Unit-4
Dimensionality Reduction

Syllabus

- **Dimensionality Reduction:**

  Singular Value Decomposition

  Principal Component Analysis

  Linear Discriminated Analysis

What is Dimensionality Reduction?

- The number of input features, variables, or columns present in a given dataset is known as dimensionality, and the process to reduce these features is called dimensionality reduction.

- A dataset contains a huge number of input features in various cases, which makes the predictive modeling task more complicated, for such cases, dimensionality reduction techniques are required to use.

Dimensionality Reduction…?

- Dimensionality reduction technique can be defined as, ***"It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information."***

- These techniques are widely used in Machine Learning for obtaining a better fit predictive model while solving the classification and regression problems.

- Handling the high-dimensional data is very difficult in practice, commonly known as the *curse of dimensionality.*

Benefits of Dimensionality Reduction..

- By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
- Less Computation training time is required for reduced dimensions of features.
- Reduced dimensions of features of the dataset help in visualizing the data quickly.
- It removes the redundant features (if present).

Two ways of Dimensionality Reduction

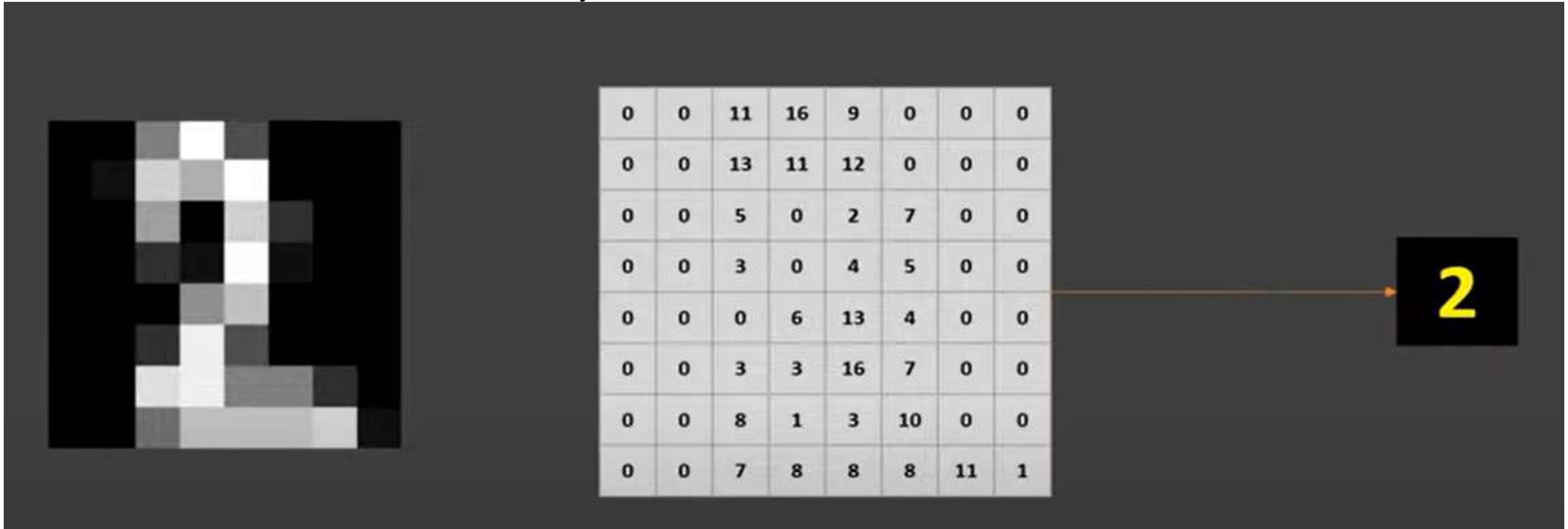
- 1. Feature Selection
- 2. Feature Extraction

Feature Selection

- Feature selection is the process of selecting the subset of the relevant features and leaving out the irrelevant features present in a dataset to build a model of high accuracy. In other words, it is a way of selecting the optimal features from the input dataset.

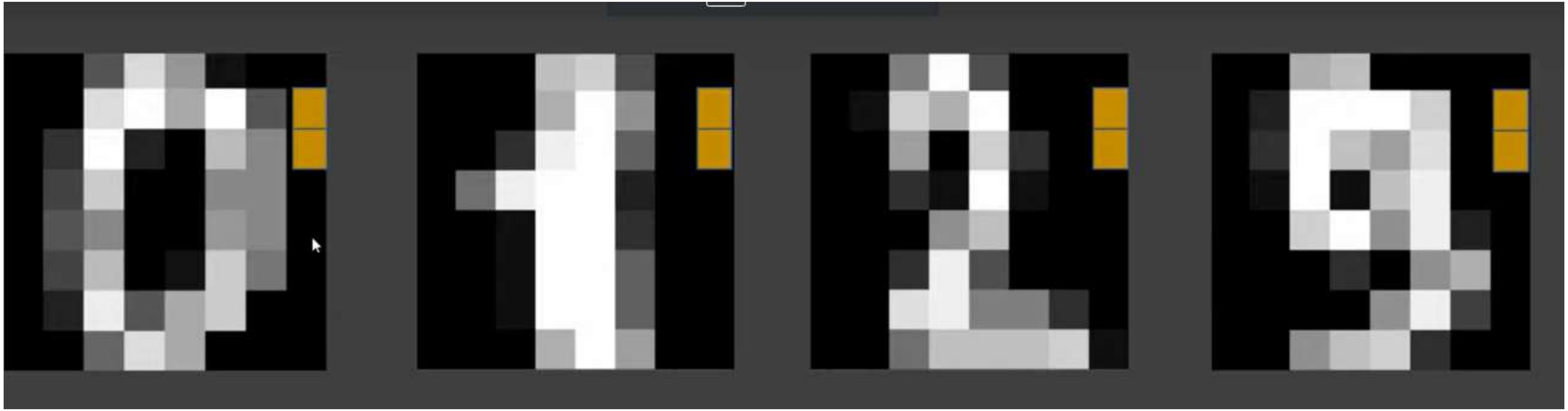
General – features reduction technique

- In this example number 2 has 64 features… but many of them are of no importance to decide the characteristics of 2, are removed first.



0	0	11	16	9	0	0	0
0	0	13	11	12	0	0	0
0	0	5	0	2	7	0	0
0	0	3	0	4	5	0	0
0	0	0	6	13	4	0	0
0	0	3	3	16	7	0	0
0	0	8	1	3	10	0	0
0	0	7	8	8	8	11	1

Remove features which are of no importance



Feature Selection – 3 Methods

- 1.Filter Method
  - **Correlation**
  - **Chi-Square Test**
  - **ANOVA**
  - **Information Gain, etc.**

- 2.Wrapper Method
  - Forward Selection
  - Backward Selection
  - Bi-directional Elimination

- 3.Embedded Method
  - **LASSO**
  - **Elastic Net**
  - **Ridge Regression, etc.**

Feature Extraction

- Feature extraction is the process of transforming the space containing many dimensions into space with fewer dimensions.

- This approach is useful when we want to keep the whole information but use fewer resources while processing the information.

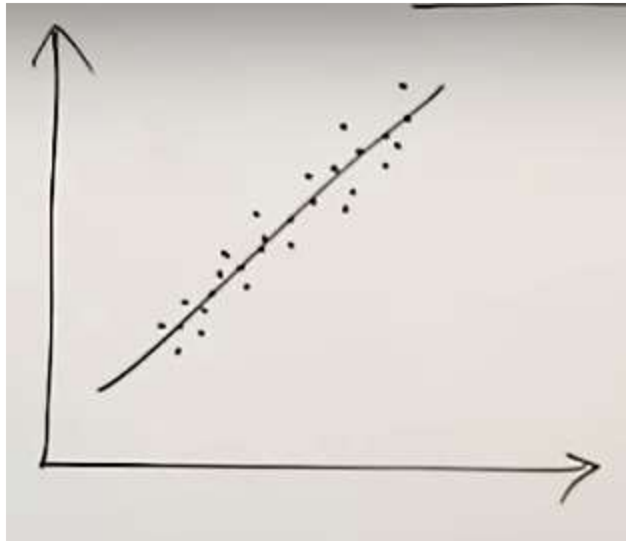
Some common feature extraction techniques are:

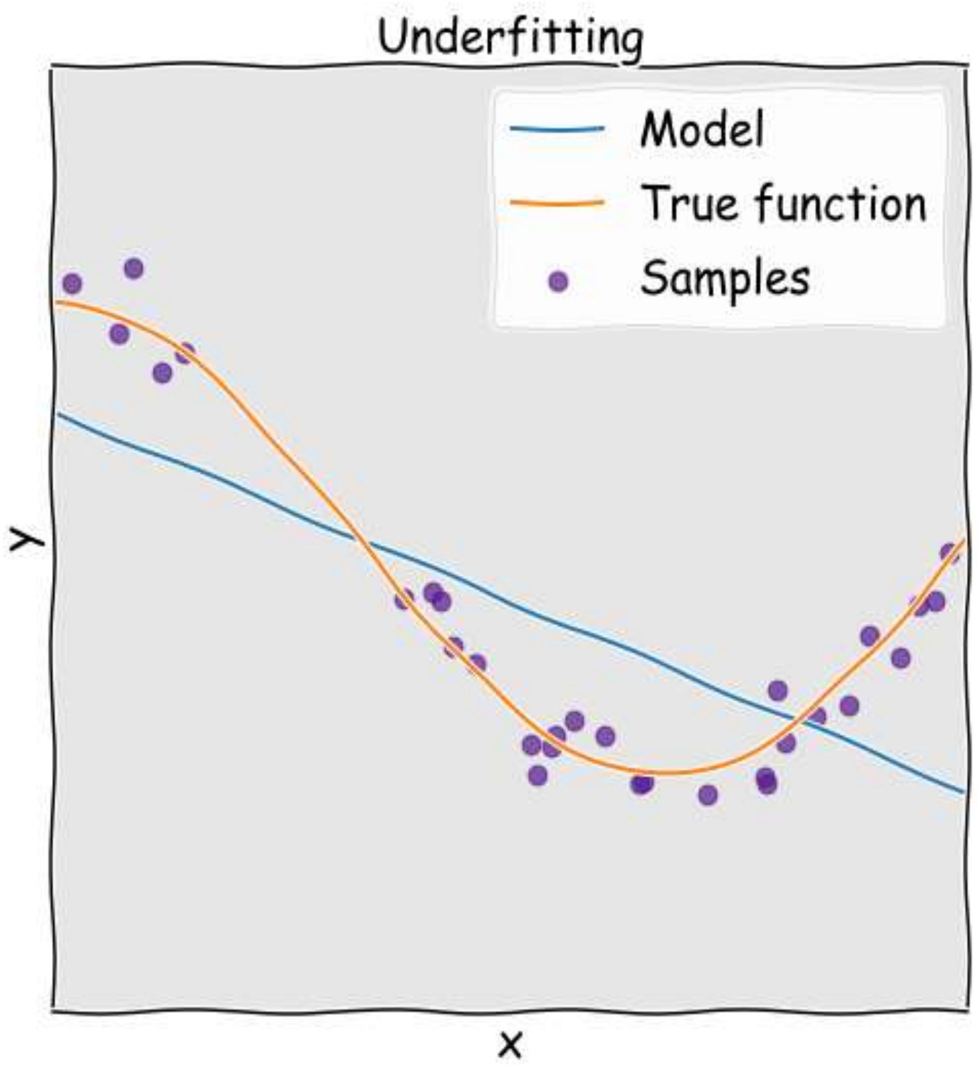
1. Principal Component Analysis (PCA)
2. Linear Discriminant Analysis (LDA)
3. Kernel PCA
4. Quadratic Discriminant Analysis (QDA)etc.

ML Model design

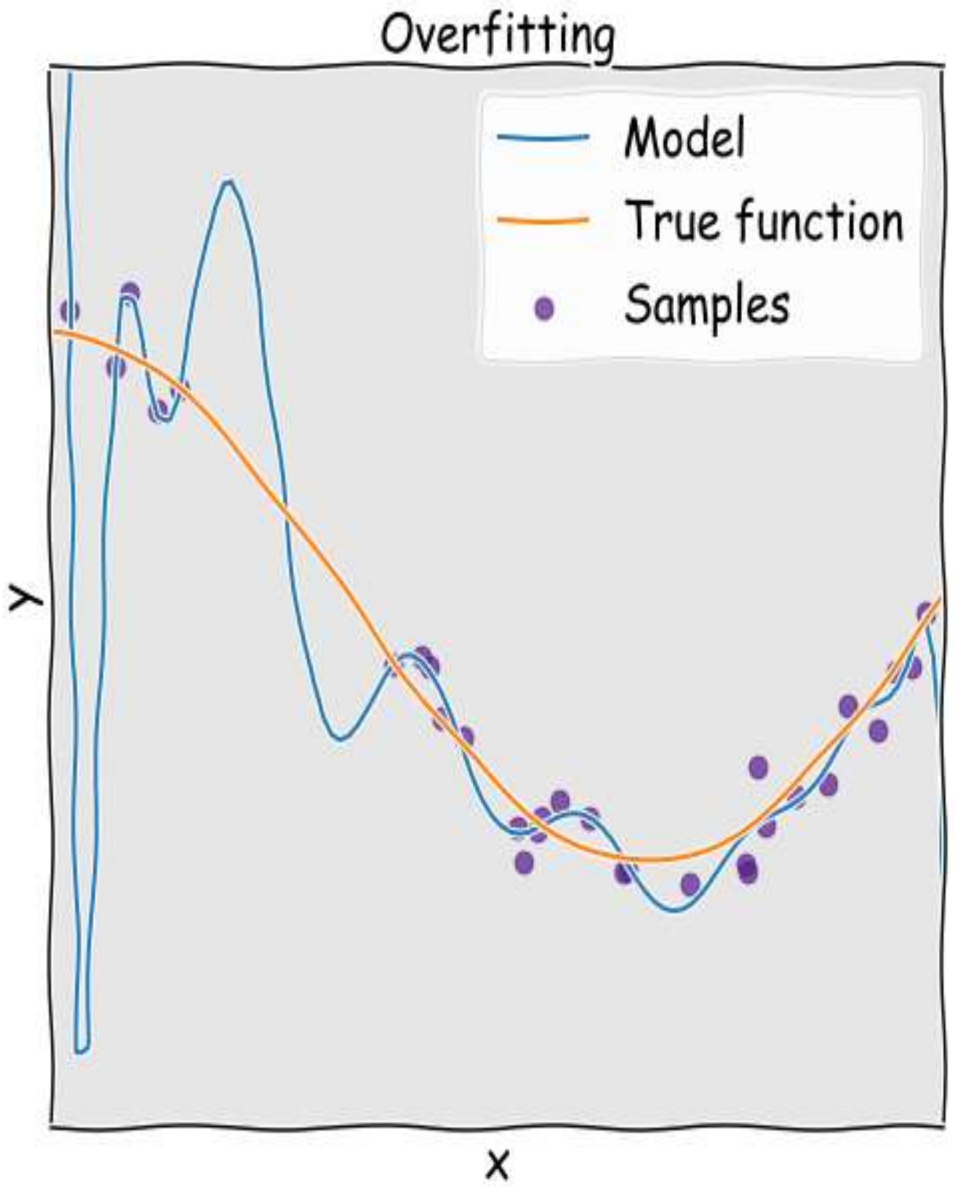
- Consider the line passing through the samples in the diagram.
- It (line) is the model/function/hypothesis generated after the training phase.

The line is trying to reach all the samples as close as possible.

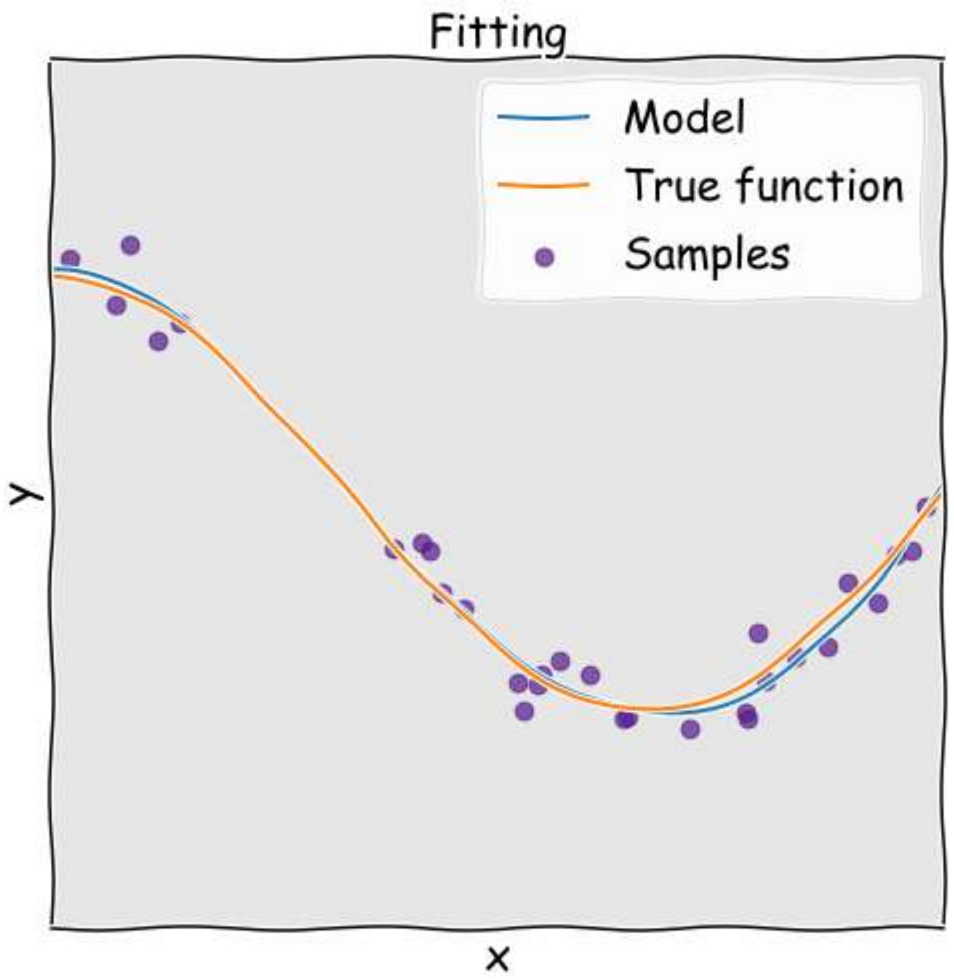




- If we have an underfitted model, this means that we do not have enough parameters to capture the trends in the underlying system.
- In general, in underfitting, model fails during testing as well as training.



- In this a complex model is built using too many features.
- During training phase, model works well. But it fails during testing.



- Under/Overfitting can be solved in different ways.
- One of the solution for overfitting is dimensionality reduction.
- Diagram shows that model neither suffers from under or overfitting.

Example to show requirement of Dimensionality reduction

- In this example important features to decide the price are town, area and plot size. Features like number of bathroom and trees nearby may not be significant, hence can be dropped.

town	area	bathroom	plot	trees nearby	price
monroe	2600	2	8500	2	550000
monroe	3000	3	9200	2	565000
monroe	3200	3	8750	2	610000
monroe	3600	4	10200	2	680000
monroe	4000	4	15000	2	725000
west windsor	2600	2	7000	2	585000
west windsor	2800	3	9000	2	615000
west windsor	3300	4	10000	1	650000
west windsor	3600	4	10500	1	710000

PCA

- PCA is a method of Dimensionality Reduction.
- PCA is a process of identifying Principal Components of the samples.
- It tries to address the problem of overfitting.

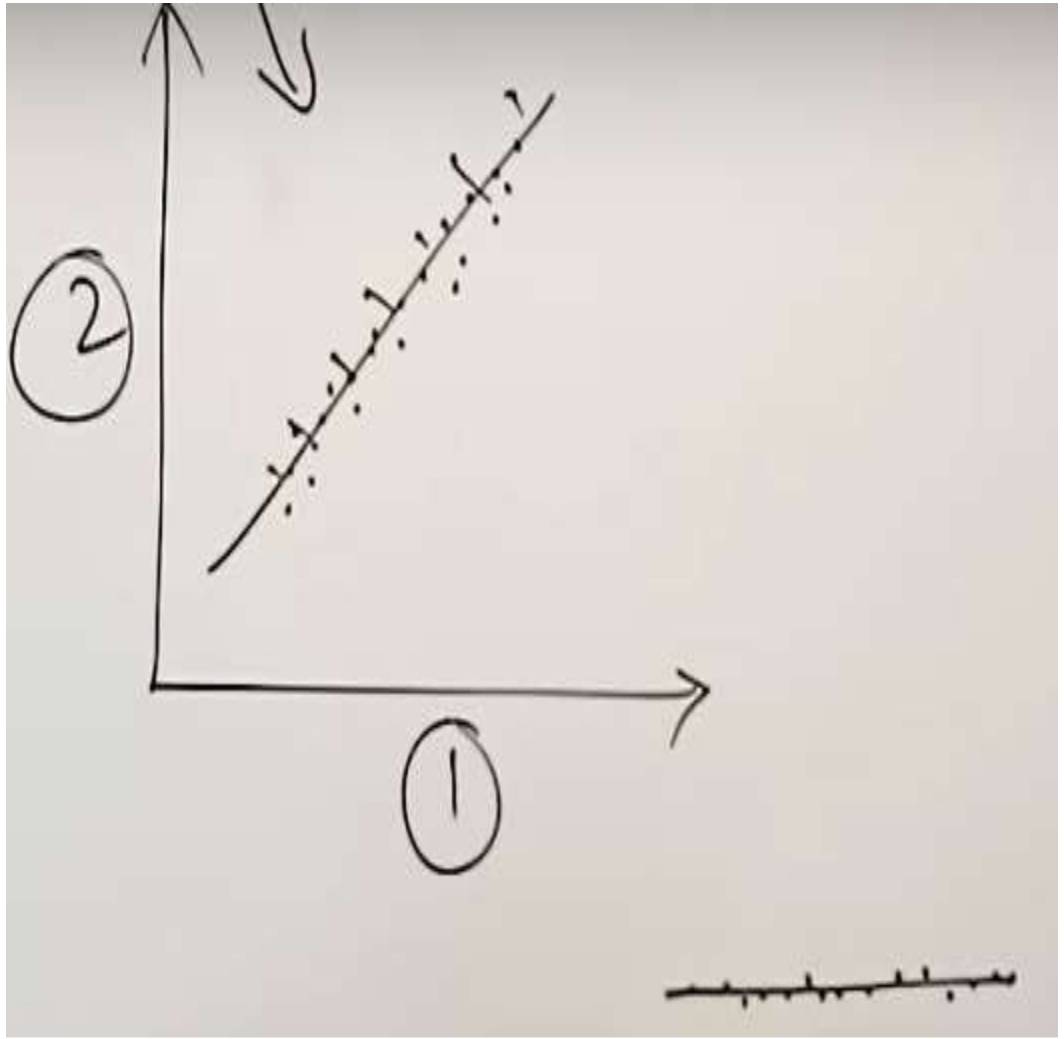
Example for PCA (from SK learn (SciKit Learn) library)

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	...	pixel_64	Digit
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	...	0.0	5
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	...	0.0	0
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	...	5.0	2
3	0.0	0.0	7.0	15.0	13.0	1.0	0.0	...	9.0	9
4	0.0	0.0	0.0	1.0	11.0	0.0	0.0	...	0.0	1

PCA (n_components=6)

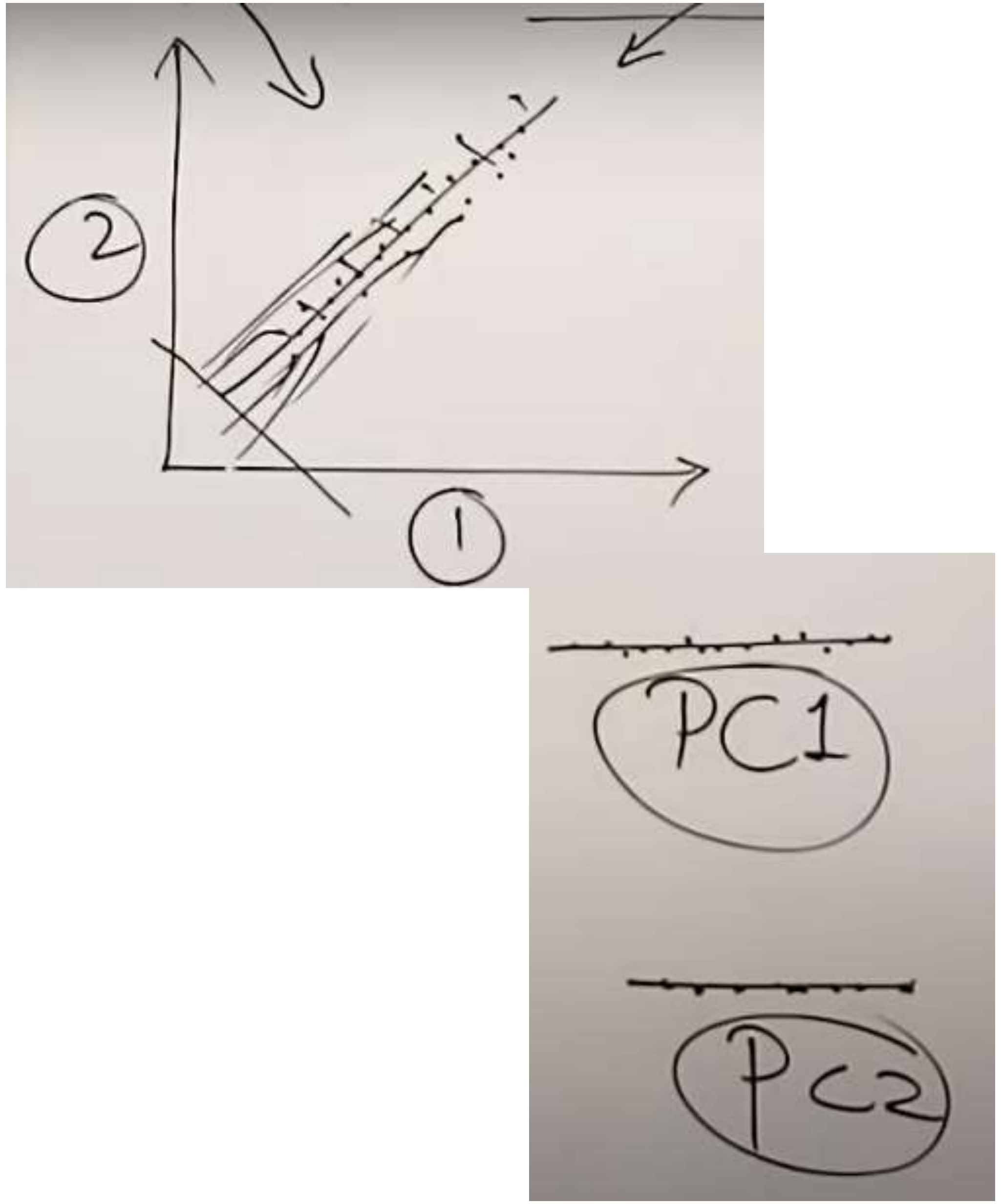
	PC 1	PC 2	PC 4	PC 5	PC 6	Digit
0	5.2	15.0	1.23	6.01	15.0	5
1	7.2	13.0	0.0	9.5	12.0	0
2	1.23	4.0	9.0	9.0	15.0	2
3	9.5	12.0	7.2	7.2	0.0	9
4	8.1	0.0	13.0	15.0	5.2	1

What does PCA do?



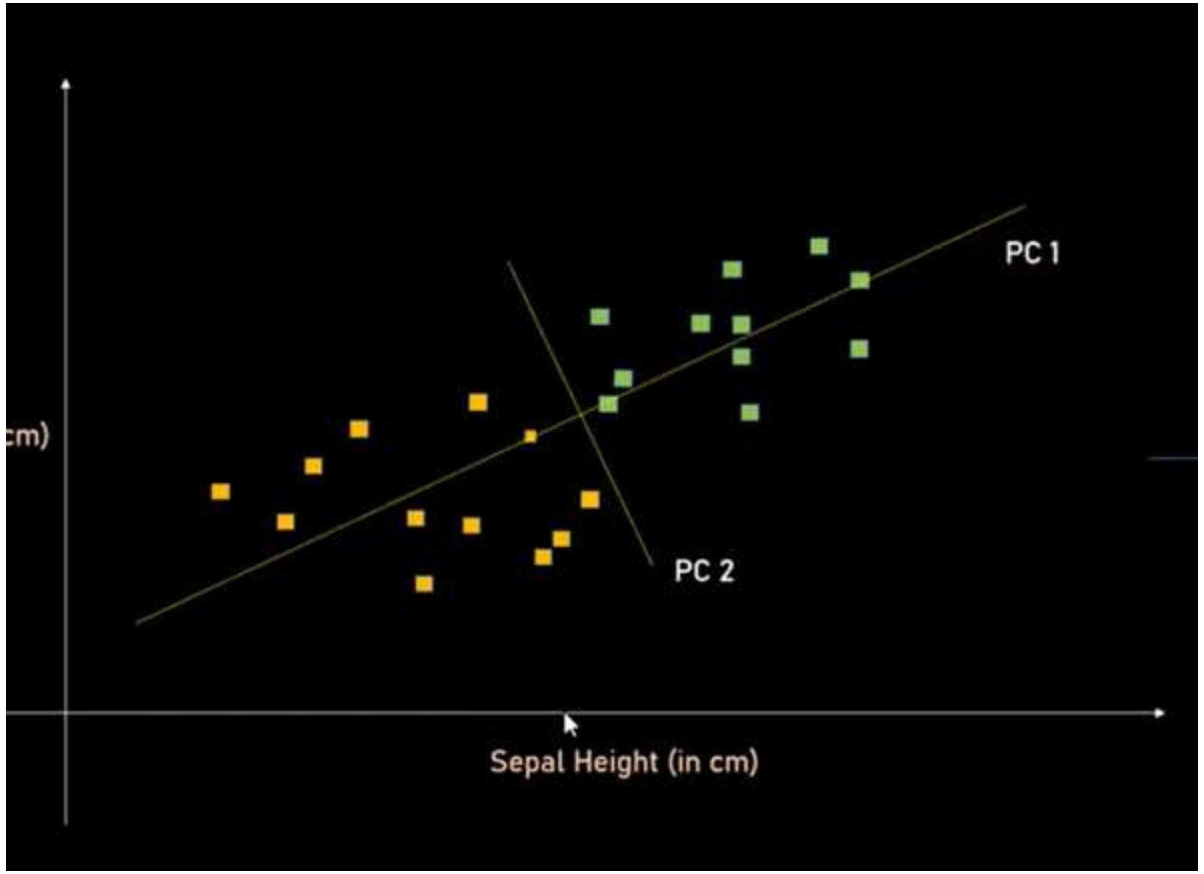
- To address overfitting, reduce the dimension, without loosing the information.
- In this example two dimension is reduced to single dimension.
- But in general, their can be multiple dimensions… and will be reduced.
- When the data is viewed from one angle, it will be reduced to single dimension and the same is shown at the bottom right corner, and this will be Principal Component 1.

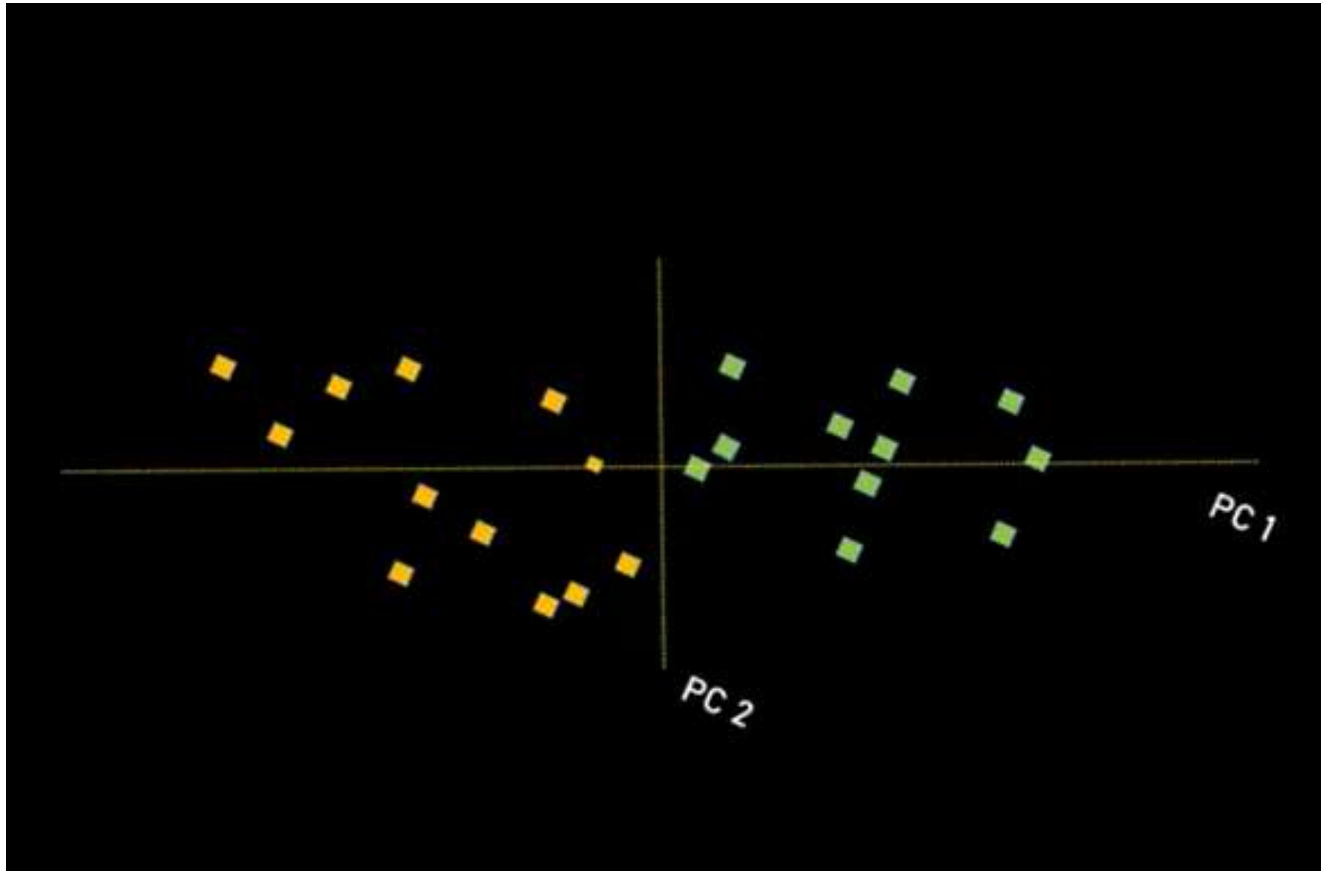
Similarly compute PC2



- Figure shows the representation of PC1 and PC2.
- Like this we have several principal components…
- Say PC1,PC2, PC3… and so on..
- In that PC1 will be of top priority.
- Each Principal Components are independent and are orthogonal. It means one PC does not depends on another…all of them are independent.

Another Example

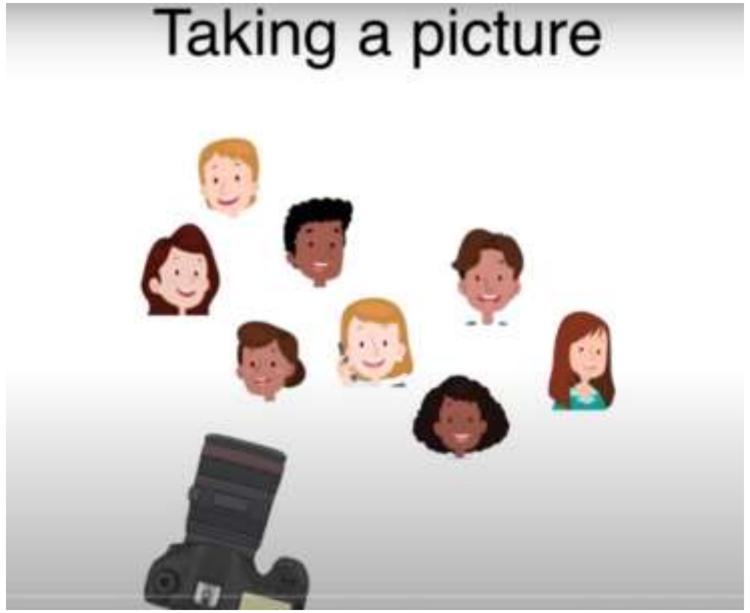




Example to illustrate the PC

Taking a picture

Multiple angles in which picture can be captured

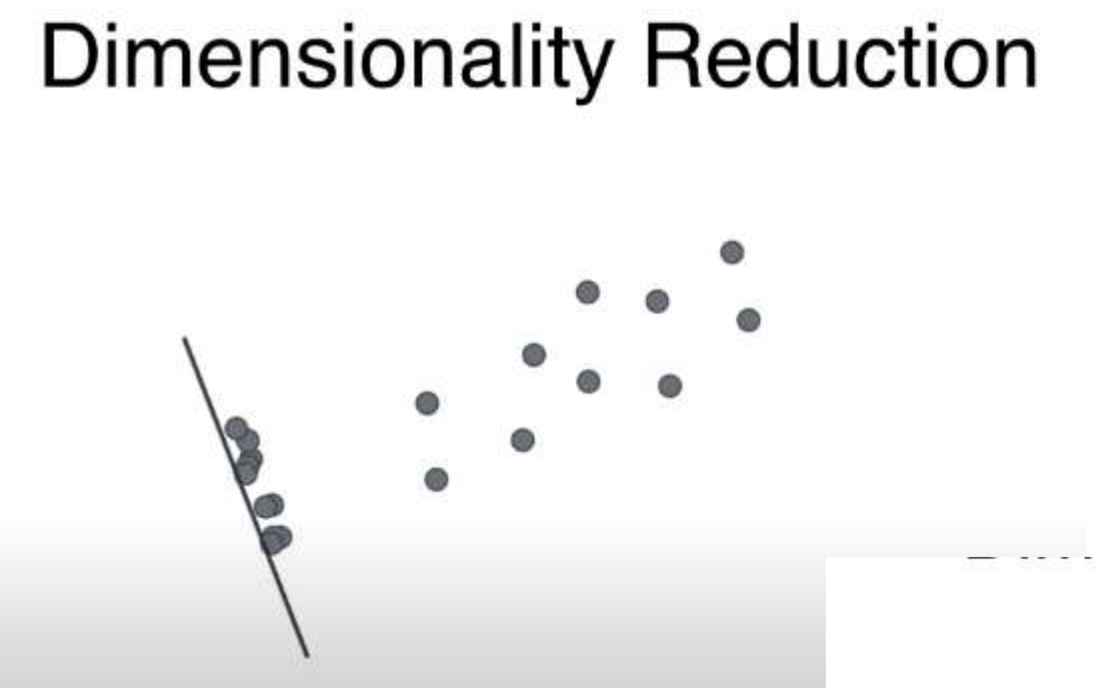


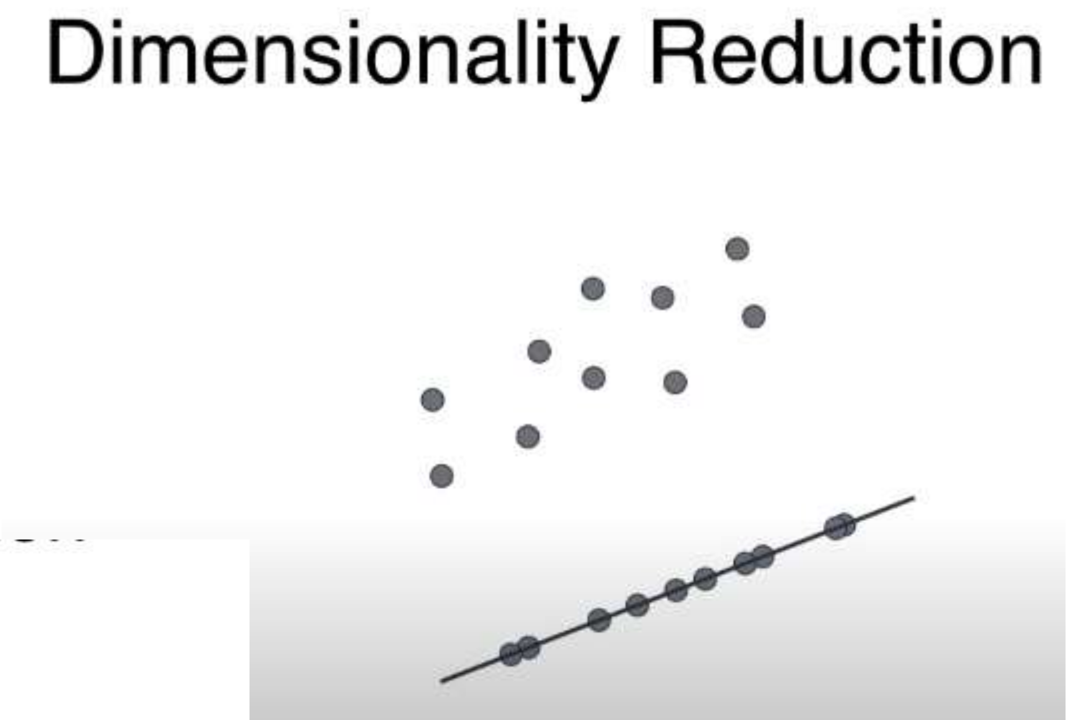
- In previous slide, the last picture gives the right angle to take the picture.
- It means, you have to identify a better angle to collect the data without loosing much information.
- The angle shown in the last picture will capture all the faces, without much overlapping and without loosing information.

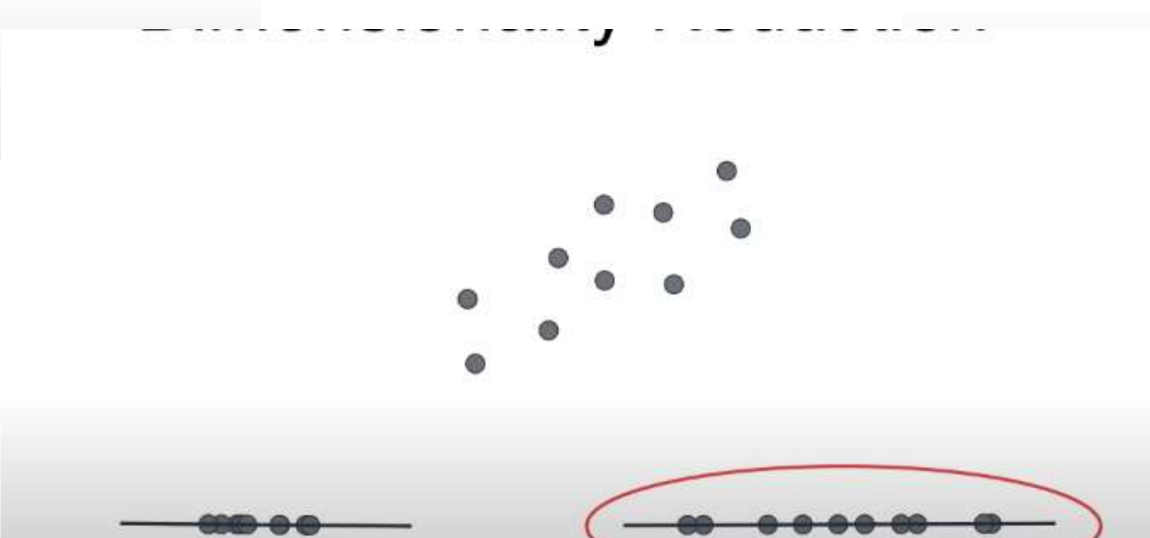
In this example the second one is the best angle to project

https://www.youtube.com/watch?v=g-Hb26agBFg **(reference video)**

https://www.youtube.com/watch?v=MLaJbA82nzk



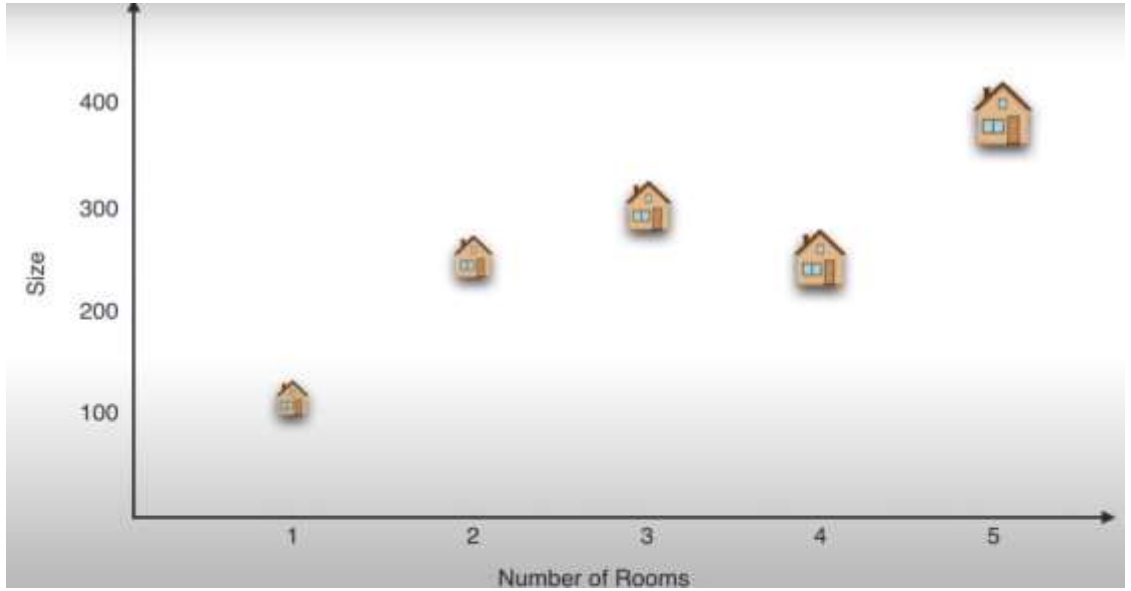




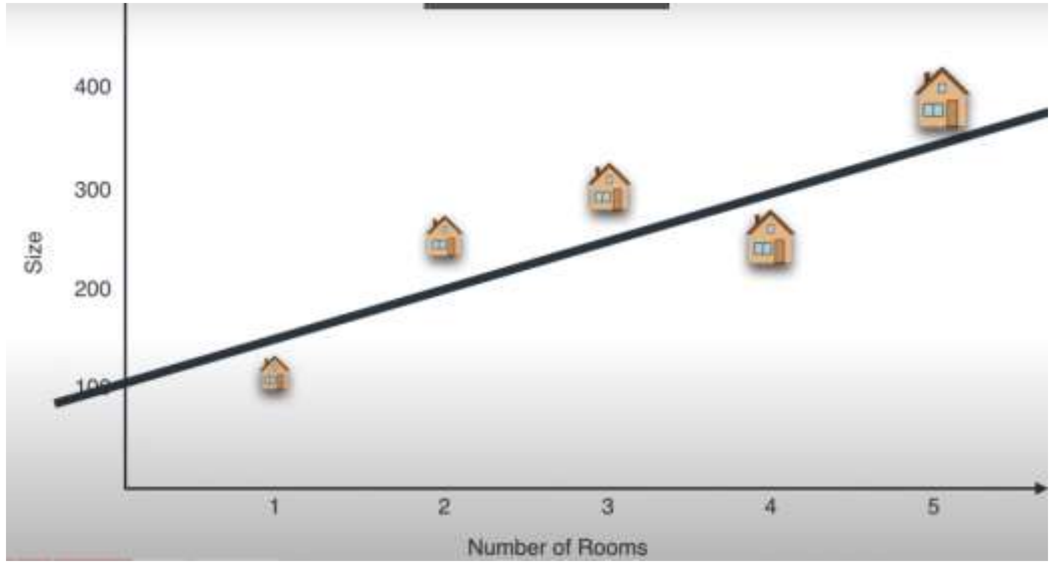
Housing Example: More rooms..more the size

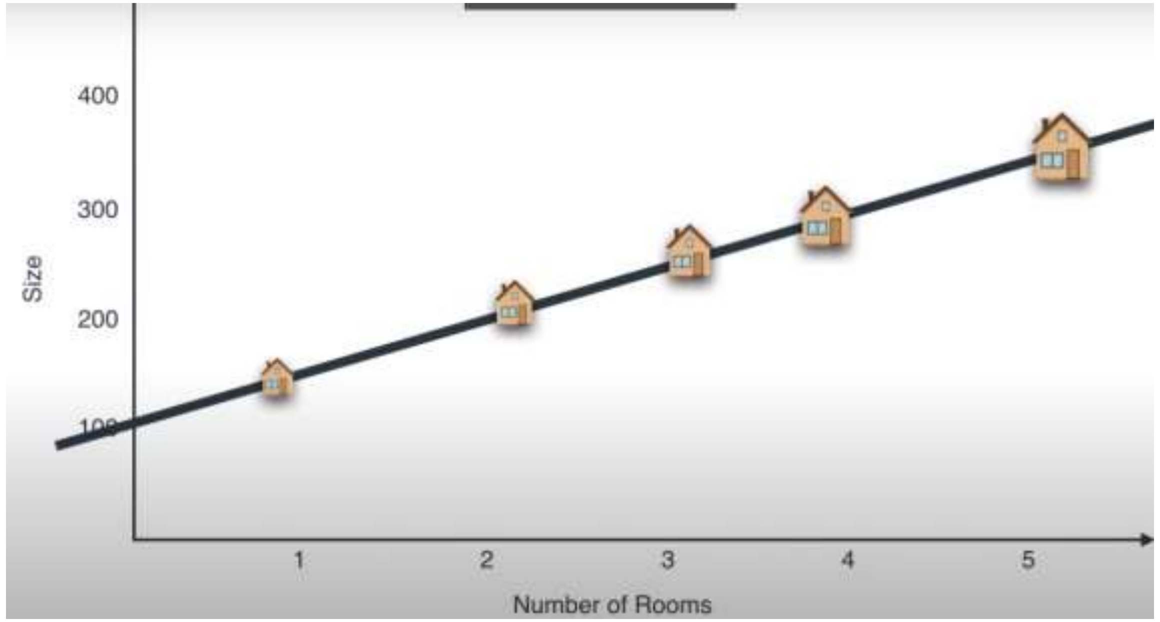
Housing Data

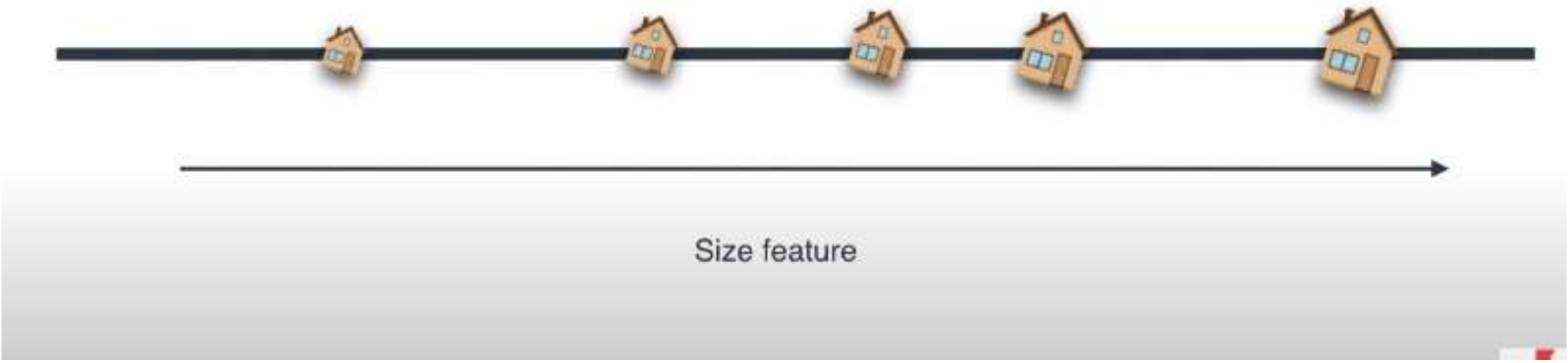
Size
Number of rooms ⟶ Size feature



Two dimension is reduced to single dimension







- PCA is a method of dimensionality reduction.
- Example shows how to convert a two dimension to one dimension.

X	Y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

How to compute PCA?

- Consider the Samples given in the table (10 Samples).
- Compute the mean of X and mean of Y independently. Similar computation has to be done for each features. (In this example only two features).
- Mean of X = 1.81 and Mean of Y = 1.91

Next Step is to compute Co-Variance Matrix.

- Covariance between (x, y) is computed as given below:

$$Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

- The following covariance Matrix to be computed is:

$$C = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix}$$

Covariance between (x and x)

X	Y	(X- Mean(X))		(x-mean(x) * (x-Mean(x)
2.5	2.4	0.69		0.476
0.5	0.7	-1.31		1.716
2.2	2.9	0.39		0.152
1.9	2.2	0.09		0.008
3.1	3	1.29		1.664
2.3	2.7	0.49		0.24
2	1.6	0.19		0.036
1	1.1	-0.81		0.656
1.5	1.6	-0.31		0.096
1.1	0.9	-0.71		0.504
			Total=	5.549
			Total/9	0.617

- Similarly compute co variance between (x,y),(y,x) and (y,y).
- Computed Co-Variance matrix is given in next slide

Final co-variance matrix

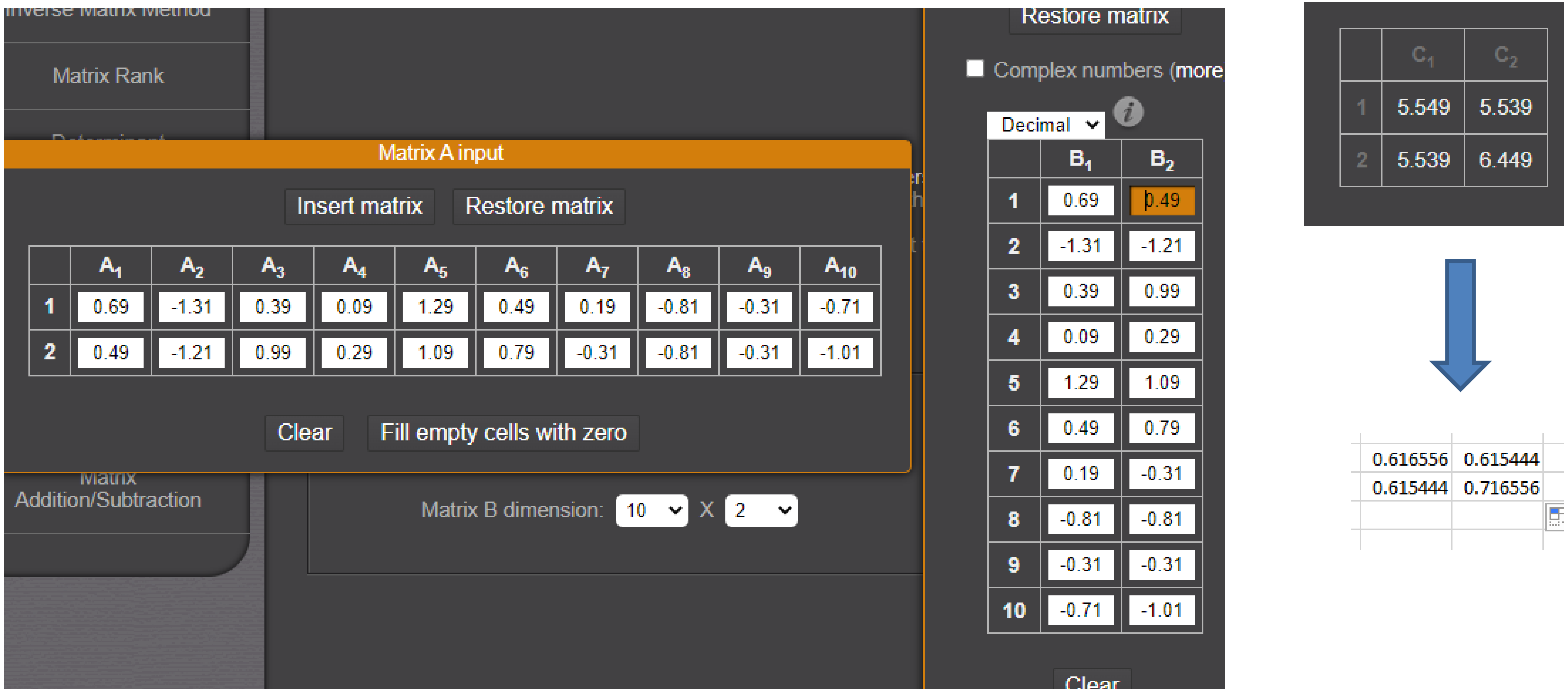
$$C = \begin{bmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{bmatrix}$$

$$= \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix}$$

Alternate Method to compute Co-variance matrix

1	Original Data			Mean Centered Data		
2	2.5	2.4		0.69	0.49	
3	0.5	0.7		-1.31	-1.21	
4	2.2	2.9		0.39	0.99	
5	1.9	2.2		0.09	0.29	
6	3.1	3		1.29	1.09	
7	2.3	2.7		0.49	0.79	
8	2	1.6		0.19	-0.31	
9	1	1.1		-0.81	-0.81	
10	1.5	1.6		-0.31	-0.31	
11	1.1	0.9		-0.71	-1.01	
12						
13	1.81	1.91				
14	(mean of X)	Mean of Y				

Consider Mean centered Matrix as A and now compute Transpose of A * A to get the Covariance matrix: Divide the resultant matrix by (n-1)



Next Step is to Compute Eigen Values using the Co-variance matrix

If A is the given matrix (in this case co-variance matrix)

We can calculate eigenvalues from the following equation:

$$|A - \lambda I| = 0$$

Where A is the given matrix

λ is the eigen value

I is the identity Matrix

$|A - \lambda I| = 0$

$$\Rightarrow \left| \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\Rightarrow \left| \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$$\Rightarrow \left| \begin{bmatrix} 0.6165-\lambda & 0.6154 \\ 0.6154 & 0.7165-\lambda \end{bmatrix} \right| = 0$$

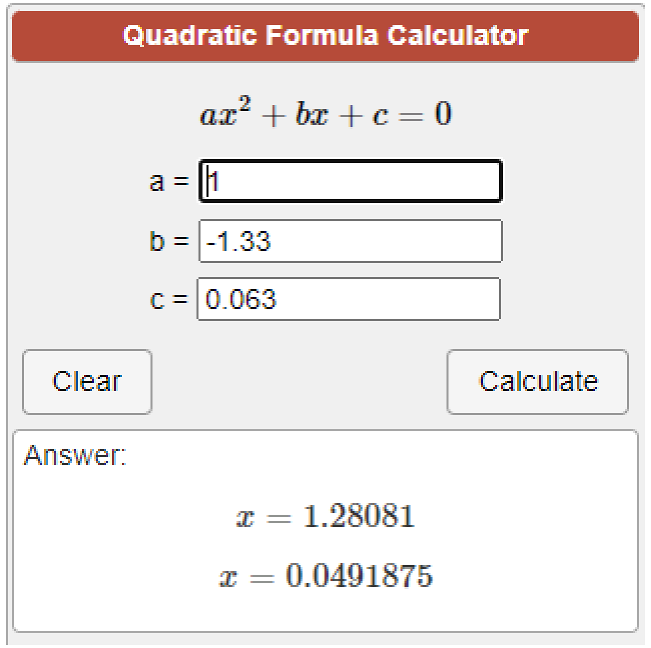
Determinant computation and finally Eigen values

$$\left[(0.6165-\lambda)*(0.7165-\lambda)-(0.6154)*(0.6154)\right]=0$$

$$\Rightarrow (0.6165*0.7165)-(0.6165\lambda)-(0.7165\lambda)+\lambda^2-(0.6154)*(0.6154)=0$$

$$\Rightarrow \boxed{\lambda^2-1.333\lambda+0.0630=0}$$

$$\boxed{\begin{aligned} a &= 1 \\ b &= -1.33 \\ c &= 0.0630 \end{aligned}}$$



$$\boxed{\begin{aligned} \lambda_1 &= 1.2840 \\ \lambda_2 &= 0.490 \end{aligned}}$$

- Compute Eigen vector for each of the eigen value.

$$\boxed{CV = \lambda V}$$

- Consider the first eigen value **λ1** = 1.284
- C is the covariance matrix
- V is the eigen vector to be computed.

$$\begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 1.2840 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$\begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 1.9840 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$0.6165x_1 + 0.6154y_1 = 1.9840x_1$$

$$0.6154x_1 + 0.7169y_1 = 1.9840y_1$$

$$0.6154y_1 = 1.9840x_1 - 0.6165x_1$$

$$0.6154y_1 = 0.6675x_1$$

$0.6675x_1 = 0.6154y_1$

$x_1 = \frac{0.6154}{0.6675} y_1$

$\boxed{x_1 = 0.9219\, y_1}$

$y_1 = 1$

$x_1 = 0.9219$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0.9219 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.92/p \\ 1/p \end{bmatrix}$$

$p = \sqrt{(0.9219)^2 + 1^2}$

$\boxed{p = 1.360}$

$$= \begin{bmatrix} 0.6952 \\ 0.7541 \end{bmatrix}$$

Now convert the two dimension data to single dimension

Sample 1 :- $x = 2.5$ | $\bar{x} = 1.81$

$y = 2.4$ | $\bar{y} = 1.91$

$$= e_1^T * \begin{bmatrix} x - mean(x) \\ y - mean(y) \end{bmatrix} \Rightarrow \text{new transformed data}$$

Final step

$\begin{bmatrix} 0.6952 & 0.7541 \end{bmatrix} \begin{bmatrix} 0.69 \\ 0.49 \end{bmatrix}$

$\Rightarrow 0.4796 + 0.3695$

$\Rightarrow 0.8491$

- Compute Eigen vector for the second eigen value.

$$\boxed{CV = \lambda V}$$

- Consider the first eigen value **λ2** = 0.0490
- C is the covariance matrix
- V is the eigen vector to be computed.

$$\begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} = 0.0490 \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}$$

- Using this we can have two linear equation:

$$\begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} = 0.0490 \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}$$

$$0.6165\,X_1 + 0.6154\,Y_1 = 0.0490\,X_1$$
$$0.6154\,X_1 + 0.7165\,Y_1 = 0.0490\,Y_1$$

$$0.5674\,X_1 = -0.6154\,Y_1$$
$$0.6154\,X_1 = -0.6674\,Y_1$$

- Use any one of the following equation… final result remains same.

$$0.5674\, X_1 = -0.6154\, Y_1$$

$$0.6154\, X_1 = -0.6674\, Y_1$$

- 0.5674 x1 = -0.6154 y1
- Divide both side by 0.5674.
- You will get : x1 = -1.0845 y1

- **$x1 = -1.0845\ y1$**

- **If $y1=1$, then $x1$ will be -1.0845**

- So in that case $(x1, y1)$ will be $(-1.0845,1)$. This will be the initial eigen vector. Needs normalization to get the final value.

- To normalize, take square-root of sum of square of each eigen vector values, and consider this as ‘x’
- Finally divide each eigen vector values by ‘x’ to get the final eigen vector.

eigen vectors are generated for the eigen value : 0.490

$$X_1 = -1.0845\, Y_1$$

$$\begin{bmatrix} -1.0845 \\ 1 \end{bmatrix} = 1.17614 + 1 = \sqrt{2.17614} = 1.47517$$

$$\Rightarrow \begin{bmatrix} -0.7351 \\ 0.6778 \end{bmatrix}$$

$$X_2 = 0.92194\, Y_2$$

$$\begin{bmatrix} 0.92194 \\ 1 \end{bmatrix} = 0.8499 + 1 = \sqrt{1.8499} = 1.3601$$

$$\Rightarrow \begin{bmatrix} 0.6778 \\ 0.7351 \end{bmatrix}$$

Describe the algorithm with an example:

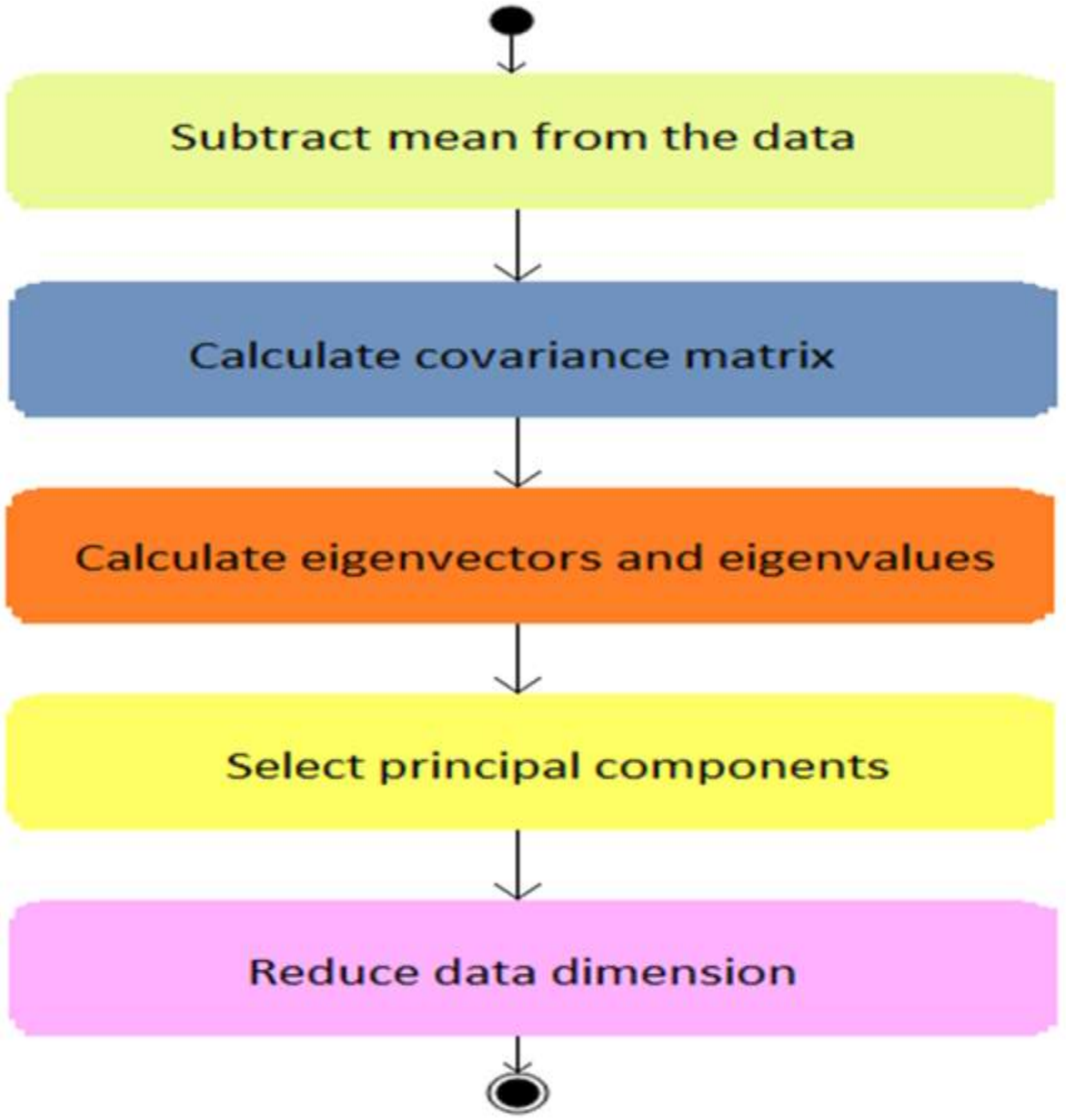
- Consider a 2-D dataset
- Cl =X1 =(x1,x2) ={(4,1),(2,4),(2,3),(3,6), (4,4)}
- C2=X2=(x1,x2) = {(9,10),(6,8),(9,5),(8,7),(10,8)}

PCA

Theory – Algorithms – steps explained

Steps/ Functions to perform PCA

- Subtract mean.
- Calculate the covariance matrix.
- Calculate eigenvectors and eigenvalues.
- Select principal components.
- Reduce the data dimension.



- Principal components is a form of multivariate statistical analysis and is one method of studying the correlation or covariance structure in a set of measurements on m variables for n observations.

- Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

- Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

- So to sum up, the idea of PCA is simple — reduce the number of variables of a data set, while preserving as much information as possible.

- What do the covariances that we have as entries of the matrix tell us about the correlations between the variables?
- It’s actually the sign of the covariance that matters

- if positive then : the two variables increase or decrease together (correlated)

- if negative then : One increases when the other decreases (Inversely correlated)

- Now, that we know that the covariance matrix is not more than a table that summaries the correlations between all the possible pairs of variables, let’s move to the next step.

Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the principal components of the data.

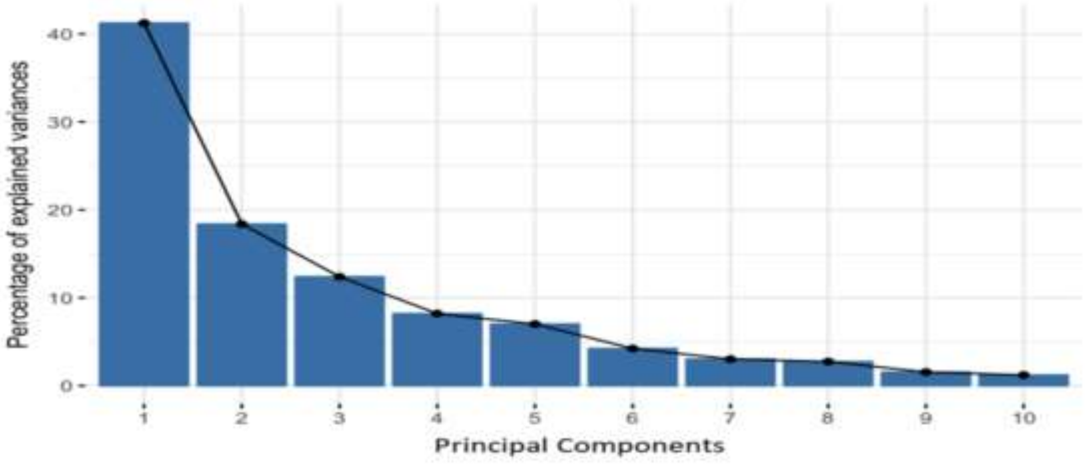
Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables.

These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components.

So, the idea is 10-dimensional data gives you 10 principal components, but PCA tries to put maximum possible information in the first component.

Then maximum remaining information in the second and so on, until having something like shown in the scree plot below.

- As there are as many principal components as there are variables in the data, principal components are constructed in such a manner that the first principal component accounts for the largest possible variance in the data set.



- Organizing information in principal components this way, will allow you to reduce dimensionality without losing much information, and this by discarding the components with low information and considering the remaining components as your new variables.

- An important thing to realize here is that, the principal components are less interpretable and don’t have any real meaning since they are constructed as linear combinations of the initial variables.

Characteristic Polynomial and characteristic equation
and
Eigen Values and Eigen Vectors

Computation for 2x2 and 3x3 Square Matrix

Eigen Values and Eigen Vectors

Definition
Let A be an $n \times n$ matrix. A *scalar* λ is called an **eigenvalue** of A if there exists a nonzero vector $\mathbf{x}$ in $\mathbf{R}^n$ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The vector $\mathbf{x}$ is called an **eigenvector** corresponding to λ.

The eigenvectors **x** and eigenvalues λ of a matrix A satisfy

$$\mathbf{Ax = \lambda x}$$

If A is an n x n matrix, then **x** is an n x 1 vector, and λ is a constant.

The equation can be rewritten as $(A - \lambda I)\,\mathbf{x} = 0$, where I is the n x n identity matrix.

Solving the equation $|A - \lambda I_n| = 0$ for λ leads to all the eigenvalues of A.

On expending the determinant $|A - \lambda I_n|$, we get a polynomial in λ. This polynomial is called the **characteristic polynomial** of A.

The equation $|A - \lambda I_n| = 0$ is called the **characteristic equation** of A.

2 X 2 Example : Compute Eigen Values

$A = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$ so $A - \lambda I = \begin{bmatrix} 1 - \lambda & -2 \\ 3 & -4 - \lambda \end{bmatrix}$

$$\det(A - \lambda I) = (1 - \lambda)(-4 - \lambda) - (3)(-2)$$
$$= \lambda^2 + 3\lambda + 2$$

Set $\lambda^2 + 3\lambda + 2$ to 0

Then = $\lambda = (-3 \;\; +/- \;\; \text{sqrt}(9-8))/2$

So the two values of λ are -1 and -2.

Example 1: Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$$

Solution Let us first derive the characteristic polynomial of A.

We get

$$A - \lambda I_2 = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -4-\lambda & -6 \\ 3 & 5-\lambda \end{bmatrix}$$

$$|A - \lambda I_2| = (-4-\lambda)(5-\lambda) + 18 = \lambda^2 - \lambda - 2$$

We now solve the characteristic equation of A.

$$\lambda^2 - \lambda - 2 = 0 \Rightarrow (\lambda - 2)(\lambda + 1) = 0 \Rightarrow \lambda = 2 \text{ or } -1$$

The eigenvalues of A are 2 and –1.

The corresponding eigenvectors are found by using these values of λ in the equation $(A - \lambda I_2)\mathbf{x} = \mathbf{0}$.

There are many eigenvectors corresponding to each eigenvalue.

For $\underline{\lambda = 2}$

We solve the equation $(A - 2I_2)\mathbf{x} = \mathbf{0}$ for $\mathbf{x}$.

The matrix $(A - 2I_2)$ is obtained by subtracting 2 from the diagonal elements of A.

We get

$$\begin{bmatrix} -6 & -6 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}$$

This leads to the system of equations

$$\begin{aligned} -6x_1 - 6x_2 &= 0 \\ 3x_1 + 3x_2 &= 0 \end{aligned}$$

giving $x_1 = -x_2$. The solutions to this system of equations are $x_1 = -r$, $x_2 = r$, where r is a scalar. Thus the eigenvectors of A corresponding to $\lambda = 2$ are nonzero vectors of the form

$$\mathbf{v}_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = r \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

For $\underline{\lambda = -1}$

We solve the equation $(A + 1I_2)x = 0$ for x.

The matrix $(A + 1I_2)$ is obtained by adding 1 to the diagonal elements of A. We get

$$\begin{bmatrix} -3 & -6 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}$$

This leads to the system of equations

$$\begin{aligned} -3x_1 - 6x_2 &= 0 \\ 3x_1 + 6x_2 &= 0 \end{aligned}$$

Thus $x_1 = -2x_2$. The solutions to this system of equations are $x_1 = -2s$ and $x_2 = s$, where s is a scalar. Thus the **eigenvectors** of A corresponding to $\lambda = -1$ are nonzero vectors of the form

$$\mathbf{v}_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2\begin{bmatrix} -2 \\ 1 \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

- **Example 2** Calculate the eigenvalue equation and eigenvalues for the following matrix –

- $$\begin{matrix} 1 & 0 & 0 \\ 0 & -1 & 2 \\ 2 & 0 & 0 \end{matrix}$$

Solution : Let A = $\begin{matrix} 1 & 0 & 0 \\ 0 & -1 & 2 \\ 2 & 0 & 0 \end{matrix}$ and $A–λI$ = $\begin{matrix} 1-\lambda & 0 & 0 \\ 0 & -1-\lambda & 2 \\ 2 & 0 & 0-\lambda \end{matrix}$

We can calculate eigenvalues from the following equation:

|A- λI| = 0 $(1 -\lambda)$ [(- 1 –λ)(-λ) - 0] – 0 + 0 = 0

λ (1 - λ) (1 + λ) = 0

From this equation, we are able to estimate eigenvalues which are –

λ = 0, 1, -1.

Example2 : Eigenvalues 3x3 Matrix

Find the eigenvalues of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & 2 \\ 0 & 0 & 7 \end{bmatrix}$$

Solution:

$$A - \lambda I_n = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & 2 \\ 0 & 0 & 7 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\lambda & 2 & 3 \\ 0 & -4-\lambda & 2 \\ 0 & 0 & 7-\lambda \end{bmatrix}$$

$$\det(A - \lambda I_n) = 0 \rightarrow \det \begin{bmatrix} 1-\lambda & 2 & 3 \\ 0 & -4-\lambda & 2 \\ 0 & 0 & 7-\lambda \end{bmatrix} = 0$$

$$(1-\lambda)(-4-\lambda)(7-\lambda) = 0$$

$$\lambda = \{1, \quad -4, \quad 7\}$$

Example 3: Eigenvalues and Eigenvectors

Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

Solution The matrix $A - \lambda I_3$ is obtained by subtracting λ from the diagonal elements of A.Thus

$$A - \lambda I_3 = \begin{bmatrix} 5-\lambda & 4 & 2 \\ 4 & 5-\lambda & 2 \\ 2 & 2 & 2-\lambda \end{bmatrix}$$

The characteristic polynomial of A is $|A - \lambda I_3|$. Using row and column operations to simplify determinants, we get

Alternate Solution

$|A - \lambda I_3| = 0$

$$A = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

λ = Eigen values

For 3×3 matrix Eigen values can be computed using the following equations

$$\boxed{\lambda^3 - S_1\lambda^2 + S_2\lambda - S_3 = 0}$$

S_1 = Sum of principal diagonal elements

S_2 = Sum of minors of principal diagonal elements

S_3 = Determinant of A

$$A = \begin{array}{c} \begin{array}{ccc} + & - & + \end{array} \\ \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} \end{array}$$

$$S_1 = 5 + 5 + 2$$

$$S_2 = \begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix} + \begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix} + 2 \begin{vmatrix} 5 & 4 \\ 4 & 5 \end{vmatrix}$$

$$= 6 + 6$$

$$S_2 = \begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix} + \begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix} + \begin{vmatrix} 5 & 4 \\ 4 & 5 \end{vmatrix}$$

$$= (10 - 4) + (10 - 4) + (25 - 16)$$

$$= 6 + 6 + 9$$

$$= 21$$

$$\boxed{S_2 = 21}$$

$S_3 =$ Determinant of A

$$S_3 = |A| = \left| \begin{array}{ccc} + & - & + \\ \end{array} \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} \right|$$

$$= 5 * \begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix} - 4 * \begin{vmatrix} 4 & 2 \\ 2 & 2 \end{vmatrix} + 2 * \begin{vmatrix} 4 & 5 \\ 2 & 2 \end{vmatrix}$$

$$= 5 * (10 - 4) - 4 * (8 - 4) + 2 * (8 - 10)$$

$$= 5 * 6 - 4 * 4 + 2(-2)$$

$$= 30 - 16 - 4$$

$$\boxed{S_3 = 10}$$

$\lambda^3 - [illegible]\lambda^2 + [illegible]\lambda - [illegible] = 0$

$\lambda^3 - [illegible]\lambda^2 + [illegible]\lambda - 10 = 0$

Eigen values will be factors of [illegible]

Factors of [illegible] = 1, 2, 5, 10

check for both +ve as well as −ve

It's [illegible] not [illegible] none

$\lambda = \{10 \ \& \ 1\}$

$\lambda_1 = 10$

$\lambda_2 = 1$

Eigen vector:

$$(A - \lambda I_3)X = 0$$

$$A = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} \qquad \lambda = \{10, 1\} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\lambda = 10$

$$\left(\begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} - 10\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$\Rightarrow \left(\begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} -5 & 4 & 2 \\ 4 & -5 & 2 \\ 2 & 2 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$$-5x_1 + 4x_2 + 2x_3 = 0$$
$$4x_1 - 5x_2 + 2x_3 = 0$$
$$2x_1 + 2x_2 + 8x_3 = 0$$

Consider any two equations

$$-5x_1 + 4x_2 + 3x_3 = 0$$
$$4x_1 - 5x_2 + 3x_3 = 0$$

$$\frac{x_1}{(4\times3)-(-15)} = \frac{-x_2}{-(15)-3} = \frac{x_3}{25-16}$$

$$\frac{x_1}{8+10} = \frac{x_3}{18} = \frac{x_2}{9}$$

$$\frac{x_1}{18} = \frac{x_2}{18} = \frac{x_3}{9} = k$$

$$\Rightarrow \frac{x_1}{2} = \frac{x_2}{2} = \frac{x_3}{1} = 9k = k_1$$

$$\frac{x_1}{2} = \frac{x_2}{2} = \frac{x_3}{1} \; ; \quad x_1 = 2k_1, \; x_2 = 2k_1, \; x_3 = k_1$$

Hence $X_1 = \begin{bmatrix} 2k_1 \\ 2k_1 \\ k_1 \end{bmatrix}$

$$= k_1 \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

- $\lambda_2 = 1$

  Let $\lambda = 1$ in $(A - \lambda I_3)\mathbf{x} = \mathbf{0}$. We get

$$(A - 1I_3)\mathbf{x} = \mathbf{0}$$

$$\begin{bmatrix} 4 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0}$$

The solution to this system of equations can be shown to be $x_1 = -s - t$, $x_2 = s$, and $x_3 = 2t$, where s and t are scalars. Thus the eigenspace of $\lambda_2 = 1$ is the space of vectors of the form.

$$\begin{bmatrix} -s-t \\ s \\ 2t \end{bmatrix}$$

Separating the parameters *s* and *t*, we can write

$$\begin{bmatrix} -s-t \\ s \\ 2t \end{bmatrix} = s\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

Thus the eigenspace of $\lambda = 1$ is a two-dimensional subspace of $\mathbf{R}^3$ with basis

$$\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}\right\}$$

If an eigenvalue occurs as a *k* times repeated root of the characteristic equation, we say that it is of **multiplicity** *k*. Thus λ=10 has multiplicity 1, while λ=1 has multiplicity 2 in this example.

Linear Discriminant Analysis (LDA)

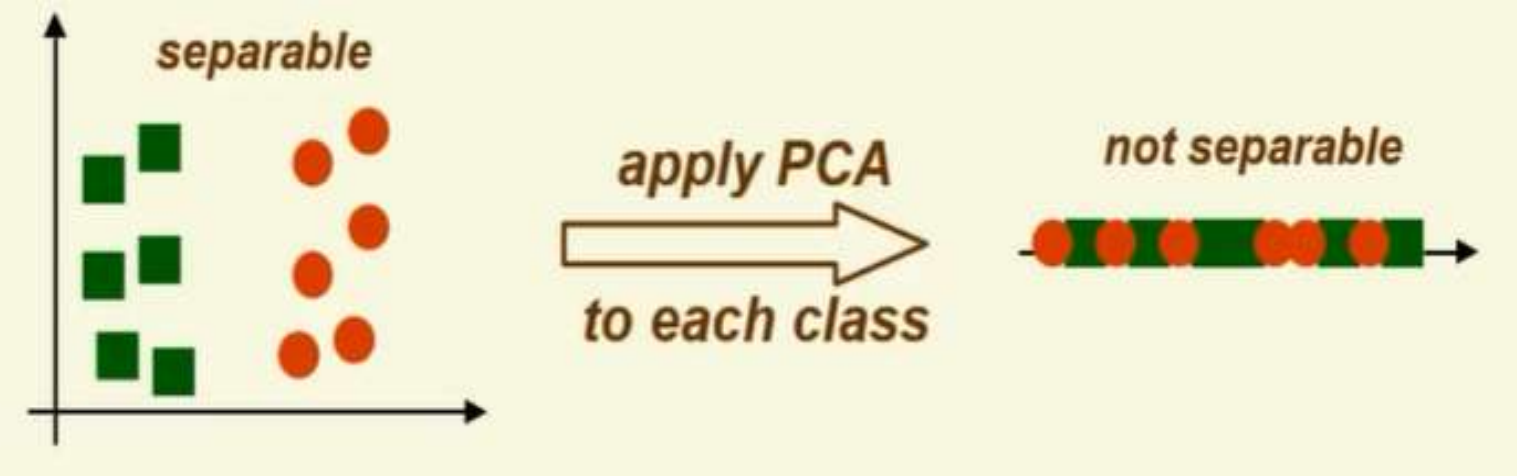
Data representation vs. Data Classification

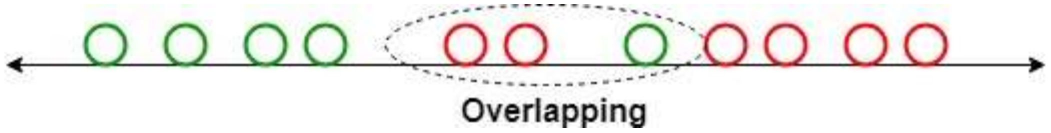
Difference between PCA vs. LDA

- PCA finds the most accurate data representation in a lower dimensional space.
- Projects the data in the directions of maximum variance.
- However the directions of maximum variance may be useless for classification
- In such condition LDA which is also called as Fisher LDA works well.
- LDA is similar to PCA but LDA in addition finds the axis that maximizes the separation between multiple classes.

LDA Algorithm

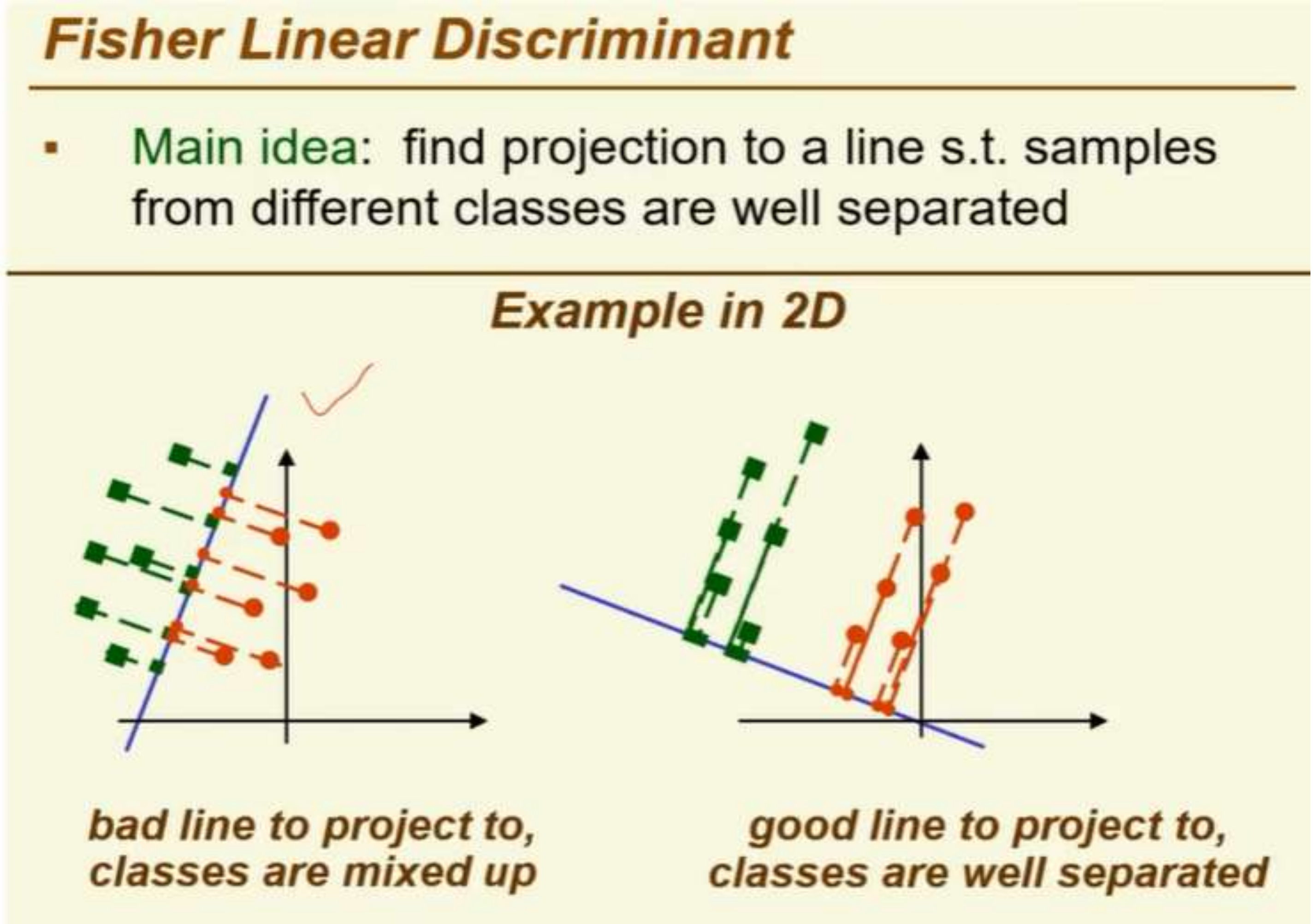
- PCA is good for dimensionality reduction.
- However Figure shows how PCA fails to classify. (because it will try to project this points which maximizes variance and minimizes the error)





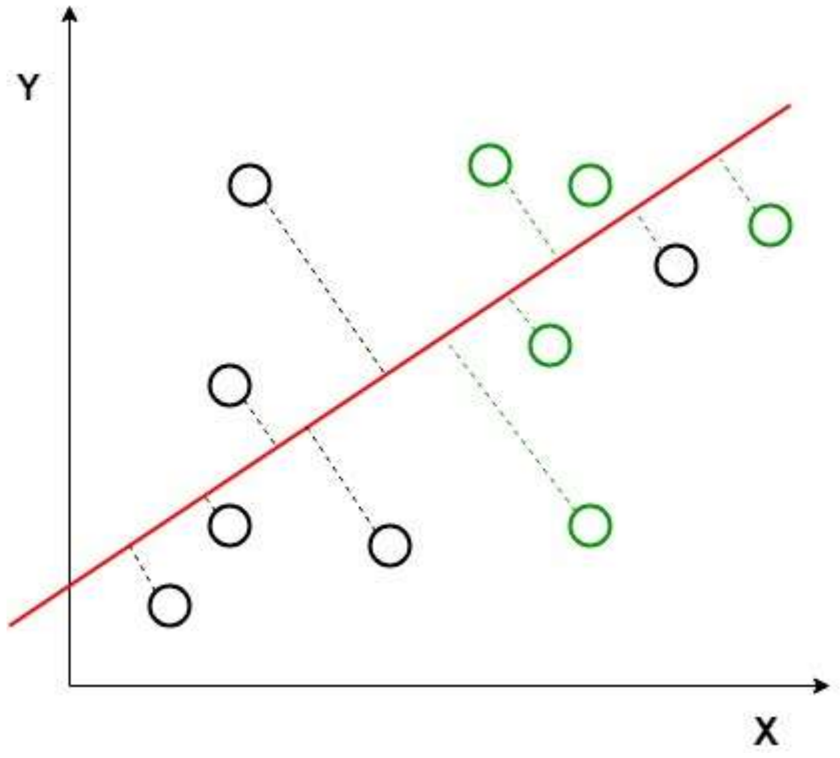
- Fisher Linear Discriminant Project to a line which reduces the dimension and also maintains the class discriminating information.

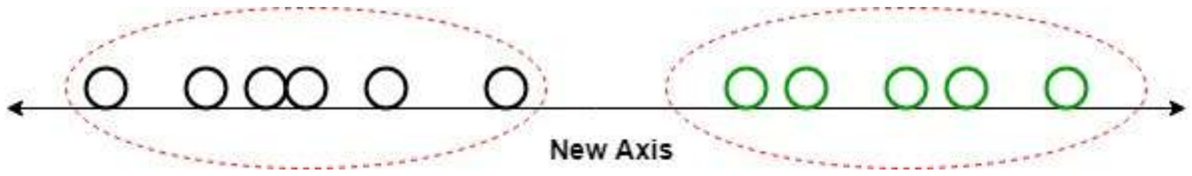
Projection of the samples in the second picture is the best:



- Two criteria are used by LDA to create a new axis:

1. Maximize the distance between means of the two classes.
2. Minimize the variation within each class.





Describe the algorithm with an example:

- Consider a 2-D dataset
- Cl =X1 =(x1,x2) ={(4,1),(2,4),(2,3),(3,6), (4,4)}
- C2=X2=(x1,x2) = {(9,10),(6,8),(9,5),(8,7),(10,8)}

Step 1: Compute within class scatter matrix(Sw)

- Sw= = s1+s2

- s1 is the covariance matrix for class 1 and
- s2 is the covariance matrix for class 2.

- Note : Covariance matrix is to be computed on the Mean Cantered data
- For the given example: mean of C1= (3, 3.6) and
- mean of C2=(8,4, 7.6)
- S1=Transpose of mean centred data * Mean centred data

Computed values s1,s2 and Sw

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_W = S_1 + S_2$$

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Step 2: Compute between class scatter Matrix(Sb)

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

- Mean 1 (M1) =(3,3.6)
- Mean 2 (M2)=(8,4,7.6)

- (M1-M2) = (3-8.4, 3.6-7.6) = (-5.4, 4.0)

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$
$$= \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} \begin{pmatrix} -5.4 & -4 \end{pmatrix} = \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{pmatrix}$$

Step 3: Find the best LDA projection vector

- To do this ..compute the Eigen values and eigen vector for the largest eigen value, on the matrix which is the product of : $S_W^{-1} S_B$ = $\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix}$

- In this example, highest eigen value is : 15.65 ($\lambda = 15.65$)

Compute inverse of S_W^{-1}

- = $S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$

S_W^{-1} is found by using the formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

S_w^{-1}

So, $S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$

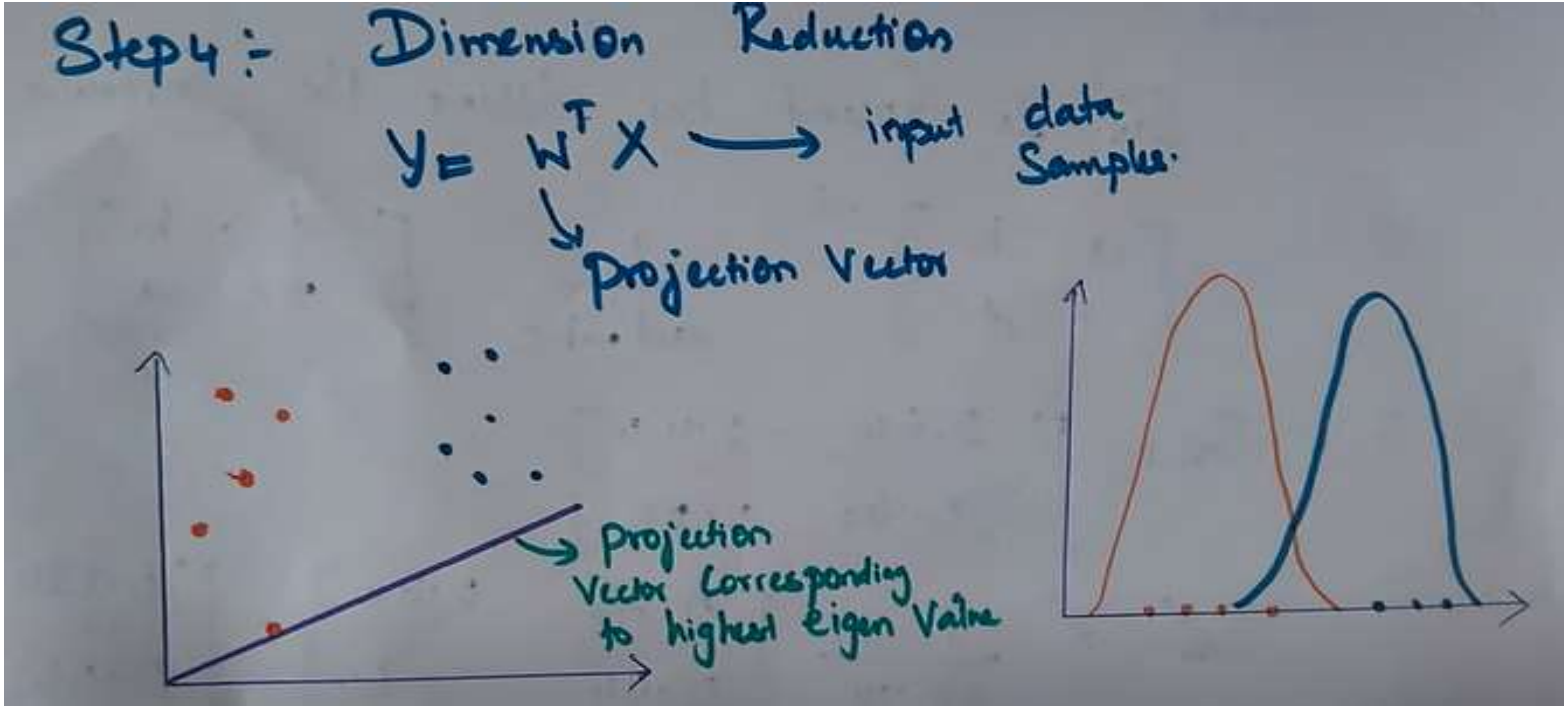
$$S_W^{-1} = \frac{1}{13.74} \begin{bmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{bmatrix} = \begin{bmatrix} 0.384 & 0.032 \\ 0.032 & 0.192 \end{bmatrix}$$

Eigen vector computed for Eigen value: 15.65

$$\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 15.65 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

we get $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$

Step 4: Dimension Reduction



Summary of the Steps

- Step 1 - Computing the within-class and between-class scatter matrices.
- Step 2 - Computing the eigenvectors and their corresponding eigenvalues for the scatter matrices.
- Step 3 - Sorting the eigenvalues and selecting the top k.
- Step 4 - Creating a new matrix that will contain the eigenvectors mapped to the k eigenvalues.
- Step 5 - Obtaining new features by taking the dot product of the data and the matrix from Step 4.

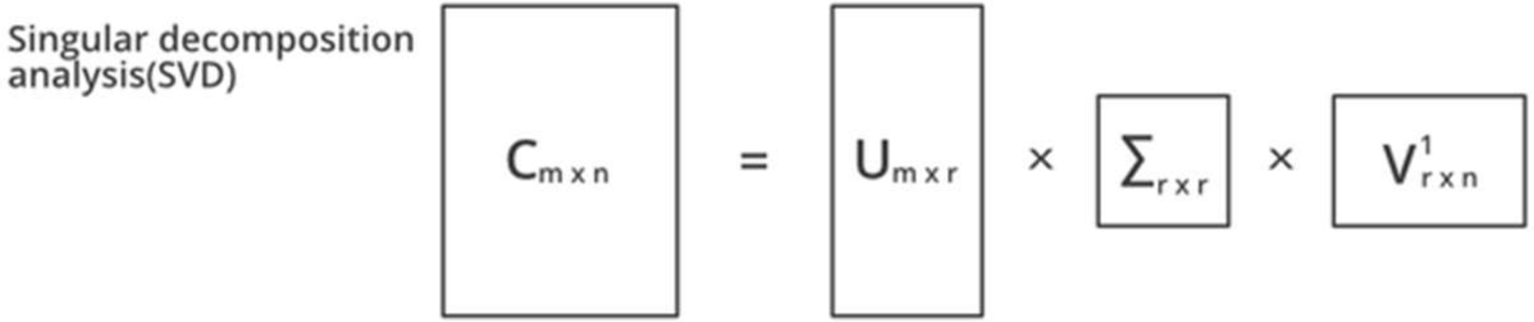
Singular Value Decomposition (SVD)

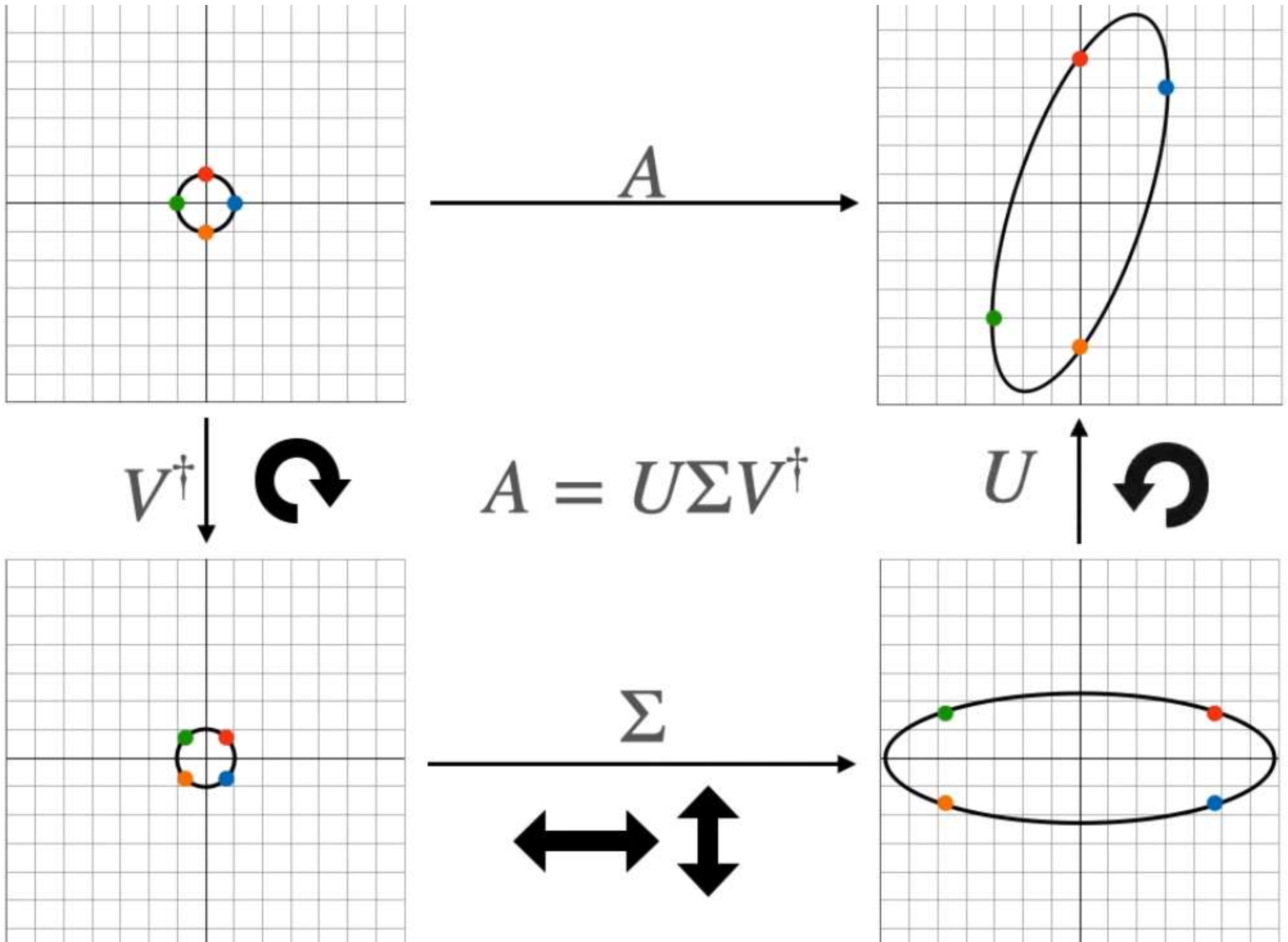
What is singular value decomposition explain with example?

- The singular value decomposition of a matrix A is **the factorization of A into the product of three matrices A = UDV^T where the columns of U and V^T are orthonormal and the matrix D is diagonal with positive real entries**. The SVD is useful in many tasks.
- Calculating the SVD consists of finding the eigenvalues and eigenvectors of AA^T and A^TA.
- The eigenvectors of A^TA make up the columns of V , the eigenvectors of AA^T make up the columns of U.
- Also, the singular values in **S** are square roots of eigenvalues from AA^T or A^TA.
- The singular values are the diagonal entries of the S matrix and are arranged in descending order. The singular values are always real numbers.
- If the matrix A is a real matrix, then U and V are also real.

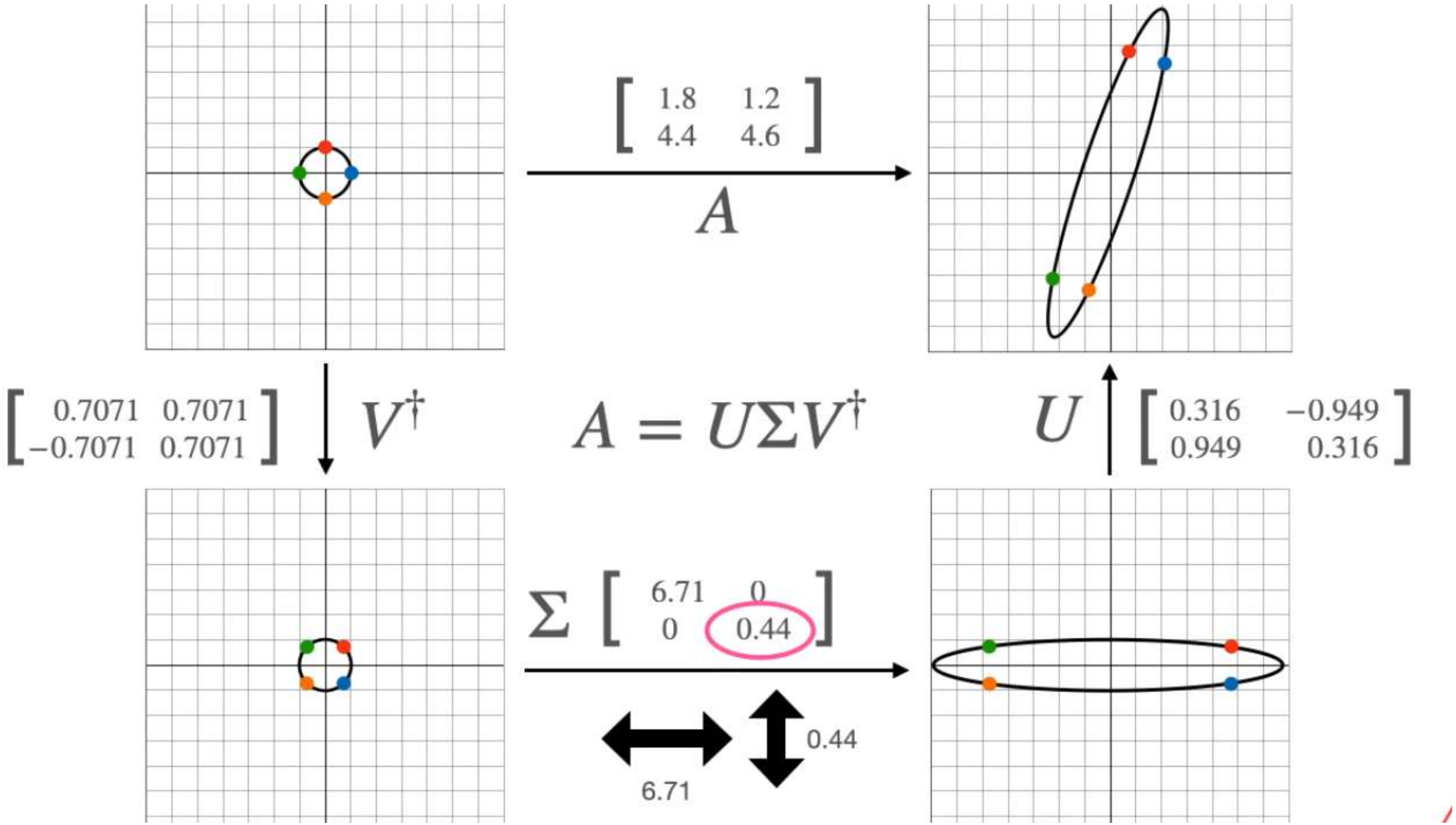
where:

- U: ***mxr*** matrix of the orthonormal eigenvectors of AA^T.
- V^T: transpose of a ***rxn*** matrix containing the orthonormal eigenvectors of A^TA.
- W: a ***rxr*** diagonal matrix of the singular values which are the square roots of the eigenvalues of AA^T and A^TA .





A
$V^\dagger$
$A = U\Sigma V^\dagger$
U
Σ



$\begin{bmatrix} 1.8 & 1.2 \\ 4.4 & 4.6 \end{bmatrix}$
A
$\begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}$
$V^\dagger$
$A = U\Sigma V^\dagger$
U
$\begin{bmatrix} 0.316 & -0.949 \\ 0.949 & 0.316 \end{bmatrix}$
Σ
$\begin{bmatrix} 6.71 & 0 \\ 0 & 0.44 \end{bmatrix}$
6.71
0.44

$C = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$

SVD of $C = U \Sigma V^T$

$C^T C = \begin{pmatrix} 5 & -1 \\ 5 & 7 \end{pmatrix} \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} = \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix}$

COMPUTE EIGEN VALUES

$|C^T C - \lambda I| = \begin{vmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{vmatrix}$

$= \lambda^2 - 100\lambda + 1600$ $\quad a = 1$, $b = 100$, $c = 1600$

$\boxed{\begin{matrix} \lambda_1 = 20 \\ \lambda_2 = 80 \end{matrix}}$

EIGEN VECTORS

$\lambda_1 = 20$

$(C^T C - 20 I)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$v_1 = \begin{pmatrix} -3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}$ $\quad v_2 = \begin{pmatrix} 1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix}$

So $V = \begin{pmatrix} -3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{pmatrix}$

Σ = square roots of eigen values of $C^T C$ in the diagonal position

$$= \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{pmatrix} = \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}$$

$$C = U \Sigma V^T$$

$$CV = U \Sigma V^T V$$

$$\boxed{CV = U\Sigma}$$

$$\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \begin{pmatrix} -3/\sqrt{10} & \frac{1}{\sqrt{10}} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{pmatrix} = U\Sigma$$

$$\Rightarrow \begin{pmatrix} -\sqrt{10} & 2\sqrt{10} \\ \sqrt{10} & 2\sqrt{10} \end{pmatrix} = U \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}$$

Hence

$$\underbrace{\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}}_{} = \underbrace{\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{pmatrix}^T}_{V^T}$$

End of unit 4