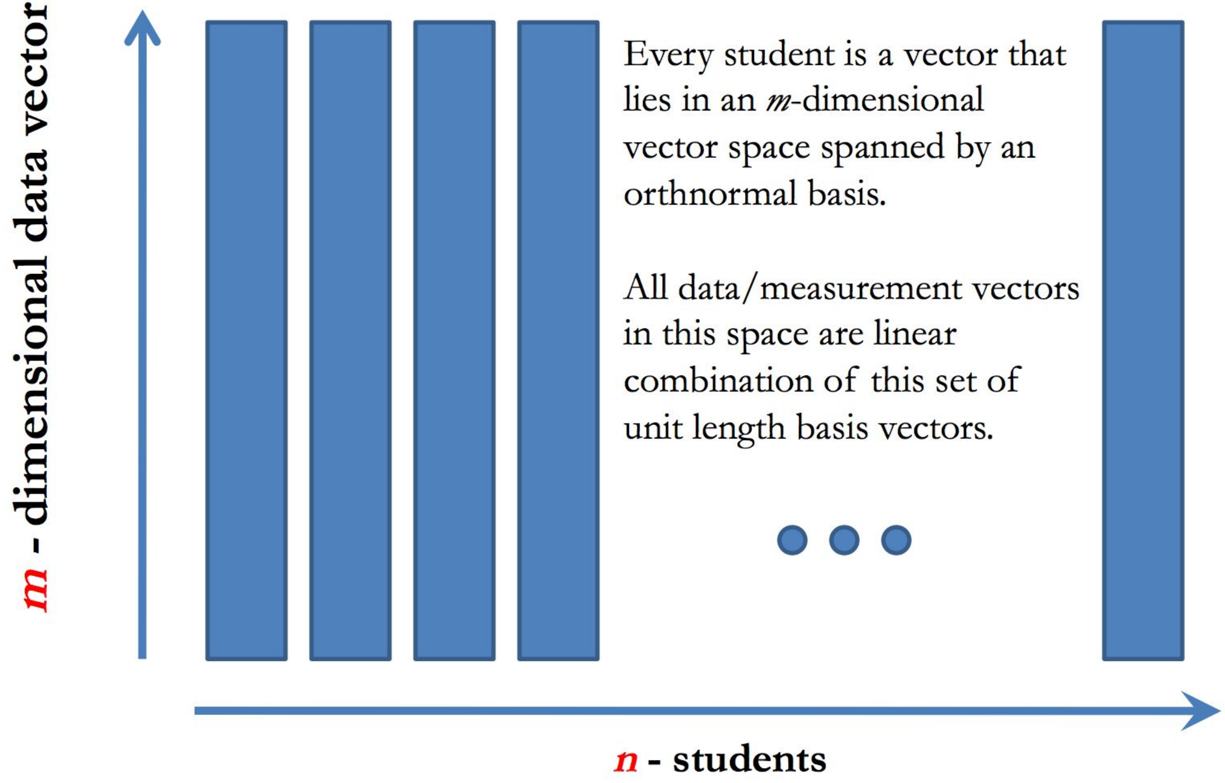
- We want to find the most important parameters that best describe a student.

# Example of a problem

- Each student has a vector of data which describes him of length $m$:
  - (180,70,'purple',84,…)

- We have $n = 100$ such vectors. Let's put them in one matrix, where each column is one student vector.
- So we have a $m$x$n$ matrix. This will be the input of our problem.

# Example of a problem




Every student is a vector that lies in an *m*-dimensional vector space spanned by an orthnormal basis.

All data/measurement vectors in this space are linear combination of this set of unit length basis vectors.

# Which parameters can we ignore?

- Constant parameter (number of heads)
  - 1,1,…,1.
- Constant parameter with some noise - (thickness of hair)
  - 0.003, 0.005,0.002,….,0.0008 ➔ low variance
- Parameter that is linearly dependent on other parameters (head size and height)
  - $Z = aX + bY$

# Which parameters can we ignore?

- Constant parameter (number of heads)
  - 1,1,…,1.
- Constant parameter with some noise - (thickness of hair)
  - 0.003, 0.005,0.002,…,0.0008 ➔ low variance
- Parameter that is linearly dependent on other parameters (head size and height)
  - $Z= aX + bY$

# Which parameters do we want to keep?

- Parameter that doesn't depend on others (e.g. eye color), i.e. uncorrelated ➔ low covariance.
- Parameter that changes a lot (grades)
  - The opposite of noise
  - High variance