Dimensionality Reduction

Some slides were adapted/taken from various sources, including Prof. Andrew Ng's Coursera Lectures, Stanford University, Prof. Kilian Q. Weinberger's lectures on Machine Learning, Cornell University, Prof. Sudeshna Sarkar's Lecture on Machine Learning, IIT Kharagpur, Prof. Bing Liu's lecture, University of Illinois at Chicago (UIC), CS231n: Convolutional Neural Networks for Visual Recognition lectures, Stanford University, Dr. Luis Serrano, Prof. Alexander Ihler and many more. We thankfully acknowledge them. Students are requested to use this material for their study only and **NOT** to distribute it.

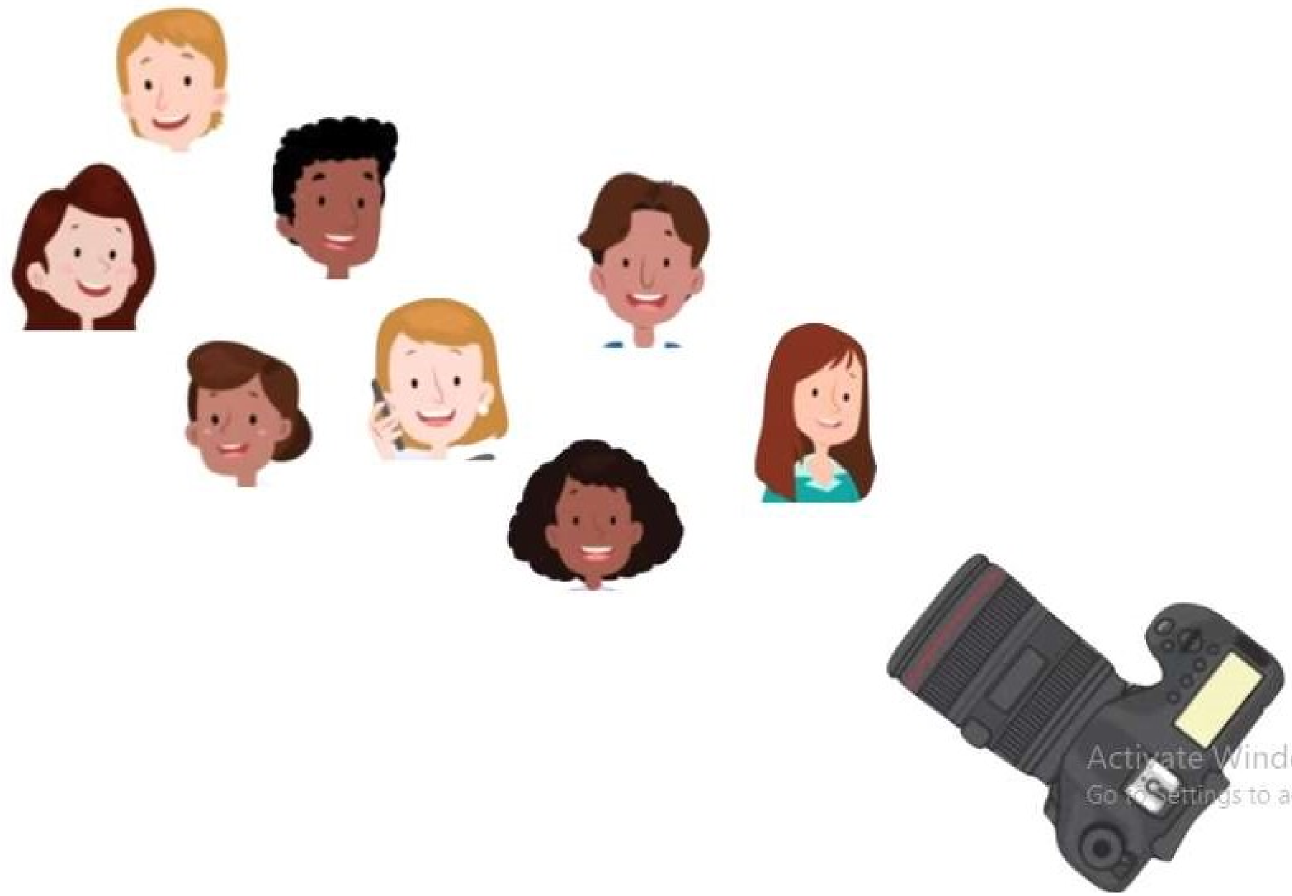
Methods

- PCA (Principal Component Analysis):
  - Find projection that maximize the variance
- ICA (Independent Component Analysis):
  - Very similar to PCA except that it assumes non-Guassian features
- Multidimensional Scaling:
  - Find projection that best preserves inter-point distances
- LDA(Linear Discriminant Analysis):
  - Maximizing the component axes for class-separation
- …
- …

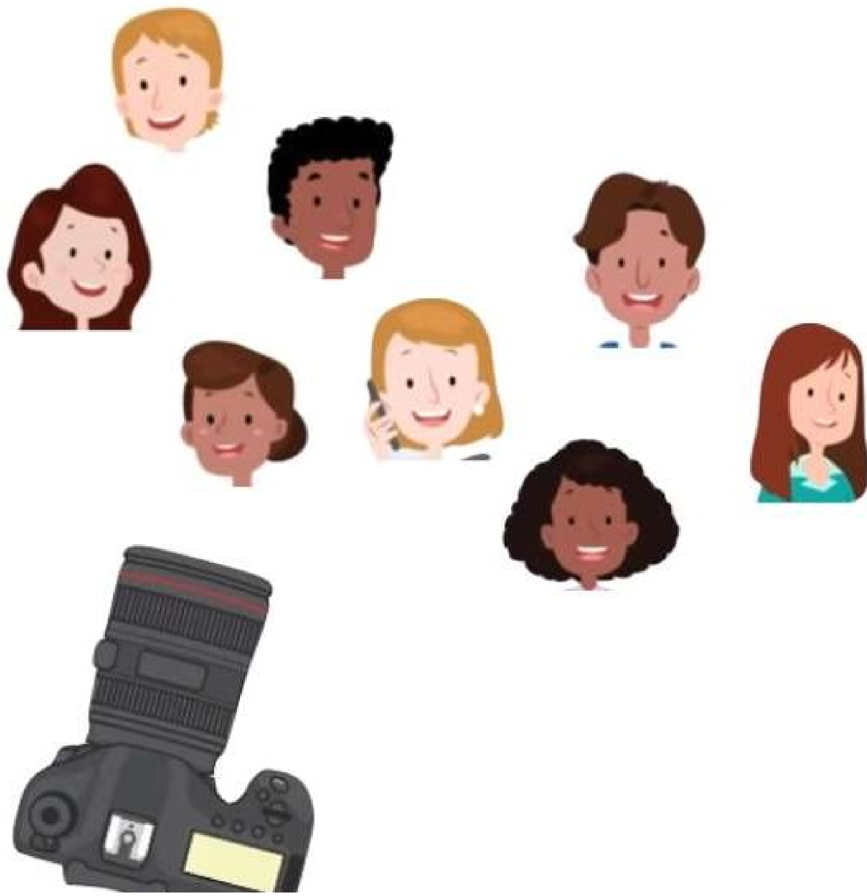
Taking a picture

Taking a picture

Taking a picture



Taking a picture

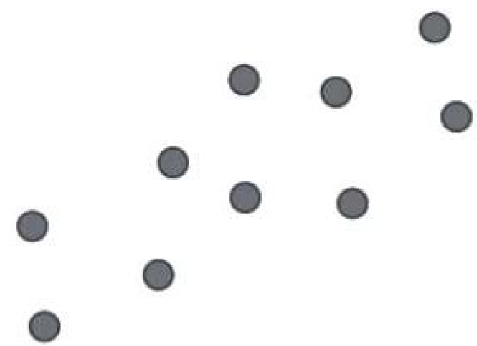


May be Best angle to take the picture

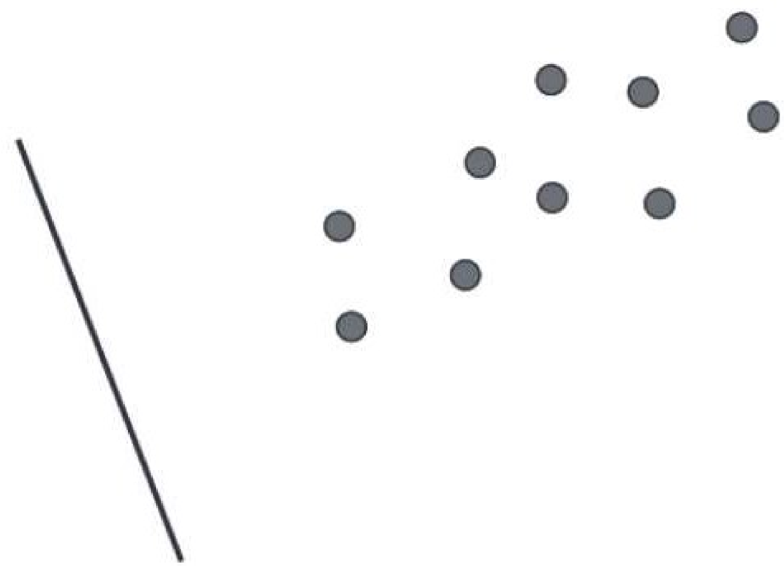
Problem: PCA

- Taking a picture (understanding/representation) of your data.

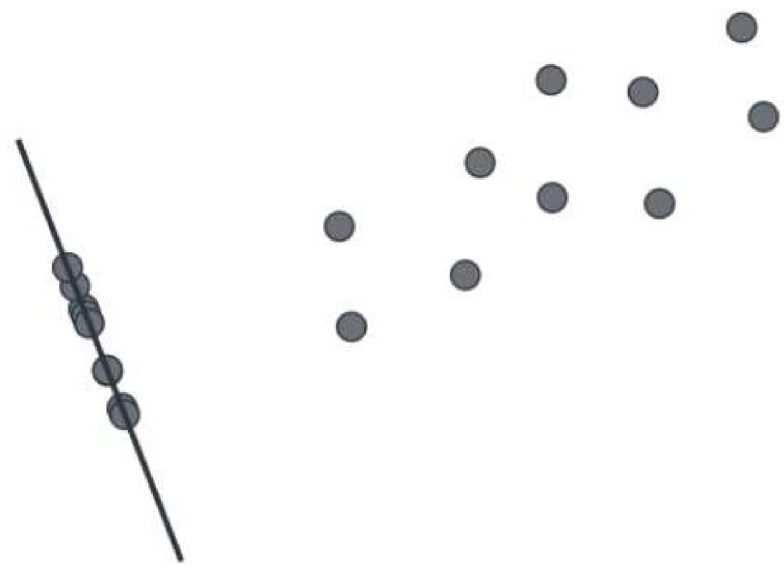
Dimensionality Reduction



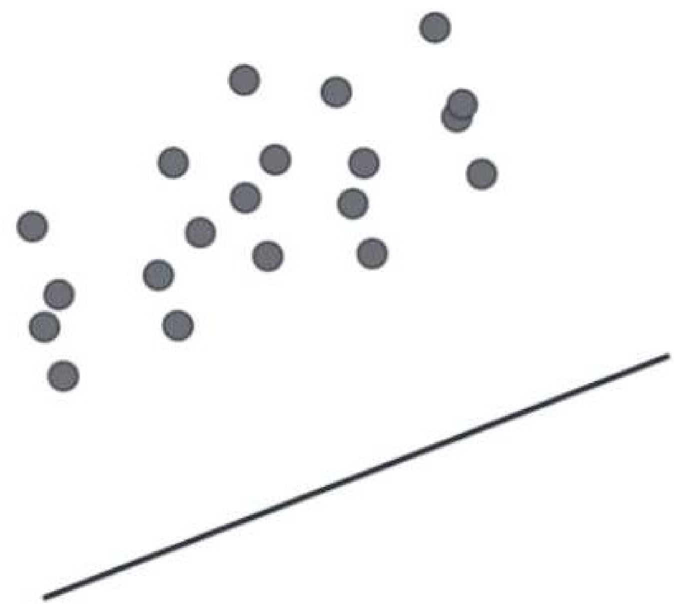
Dimensionality Reduction



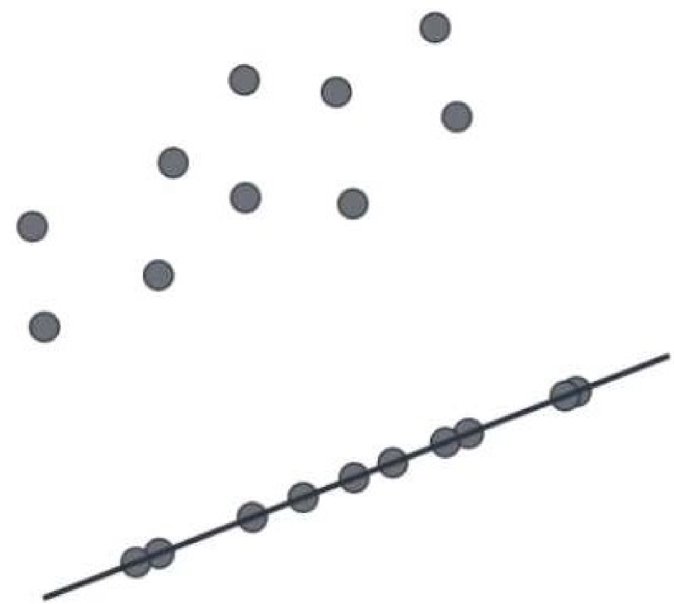
Dimensionality Reduction



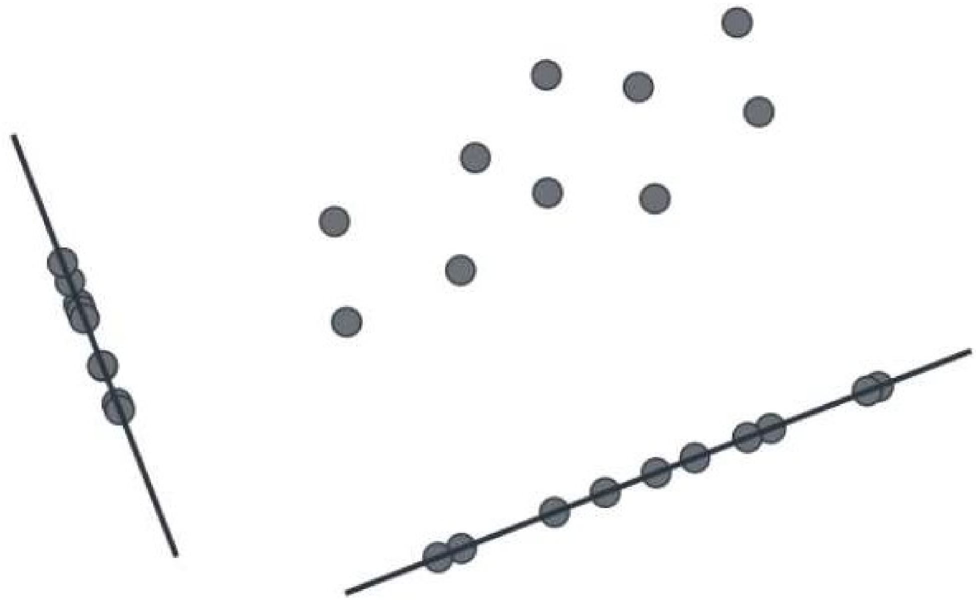
Dimensionality Reduction



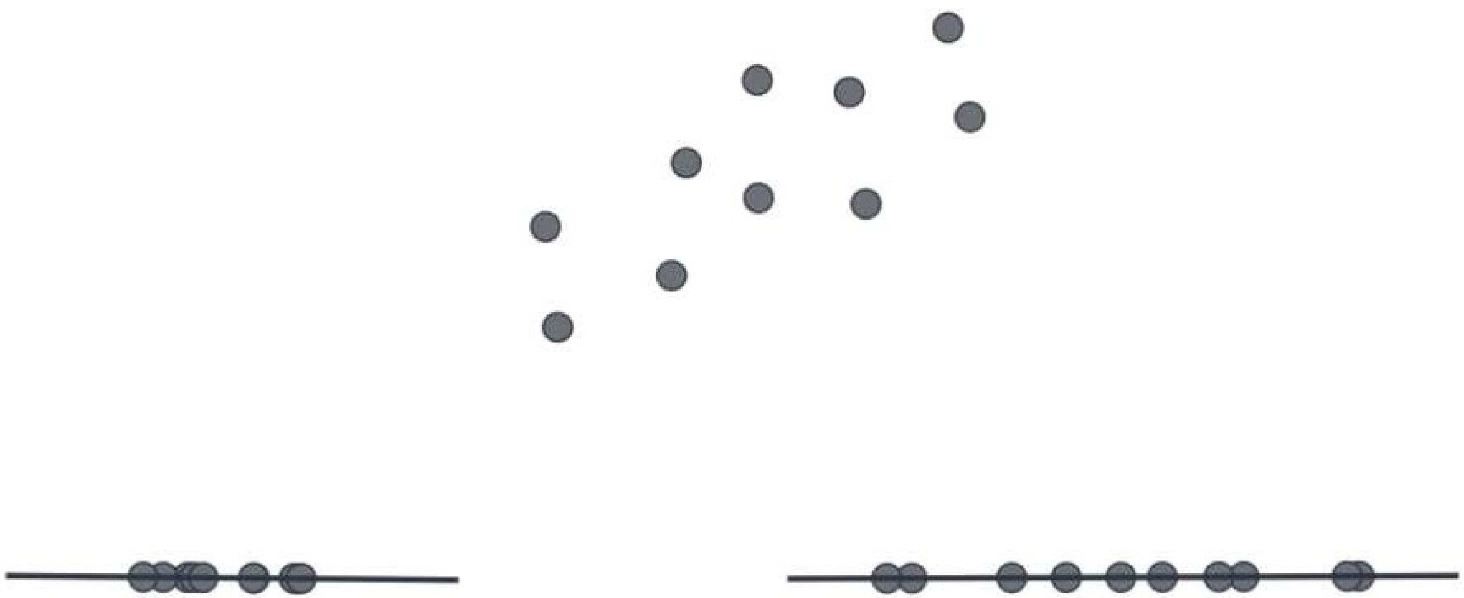
Dimensionality Reduction



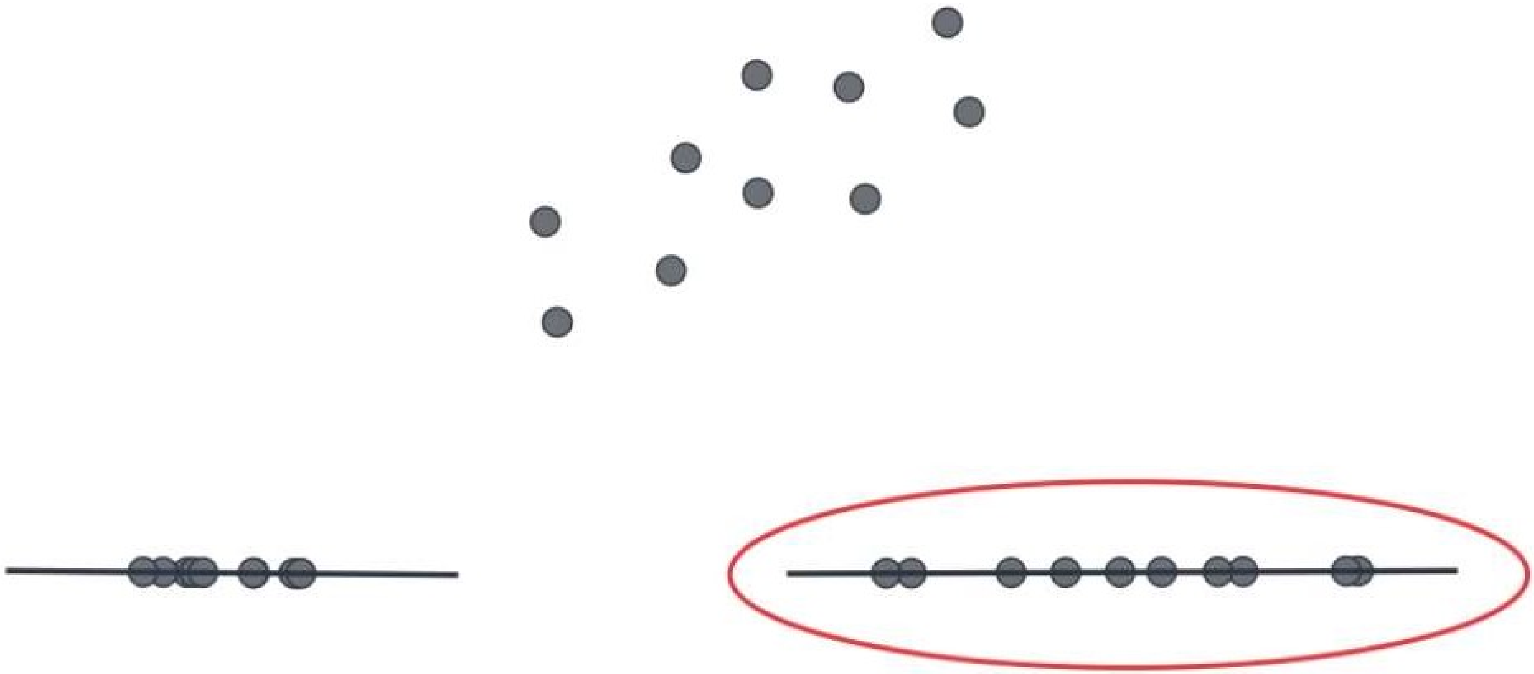
Dimensionality Reduction



Dimensionality Reduction



Dimensionality Reduction



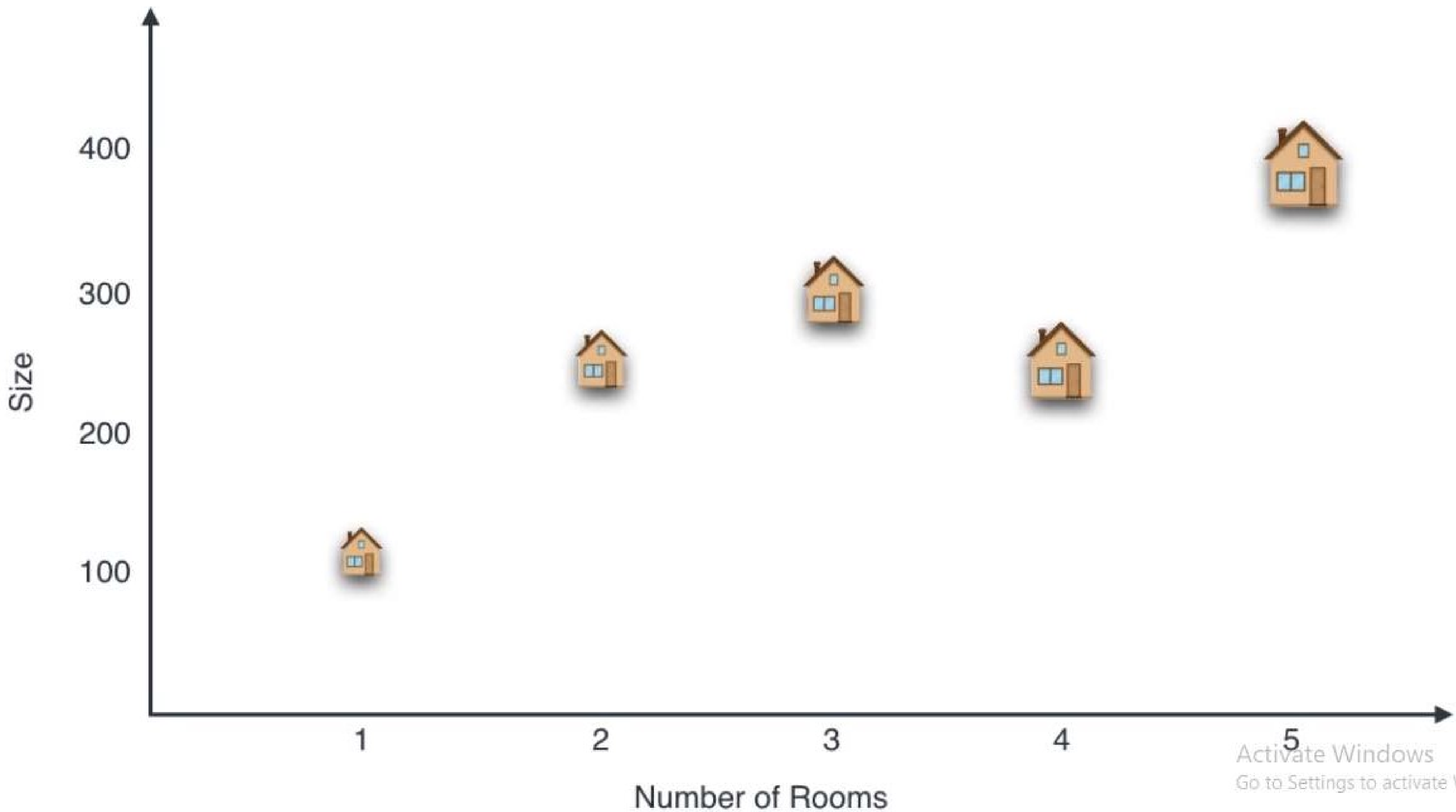
Housing Data

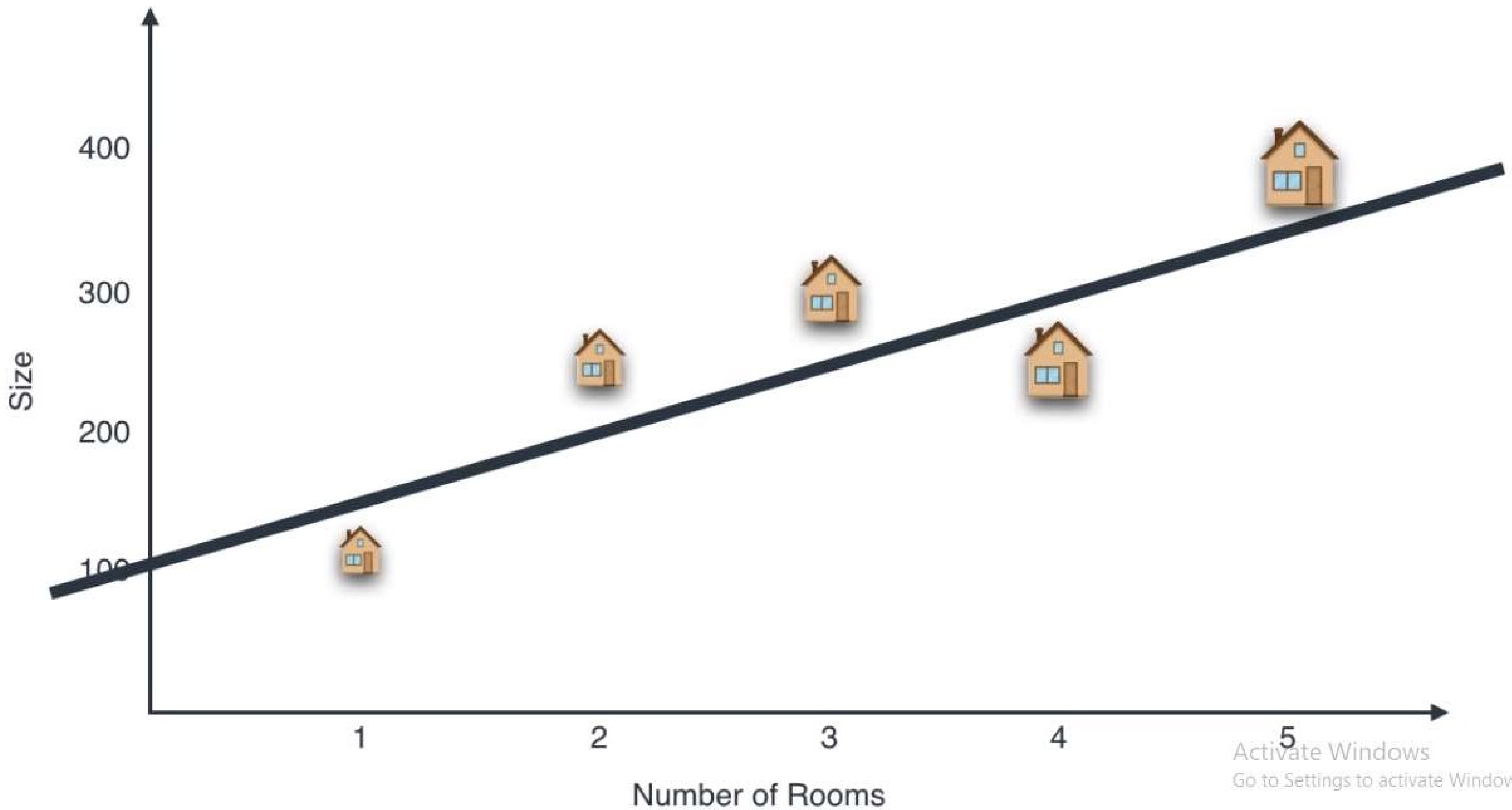
Size
Number of rooms
Number of bathrooms
Schools around
Crime rate

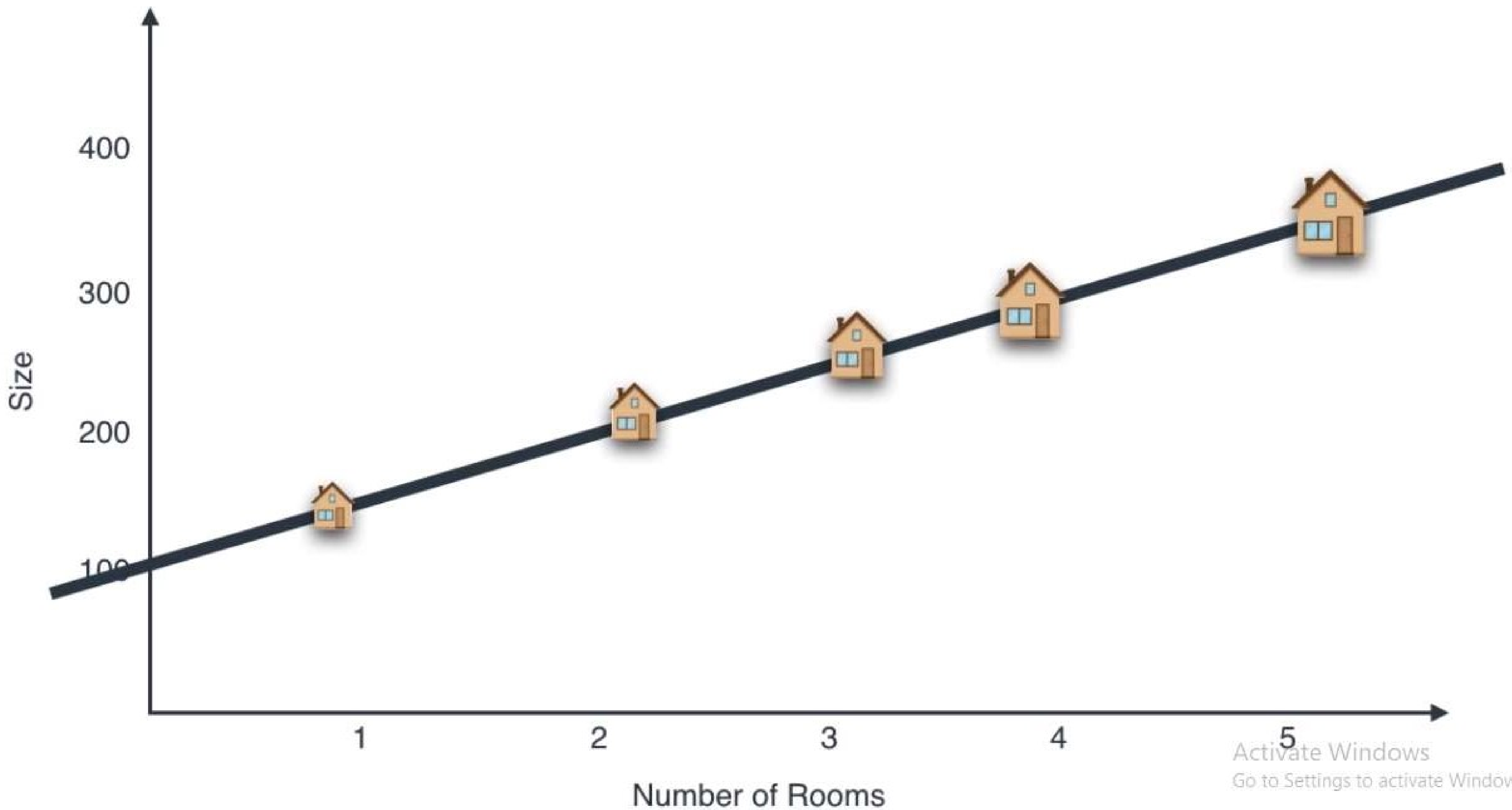
Housing Data

Size
Number of rooms → Size feature
Number of bathrooms

Schools around → Location feature
Crime rate


Size
400
300
200
100
1
2
3
4
5
Number of Rooms

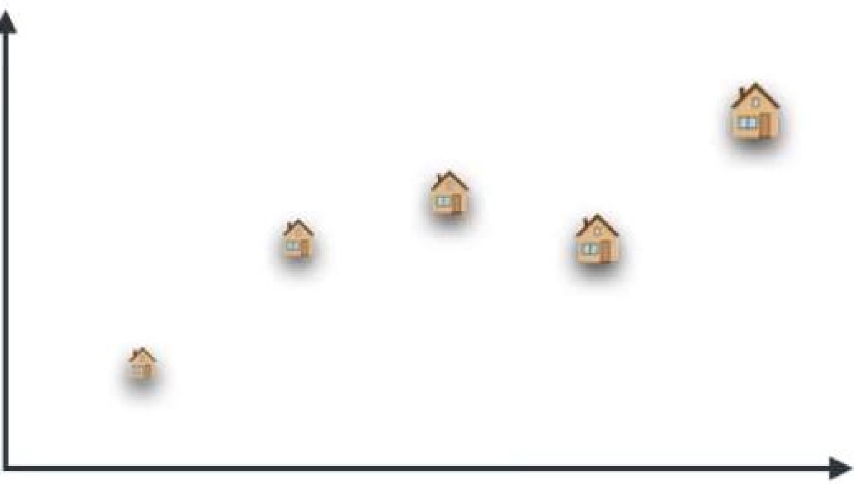

400
300
200
100
Size
1
2
3
4
5
Number of Rooms


400
300
200
100
Size
1
2
3
4
5
Number of Rooms

Size feature

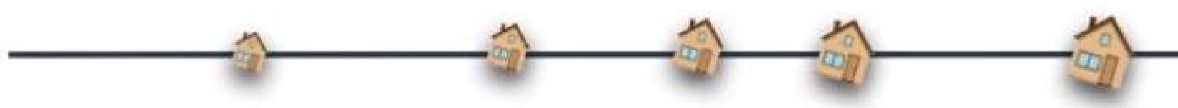
2 dimensions

size
number of rooms



1 dimension

size feature



Housing Data

5 dimensions

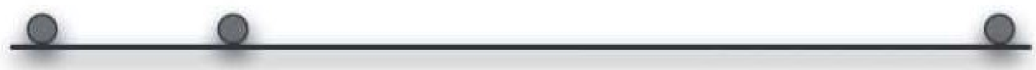
Size
Number of rooms
Number of bathrooms
Schools around
Crime rate

2 dimensions

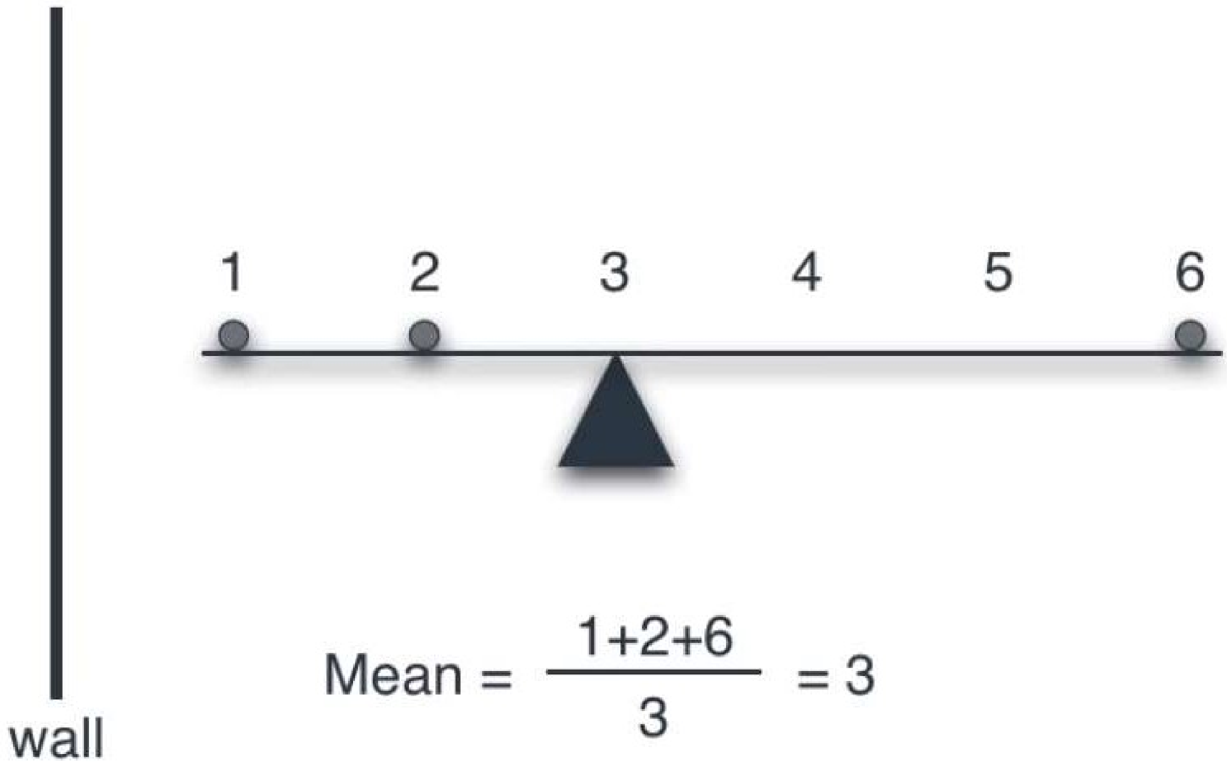
Size feature

Location feature

Mean

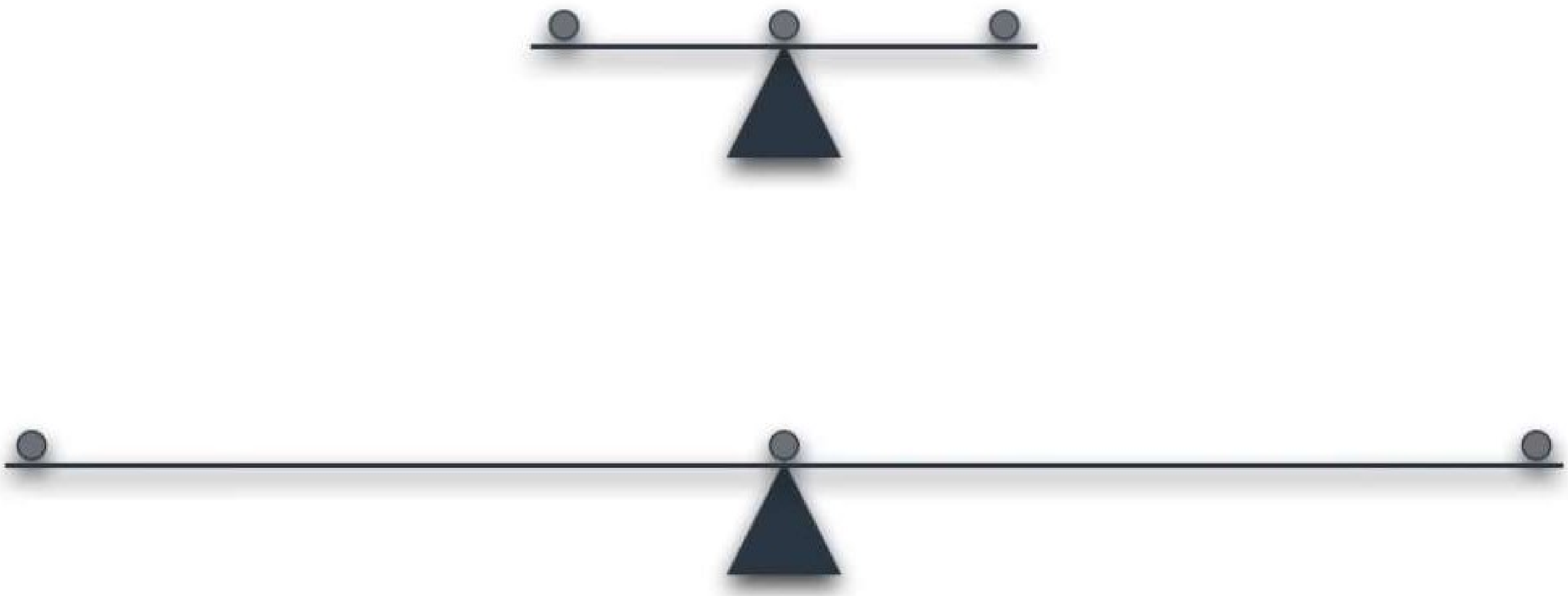


Mean



$$\text{Mean} = \frac{1+2+6}{3} = 3$$

Variance

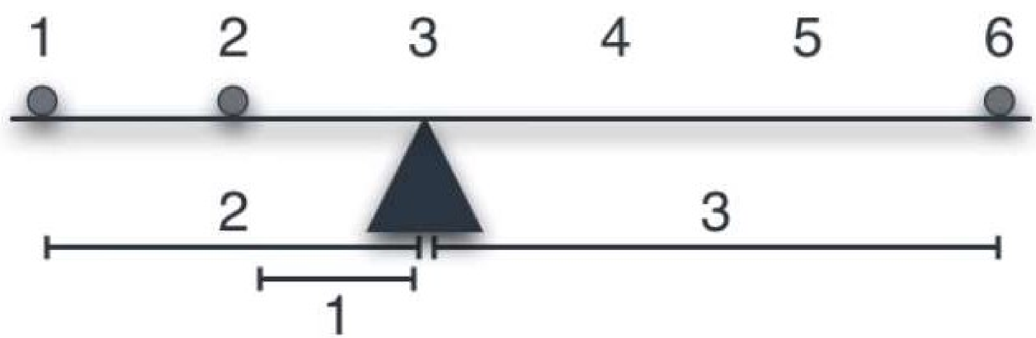


Variance

$$\text{Variance} = \frac{1^2+0^2+1^2}{3} = 2/3$$

$$\text{Variance} = \frac{5^2+0^2+5^2}{3} = 50/3$$

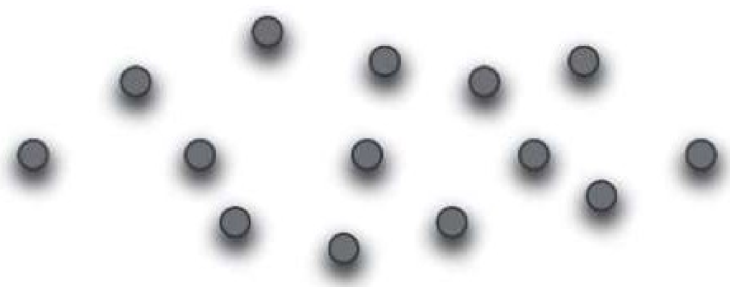
Mean



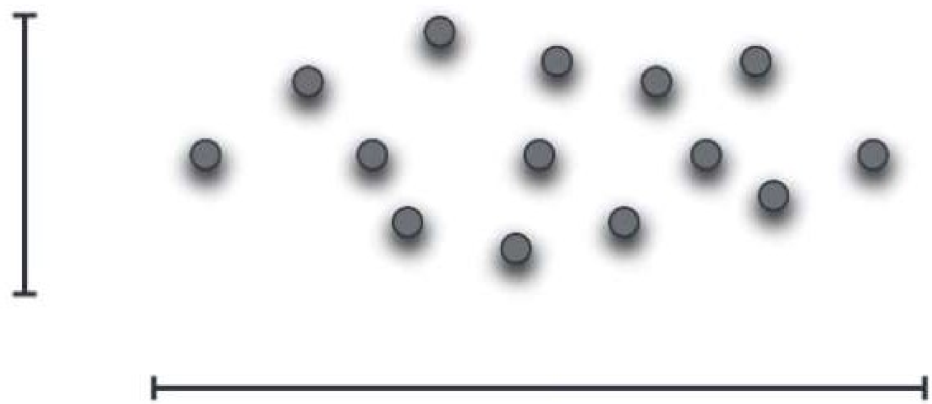
$$\text{Variance} = \frac{2^2+1^2+3^2}{3} = 14/3$$

Variance?

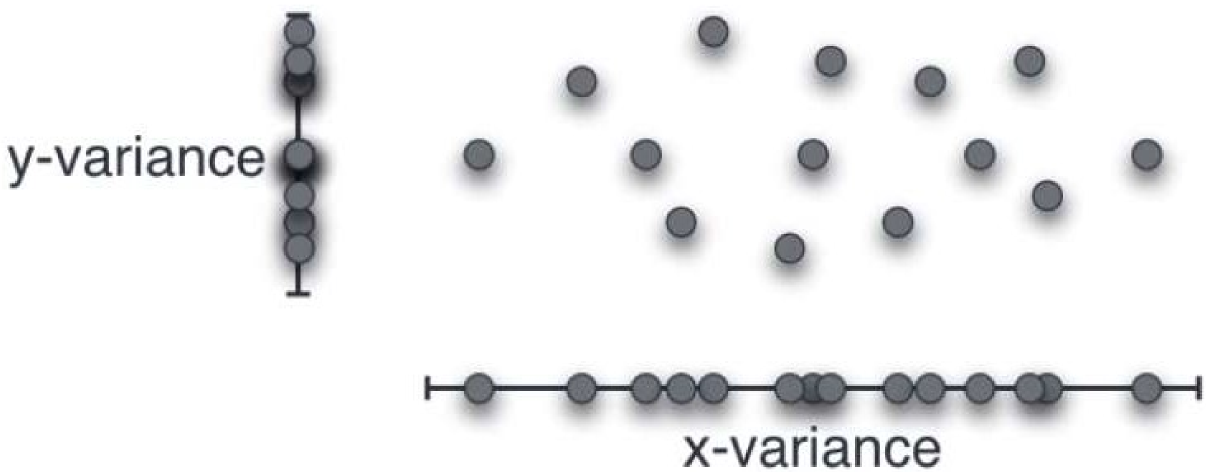
2D



Variance?



Variance?



Variance?

$$\text{x-variance} = \frac{2^2+0^2+2^2}{3} = 8/3$$

$$\text{y-variance} = \frac{1^2+0^2+1^2}{3} = 2/3$$

Fundamentally, two dataset are very different even they have same X and Y variance

Mean:
The mean is the average value of a set of numbers.
The mean is a measure of the central tendency of a dataset.

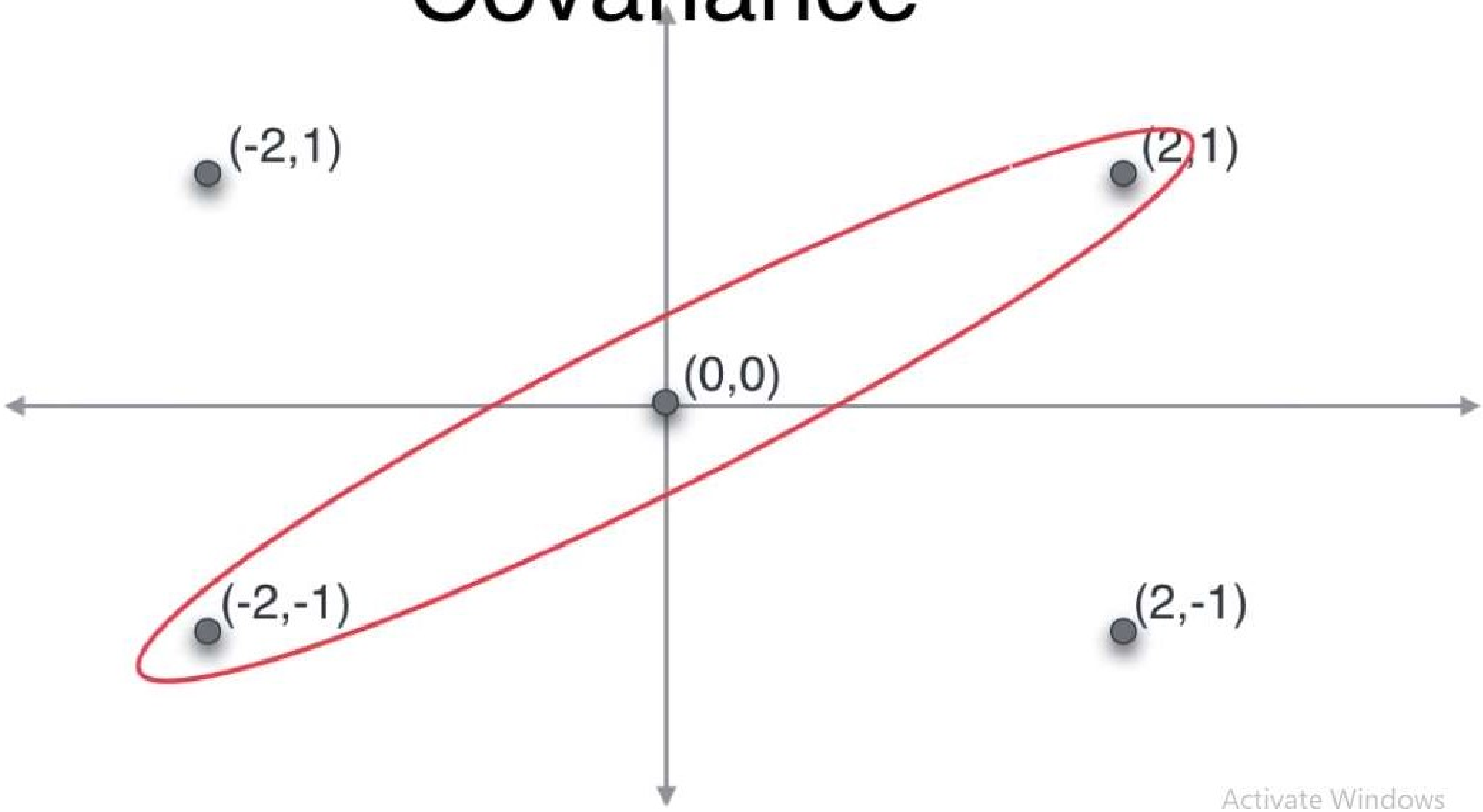
Variance:
The variance is a measure of how spread out a set of data is.
A high variance indicates that the data is spread out over a wider range of values, while a low variance indicates that the data is tightly clustered around the mean.

Covariance:
Covariance measures the degree to which two variables vary together.
It measures how much two variables change together, relative to their individual means.
If the two variables increase or decrease together, the covariance is positive. If one variable increases while the other decreases, the covariance is negative.
A covariance of 0 indicates that the two variables are independent of each other.

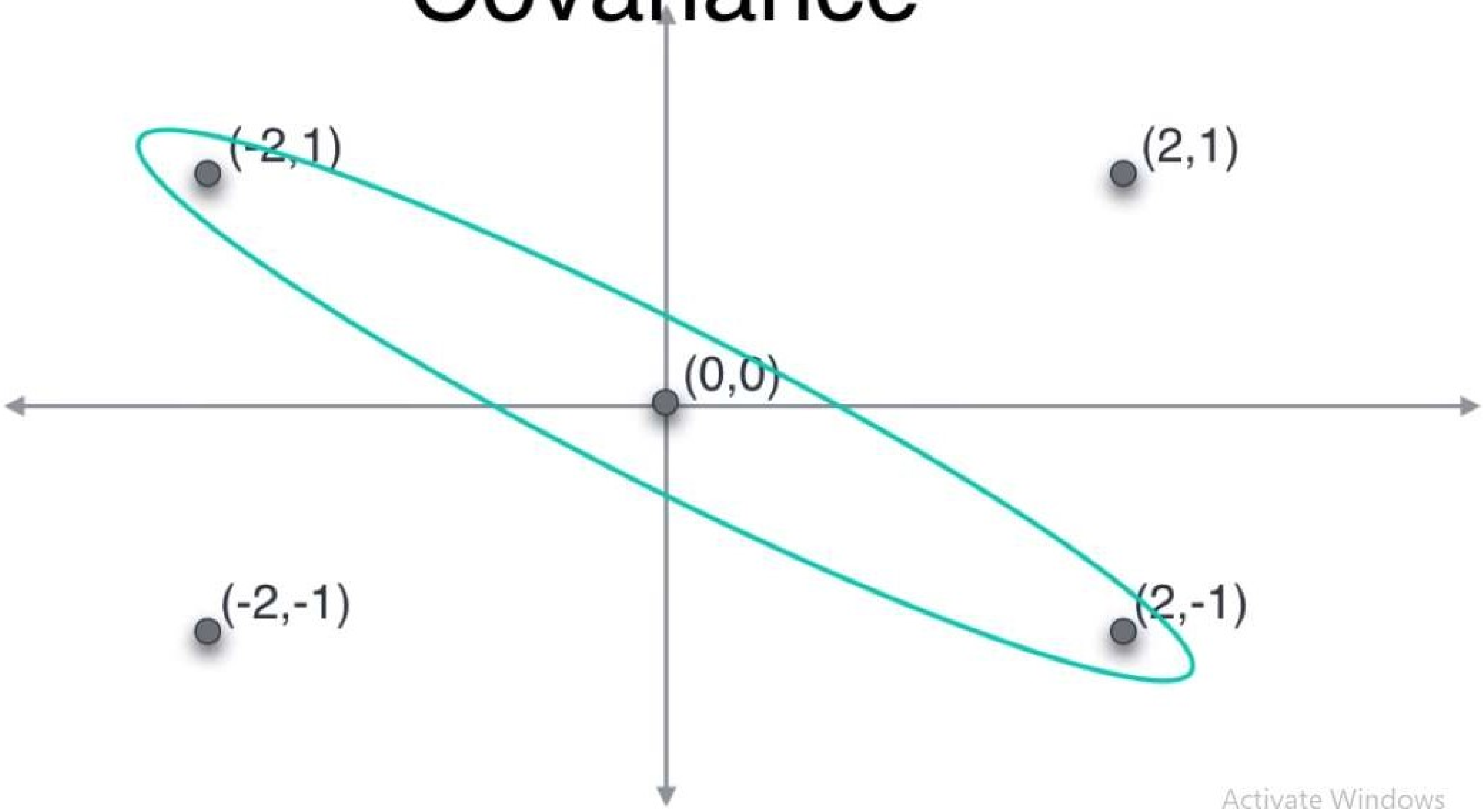
So, we need a third metrics, which is called

COVARIANCE

Covariance

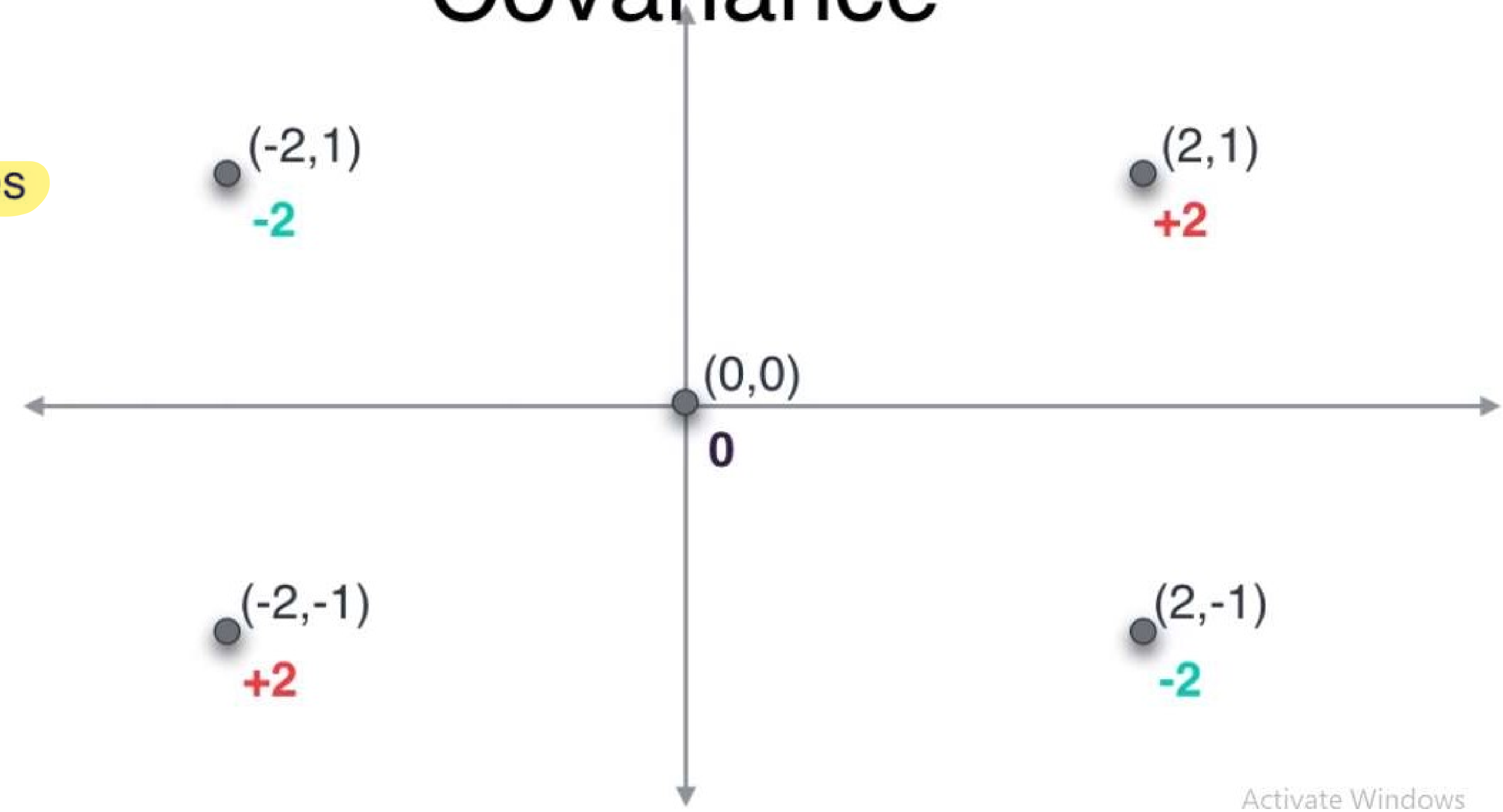


Covariance



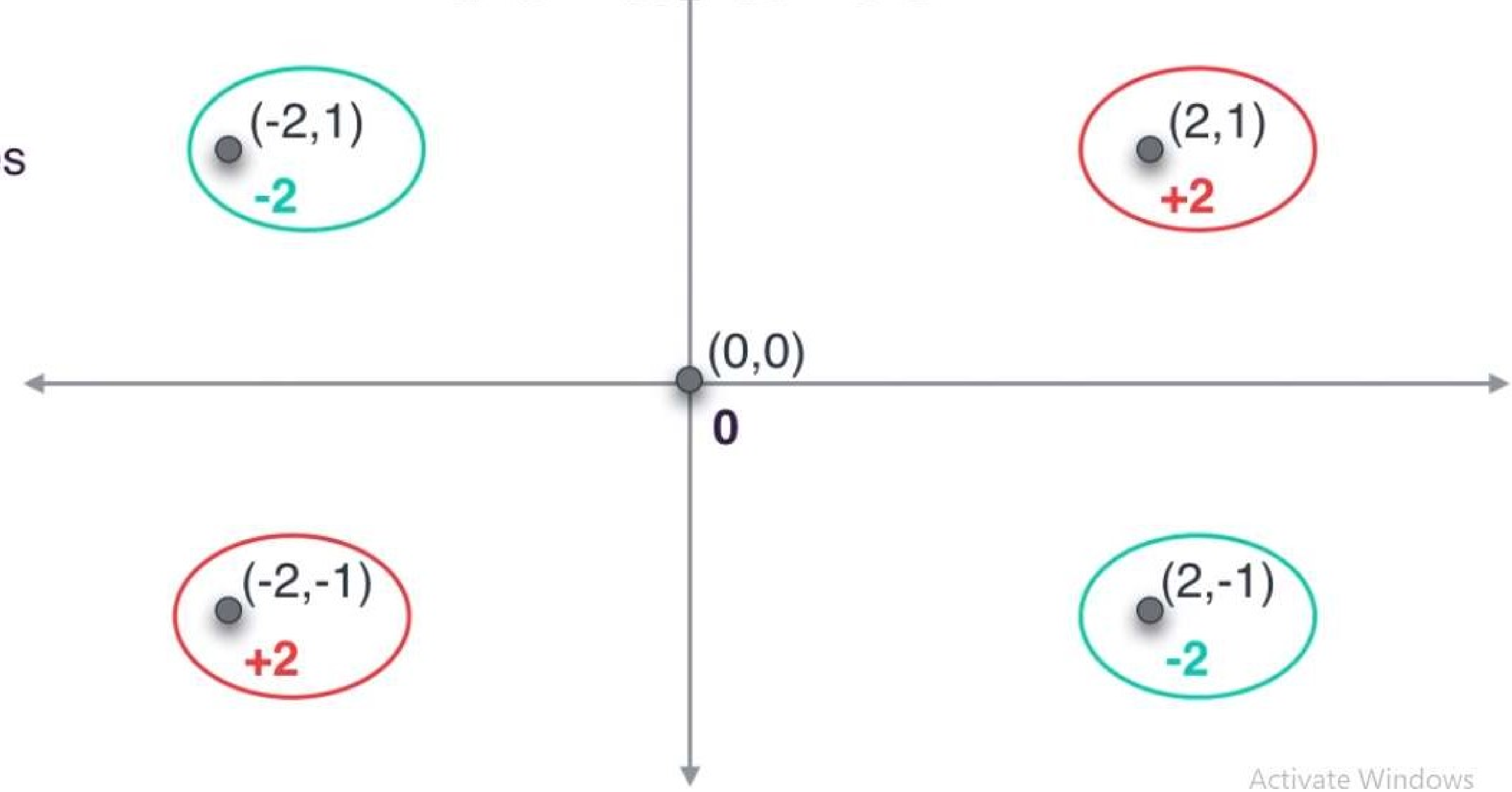
Covariance

Product of coordinates

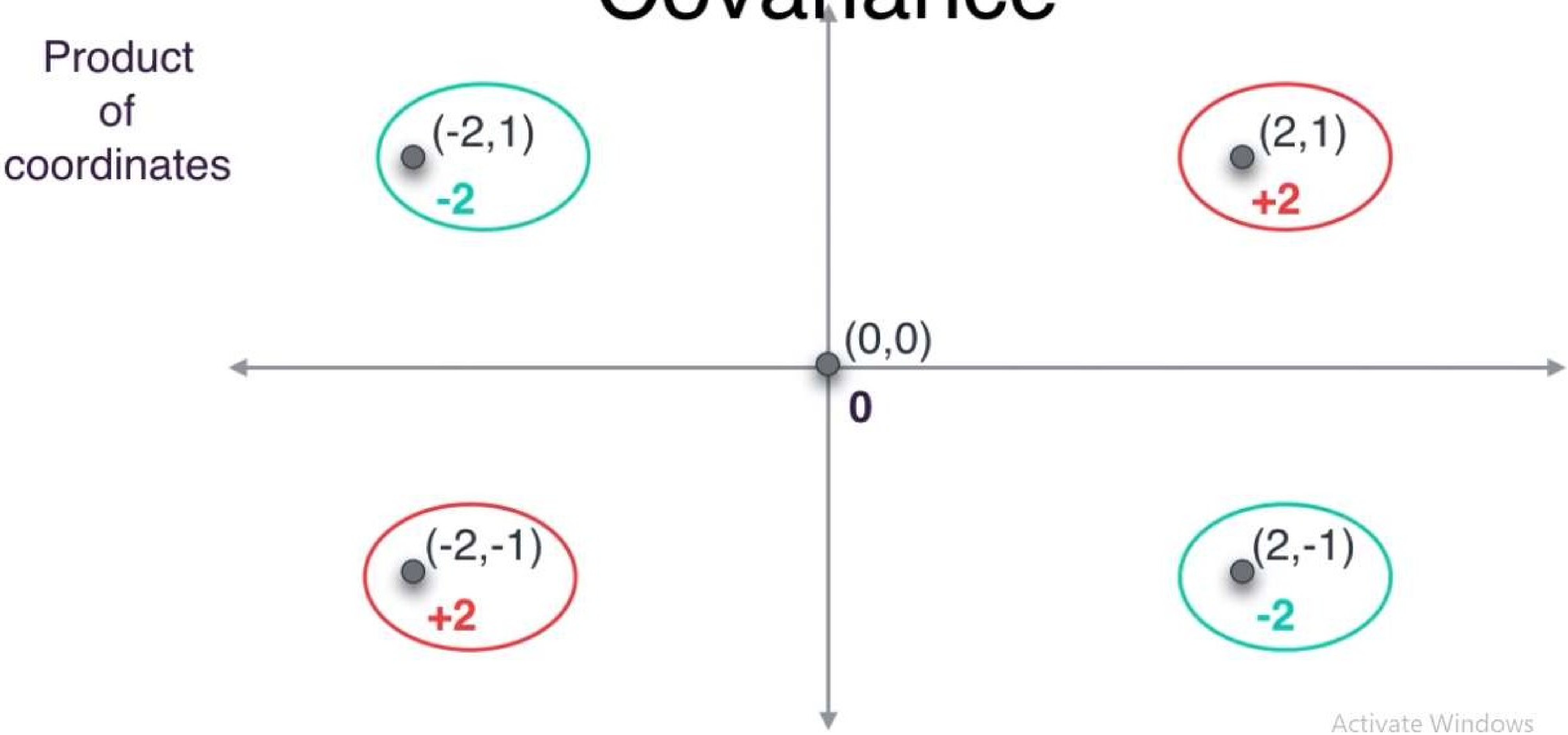


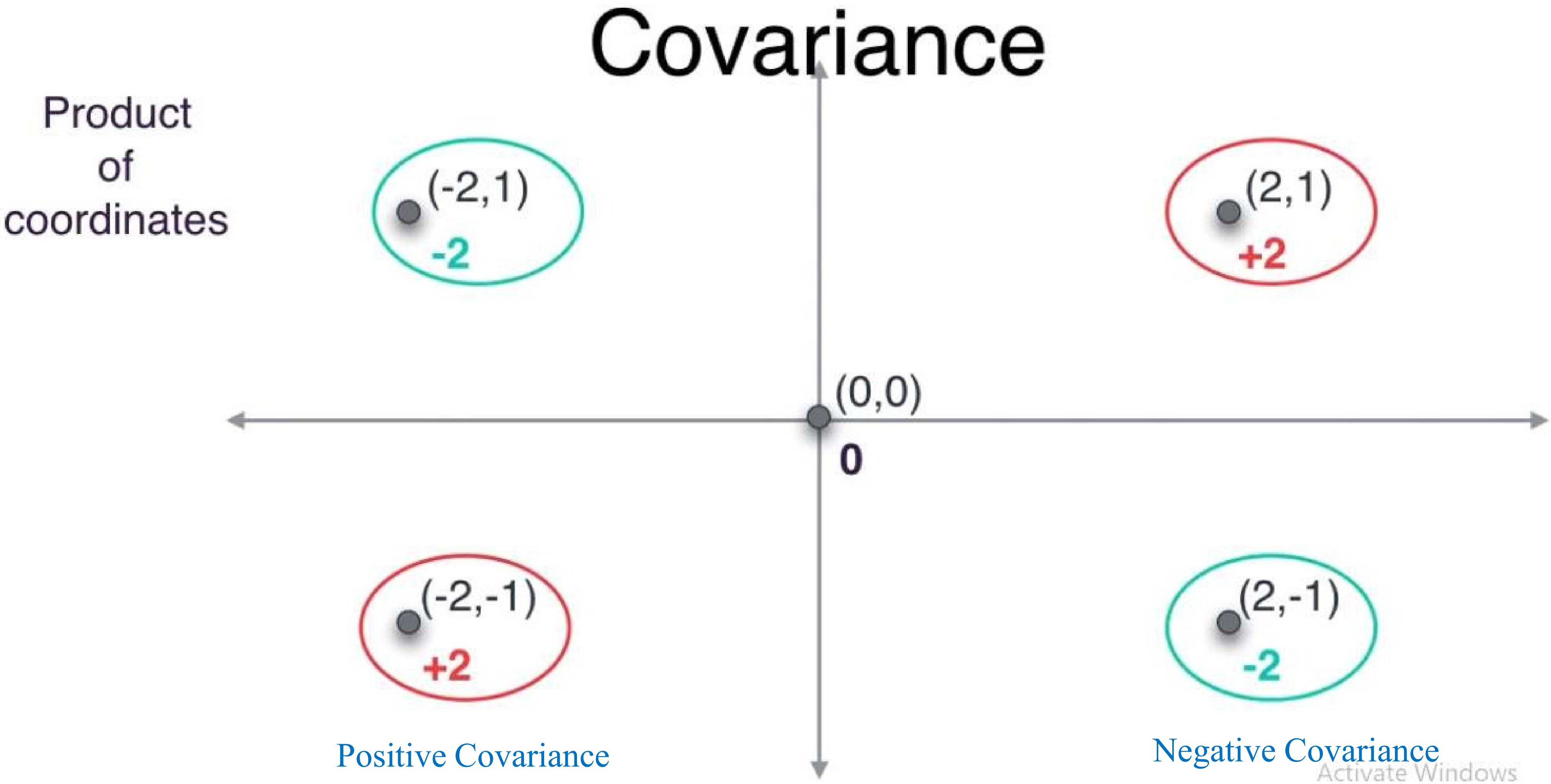
Covariance

Product of coordinates



Covariance

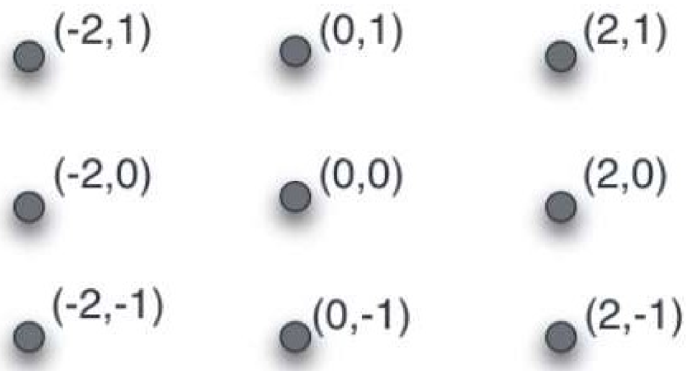




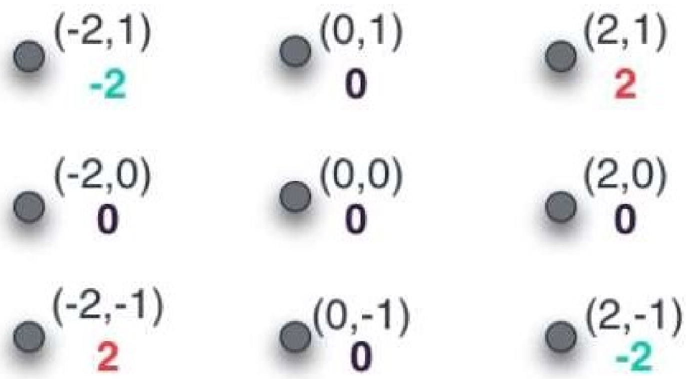
Positive Covariance

Negative Covariance

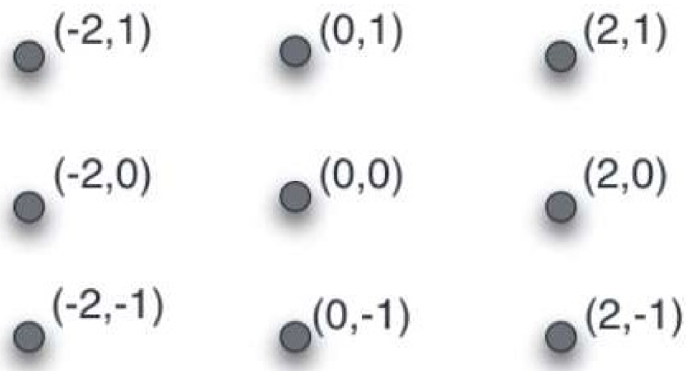
Covariance



Covariance

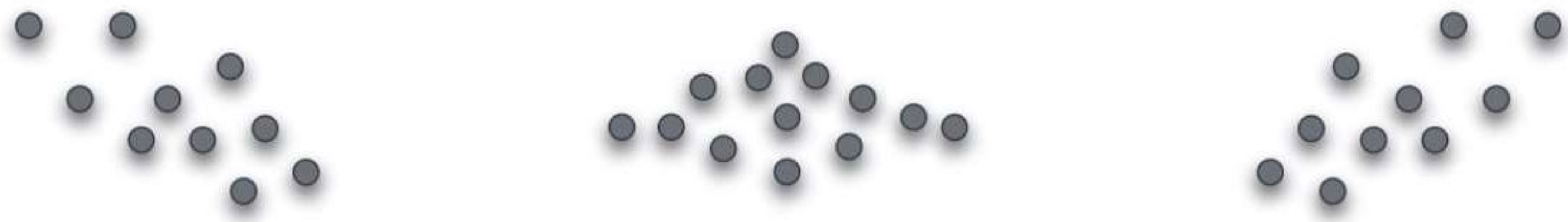


Covariance

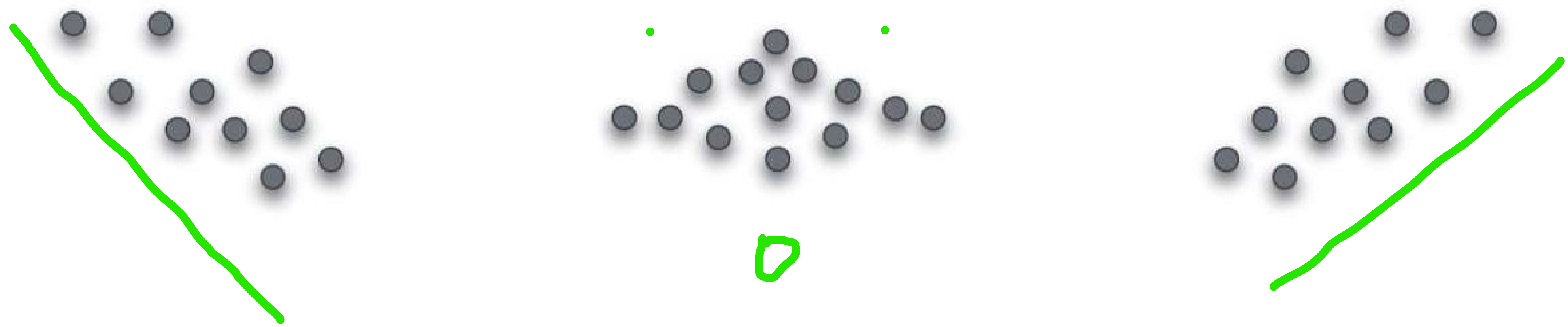


$$\text{covariance} = \frac{-2 + 0 + 2 + 0 + 0 + 0 + 2 + 0 + -2}{9} = 0$$

Covariance



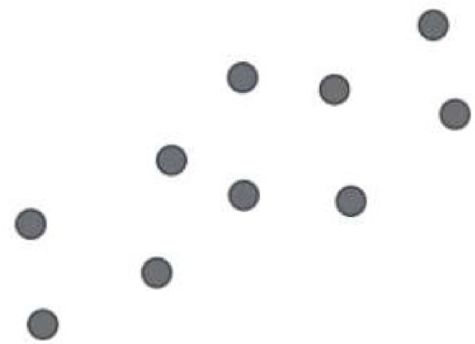
Covariance

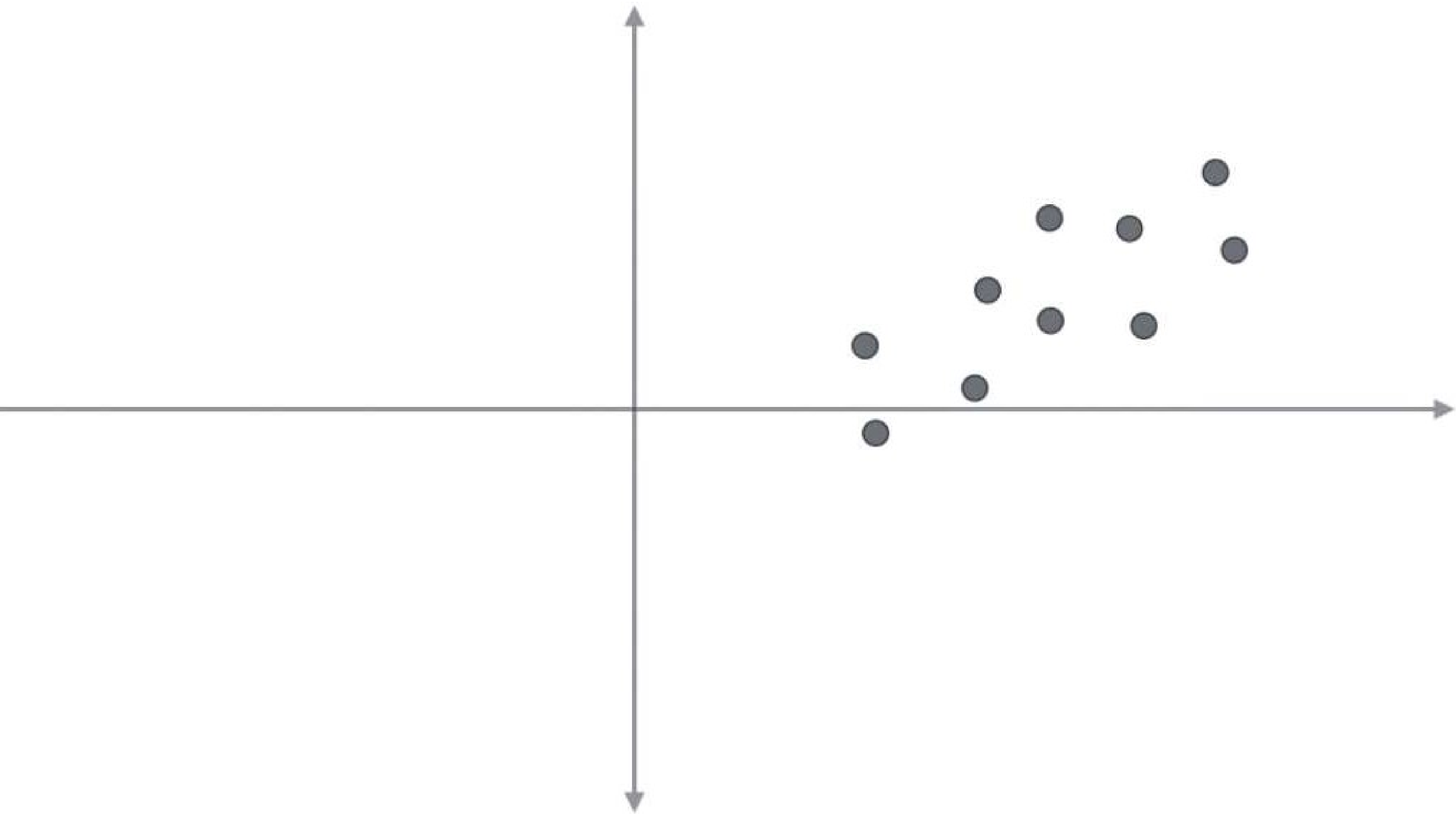


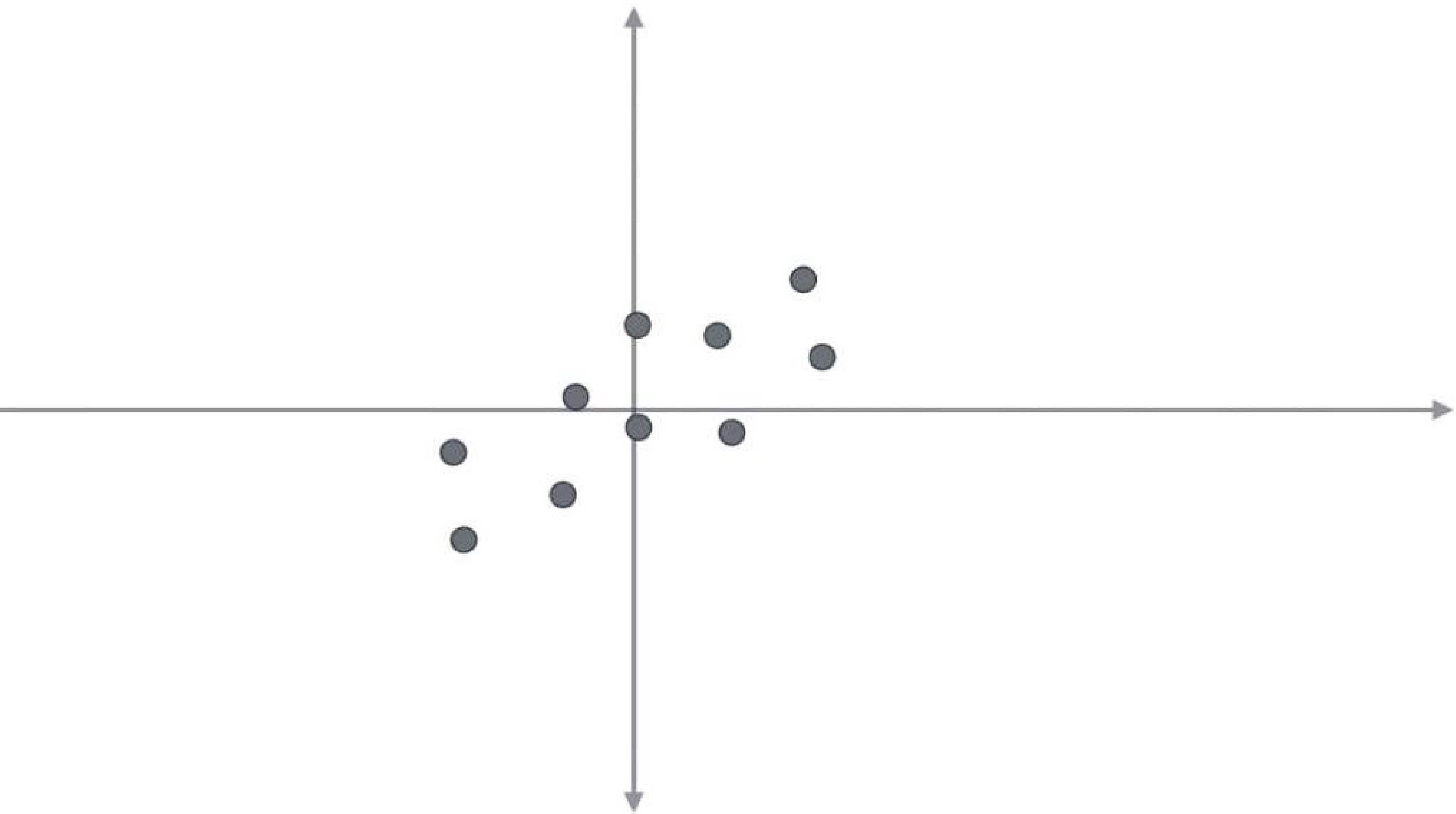
negative covariance

covariance zero (or very small)

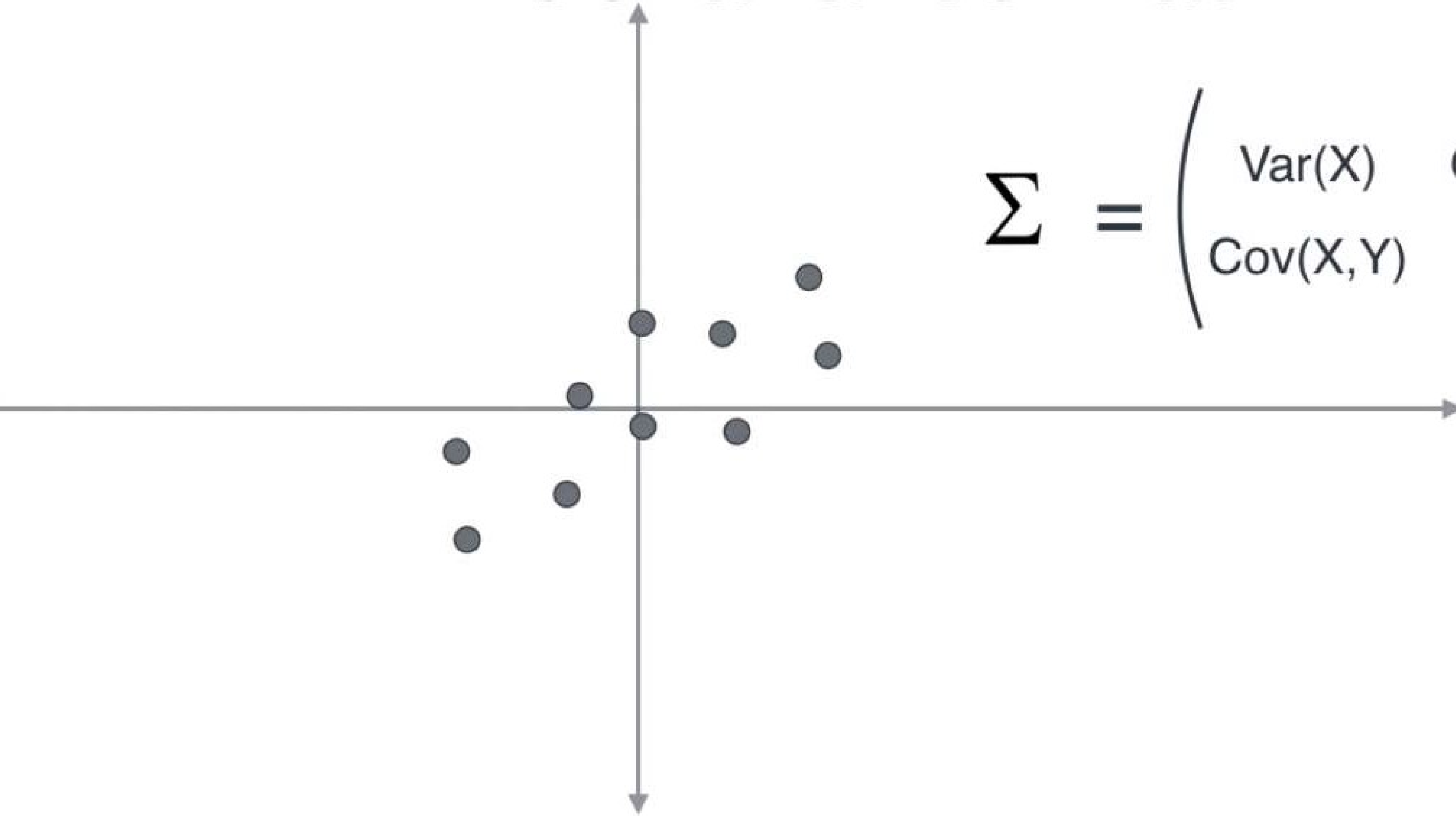
positive covariance





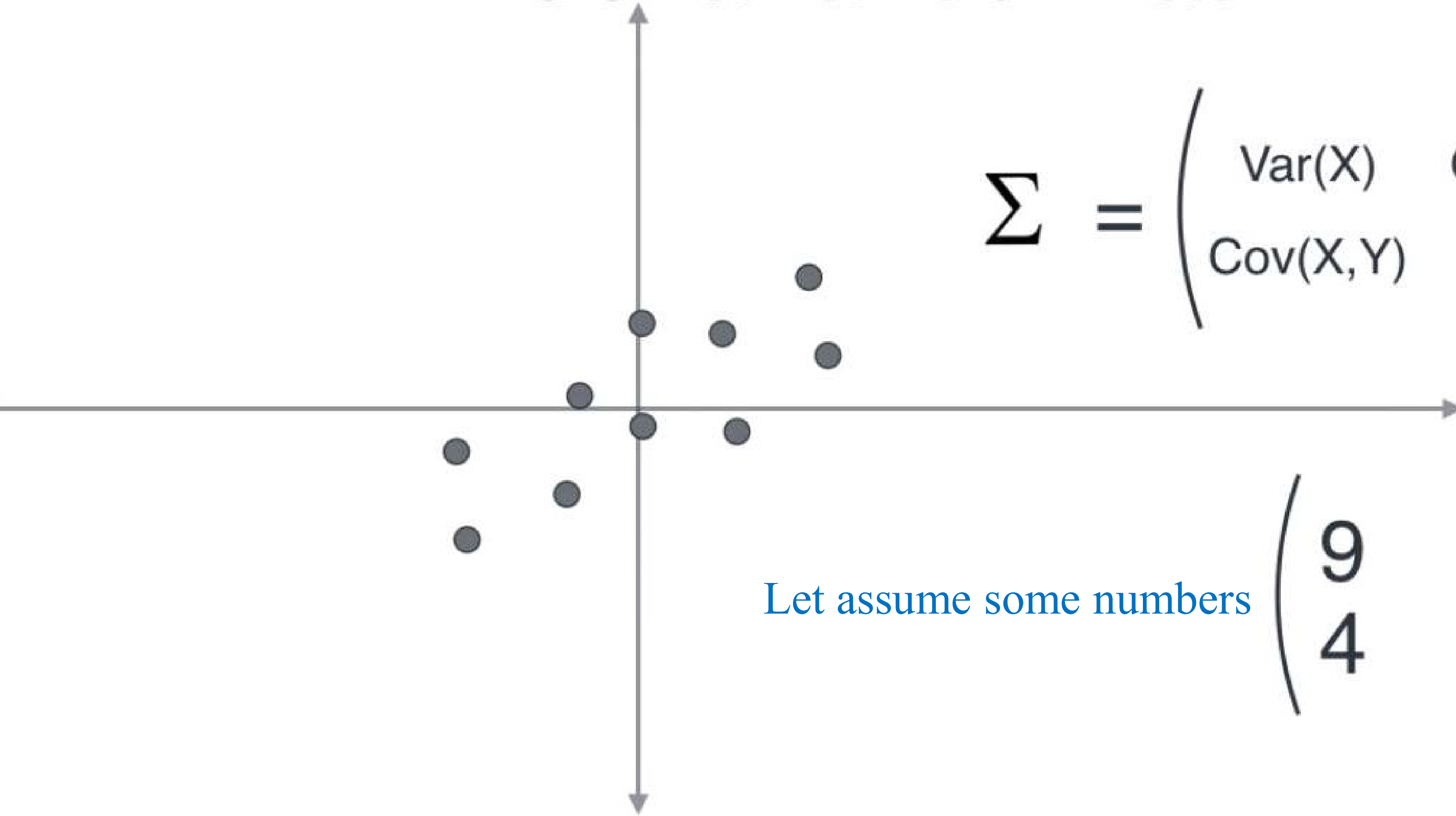


Covariance matrix



$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}$$

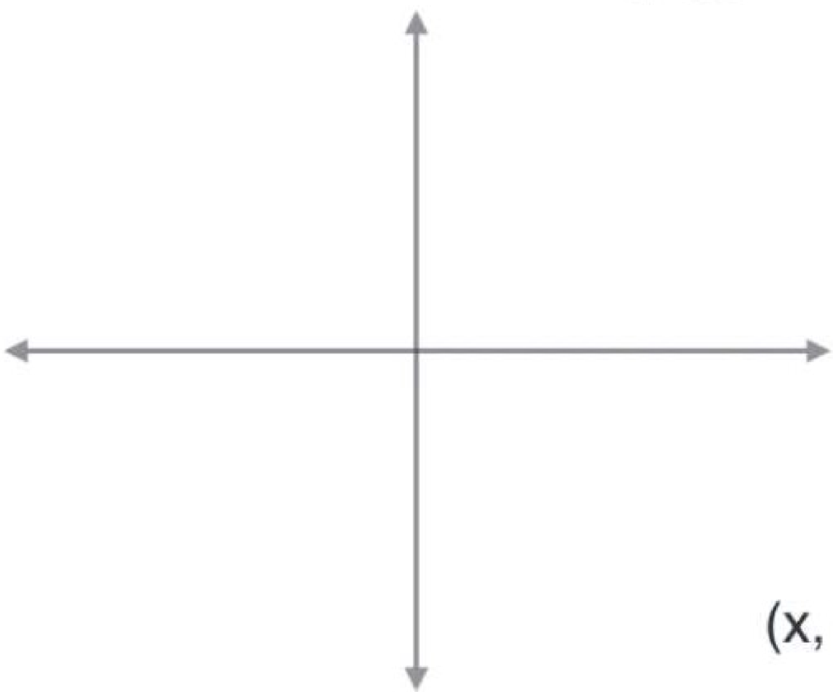
Covariance matrix



$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}$$

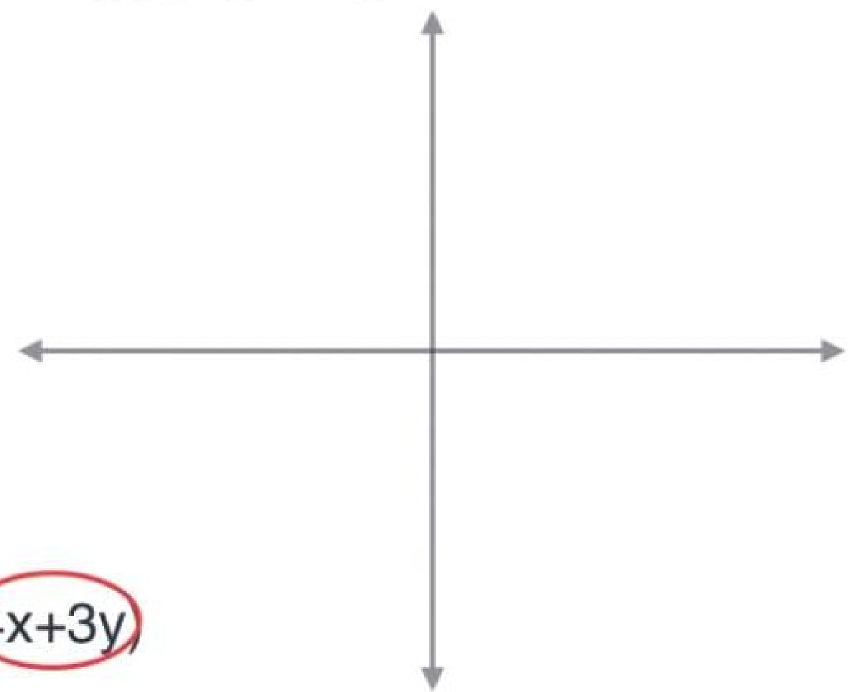
Let assume some numbers $\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$

Linear Transformations



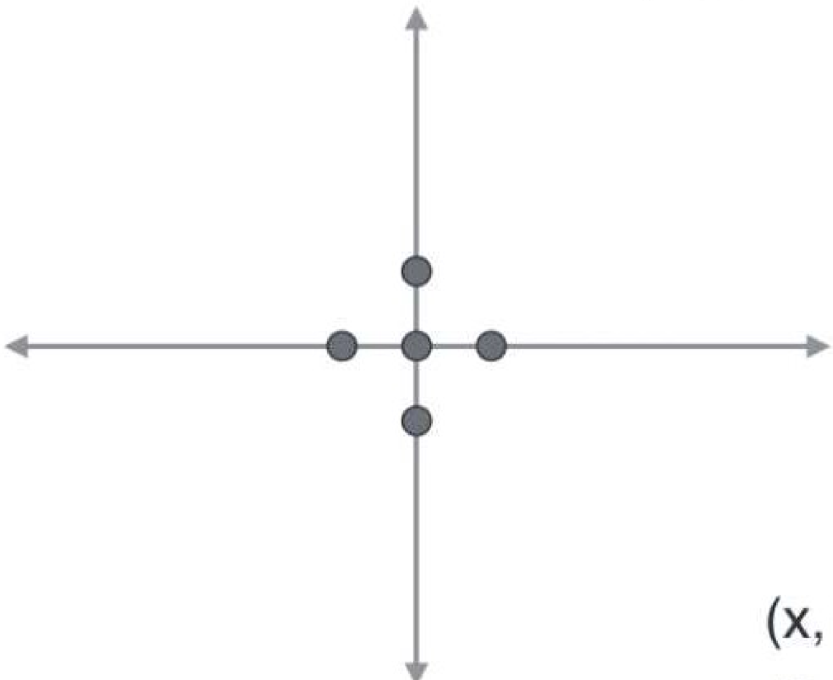
$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

$$\longrightarrow$$



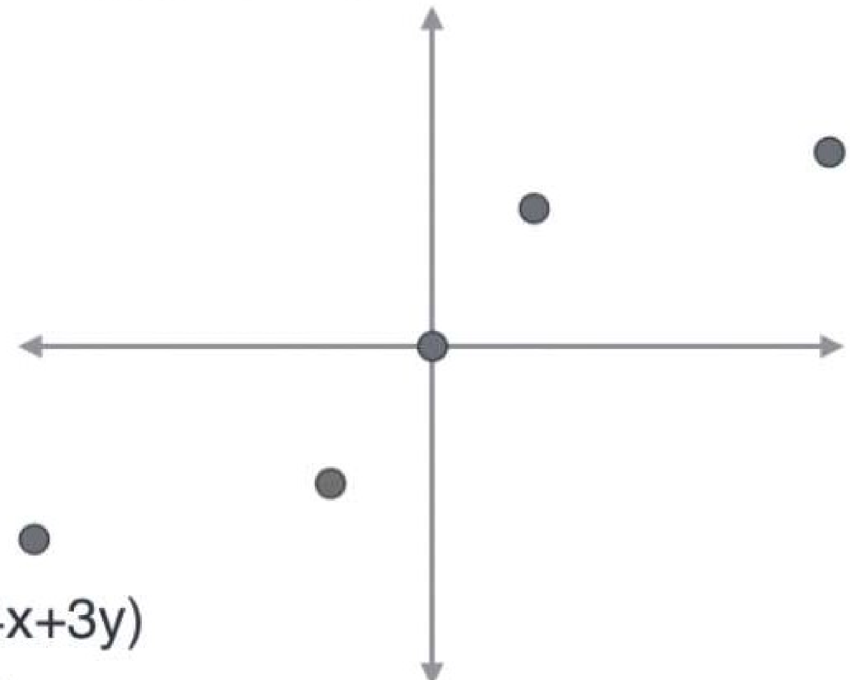
$$(x, y) \longrightarrow (9x+4y, 4x+3y)$$

Linear Transformations



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

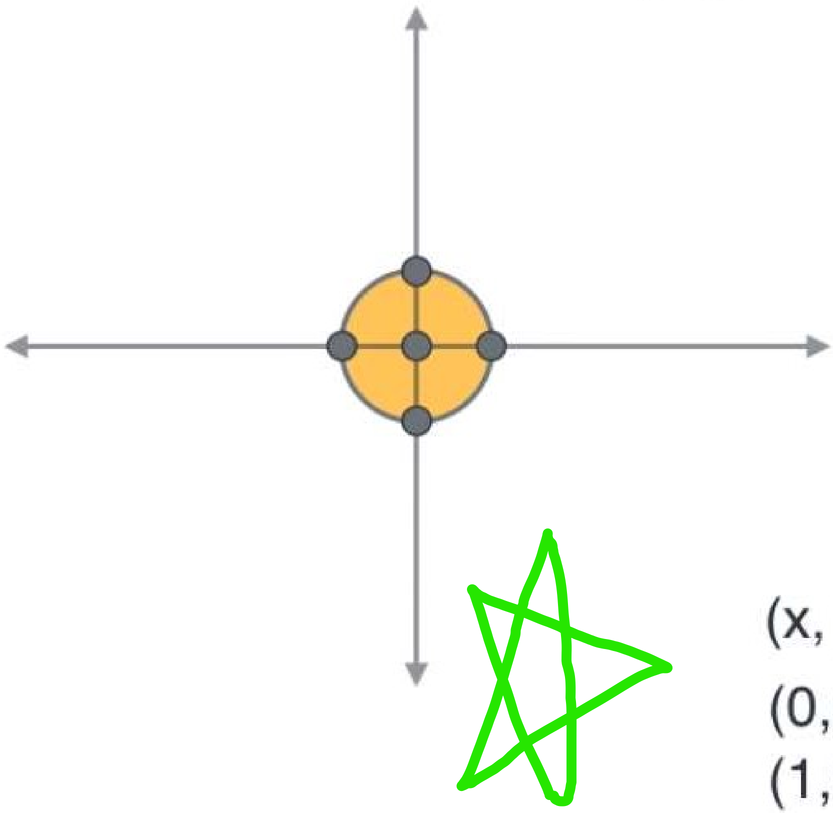
$$\longrightarrow$$



$(x, y) \longrightarrow (9x+4y,\ 4x+3y)$

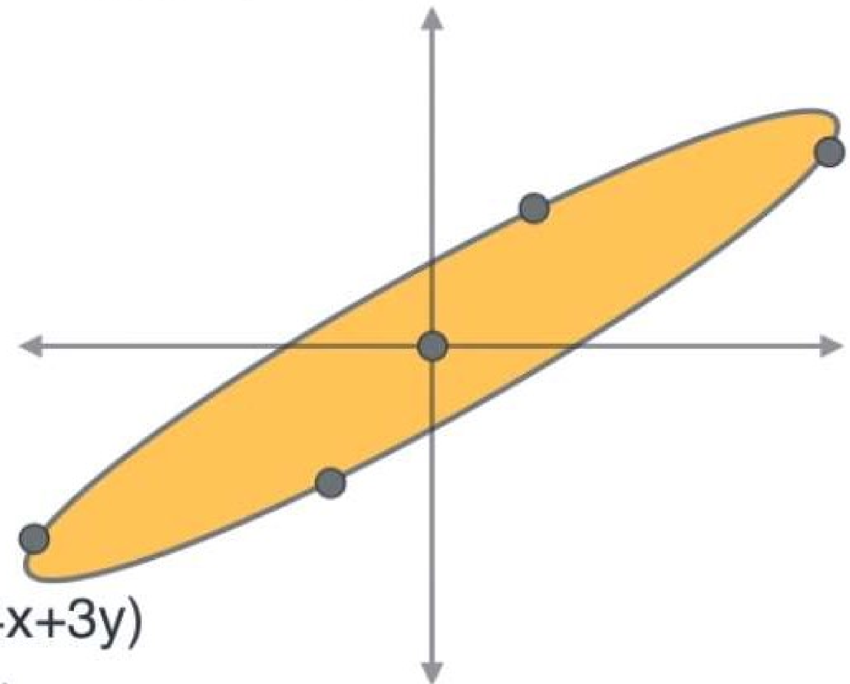
(0,0)	(0,0)
(1,0)	(9,4)
(0,1)	(4,3)
(-1,0)	(-9,-4)
(0,-1)	(-4,-3)

Linear Transformations



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

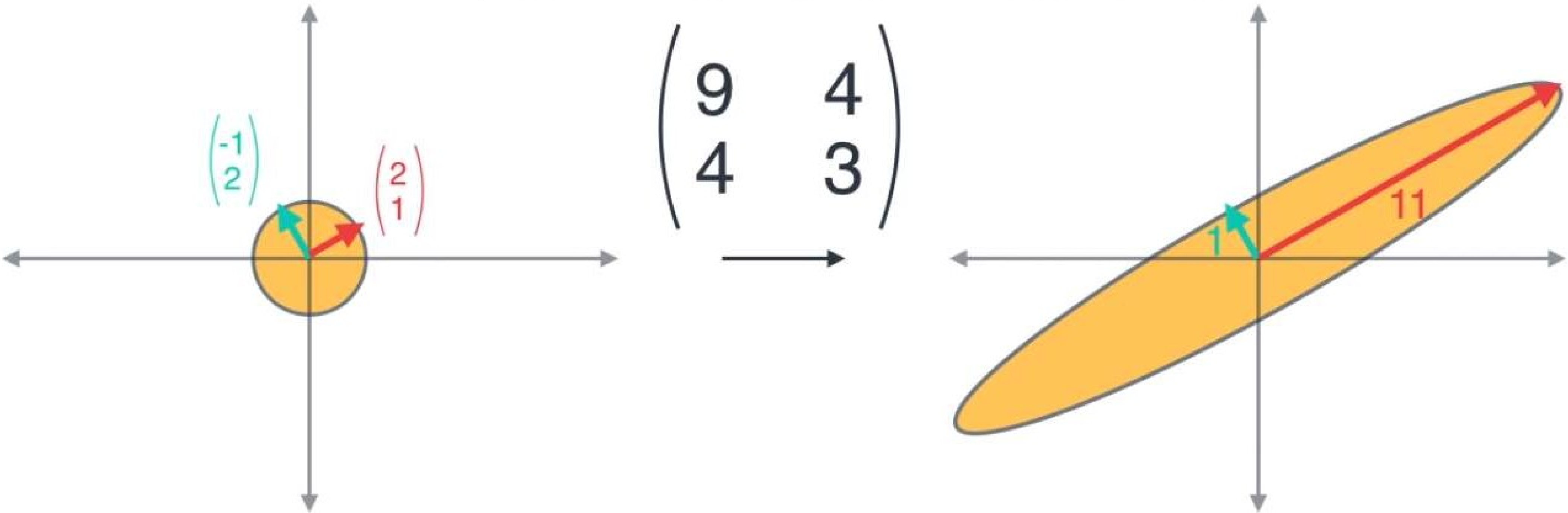
$$\longrightarrow$$



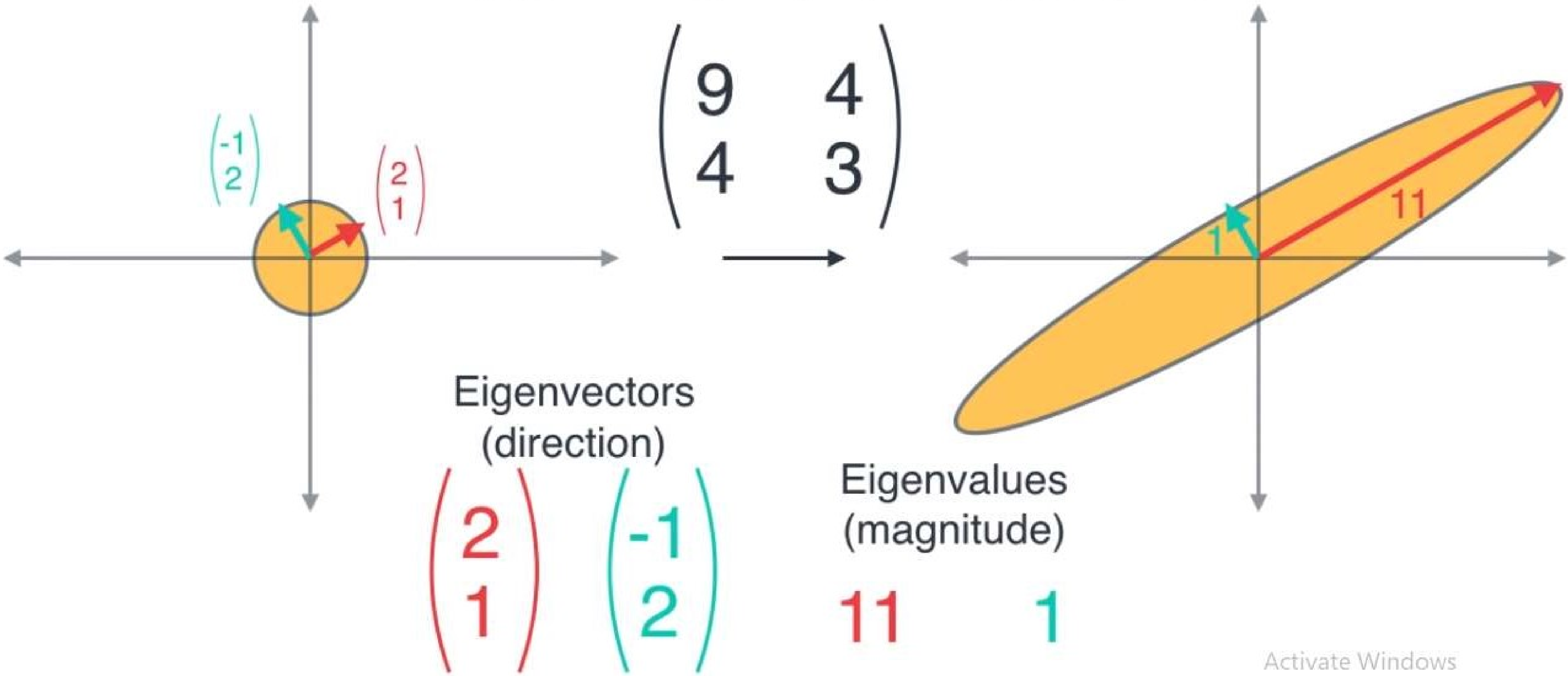
$(x, y) \longrightarrow (9x+4y,\ 4x+3y)$

(0,0)	(0,0)
(1,0)	(9,4)
(0,1)	(4,3)
(-1,0)	(-9,-4)
(0,-1)	(-4,-3)

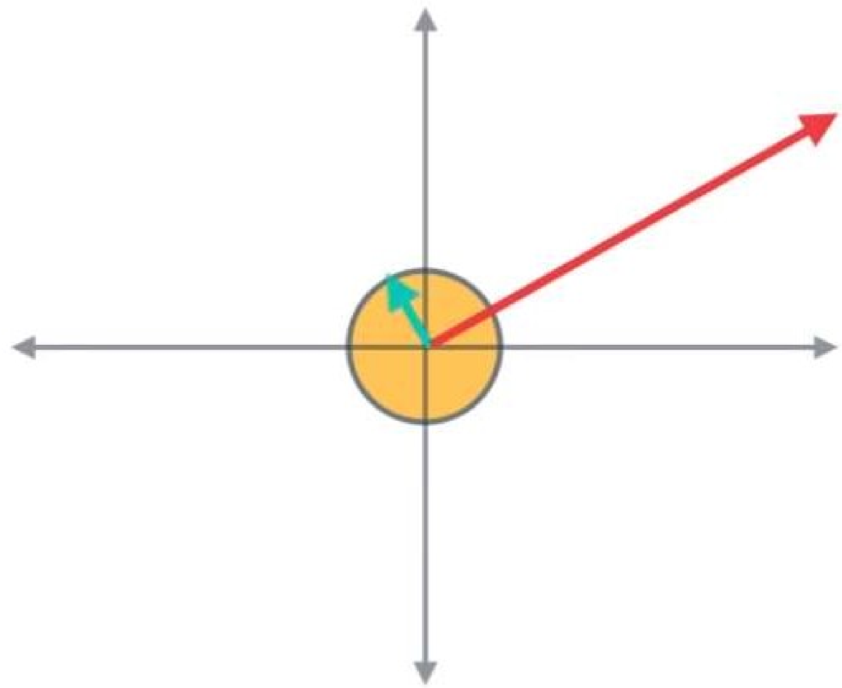
Linear Transformations



Linear Transformations



Linear Transformations



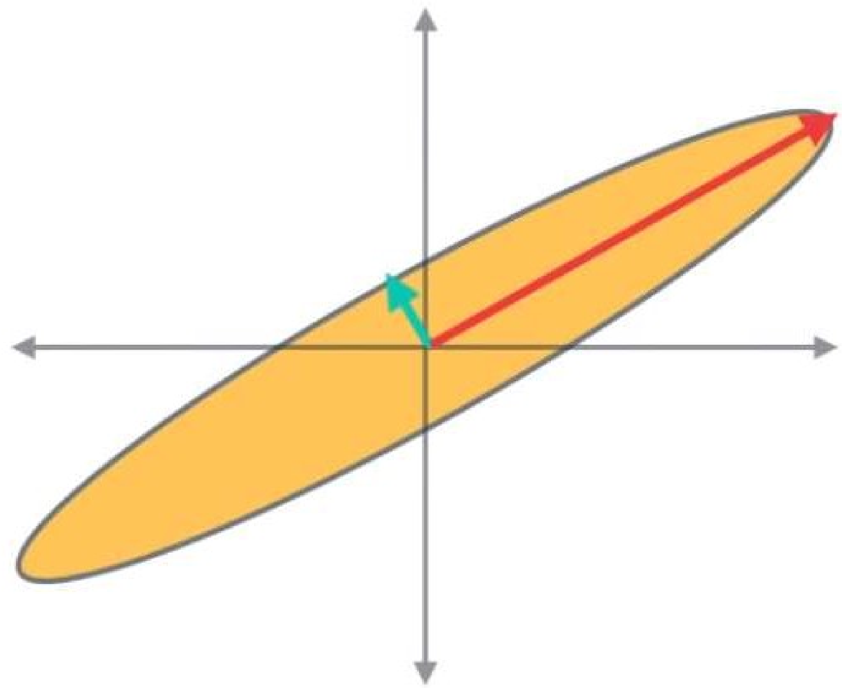
Eigenvectors

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

Eigenvalues

11 1

Linear Transformations



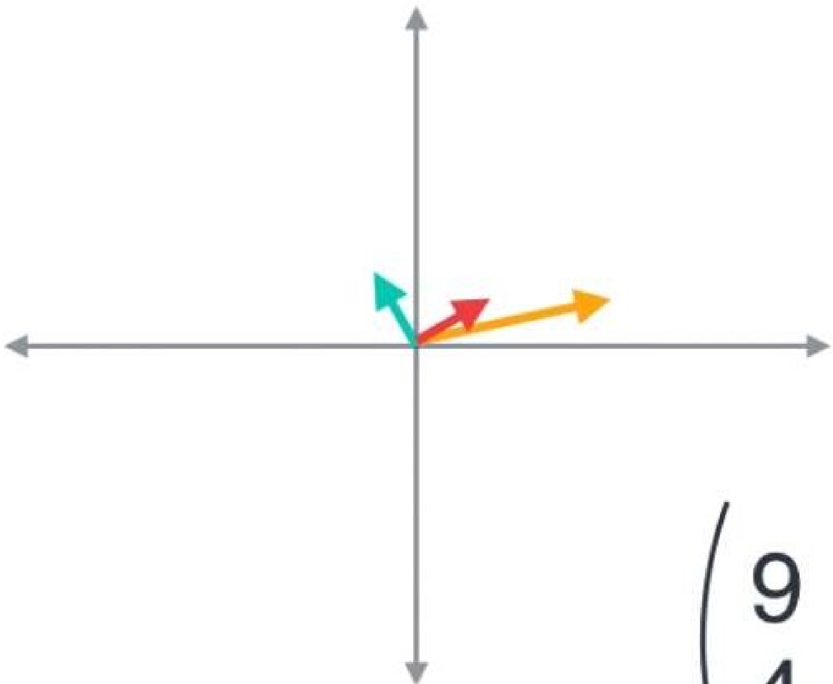
Eigenvectors
(direction)

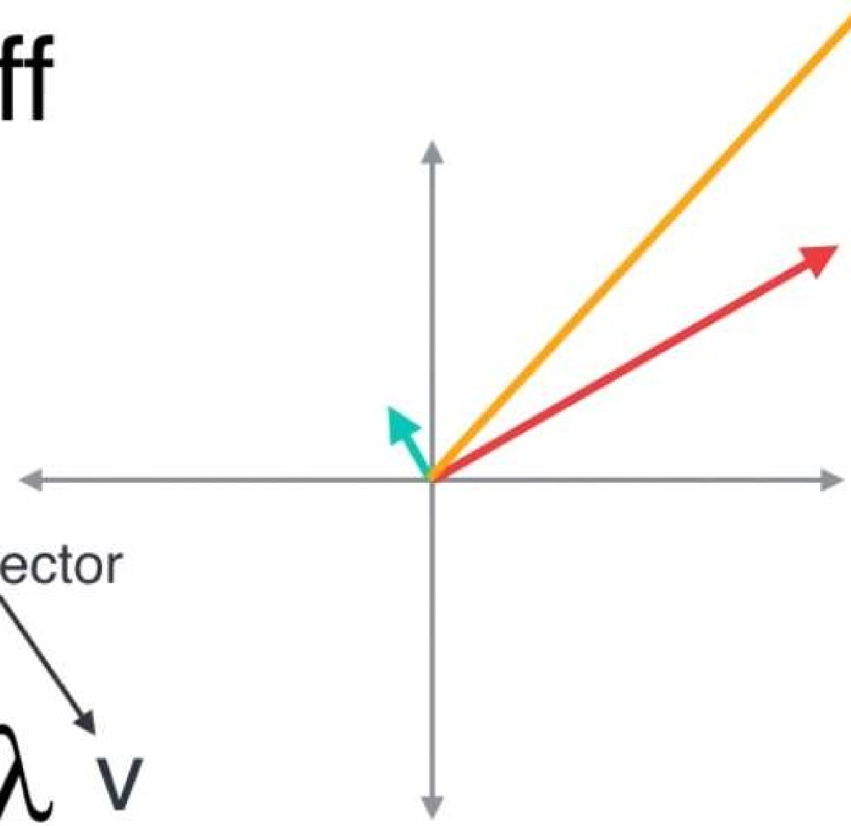
$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Eigenvalues
(magnitude)

11 1

Eigenstuff

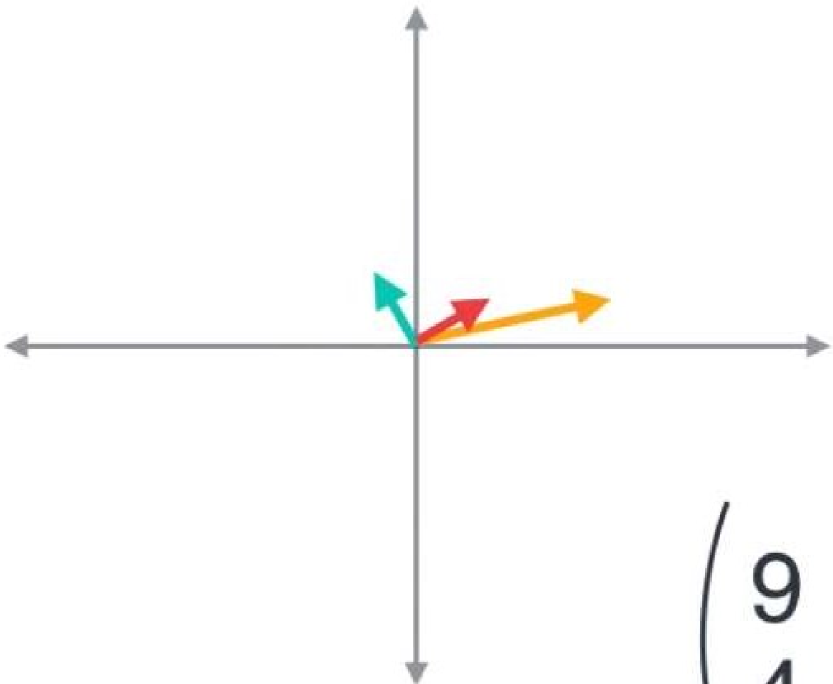


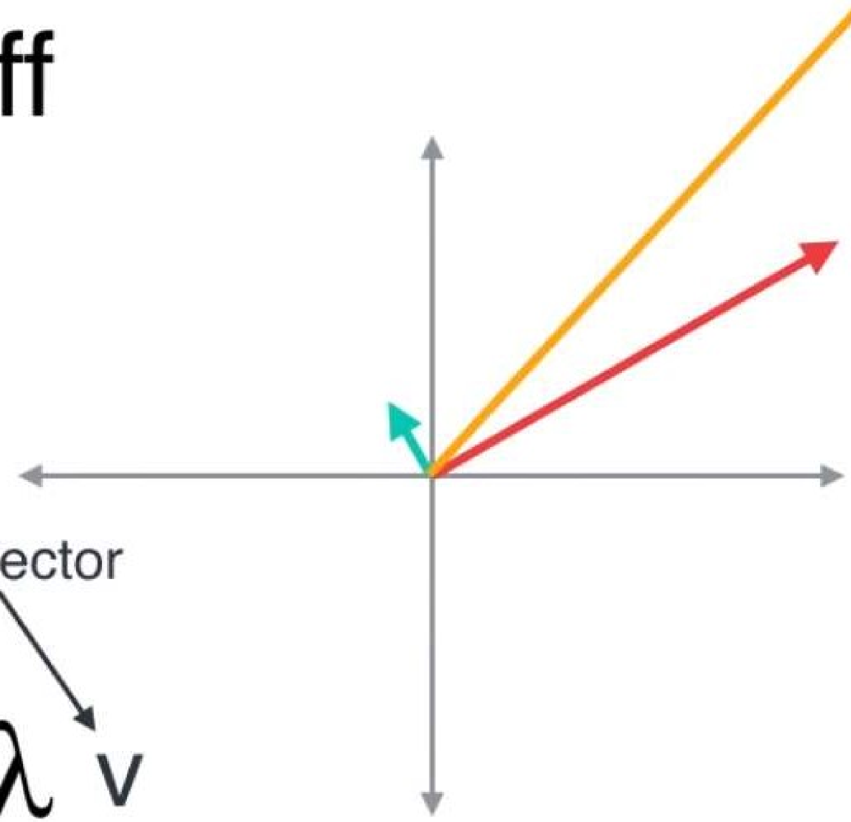


Eigenvector

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

Eigenstuff





Eigenvector

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

Eigenvalue

Eigenvalues

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

Characteristic Polynomial

$$\begin{vmatrix} x-9 & -4 \\ -4 & x-3 \end{vmatrix} = (x-9)(x-3) - (-4)(-4) = x^2 - 12x + 11$$

$$= (x-11)(x-1)$$

Eigenvalues 11 and 1

Eigenvectors

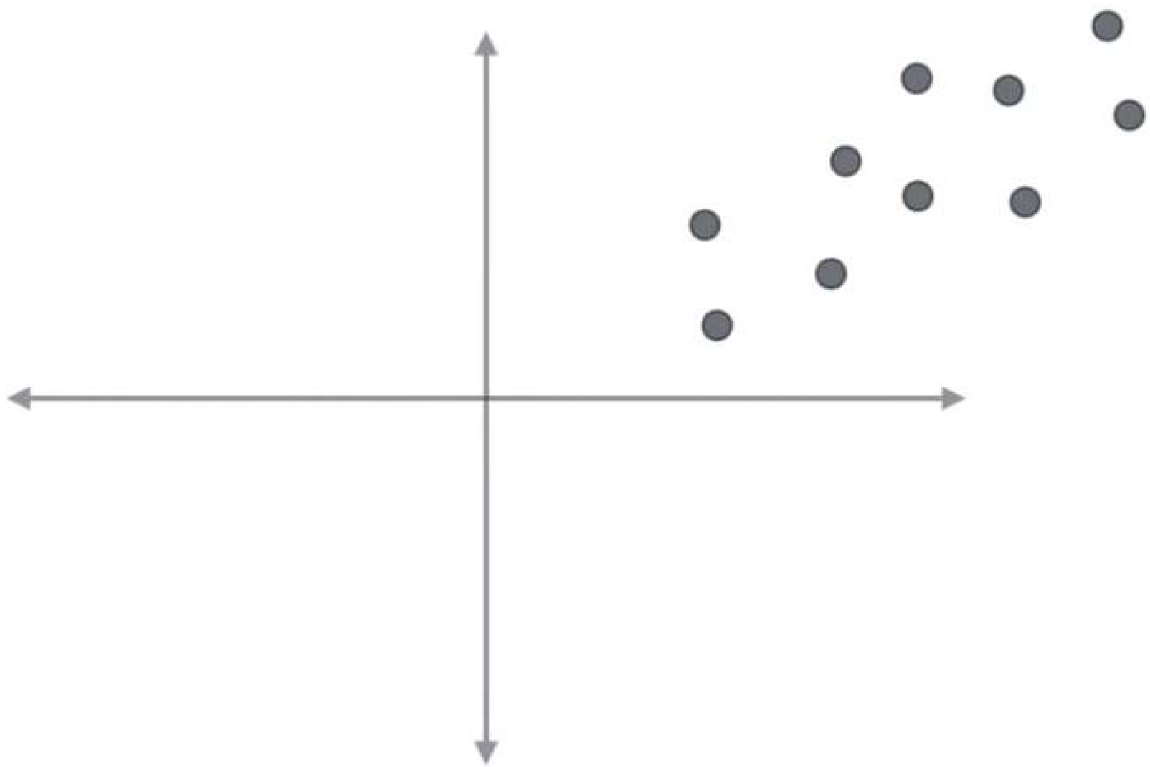
$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = 11\begin{pmatrix} u \\ v \end{pmatrix} \qquad \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = 1\begin{pmatrix} u \\ v \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

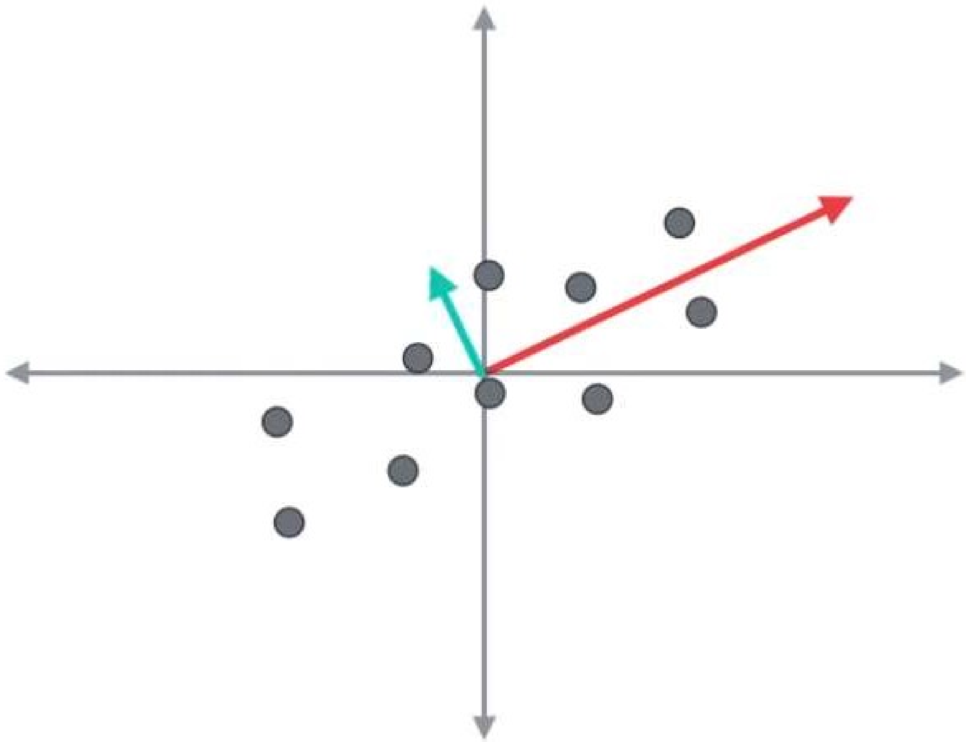
4, 2 ?

it is a direction vector,
doesn't make any difference

Principal Component Analysis (PCA)



Principal Component Analysis (PCA)



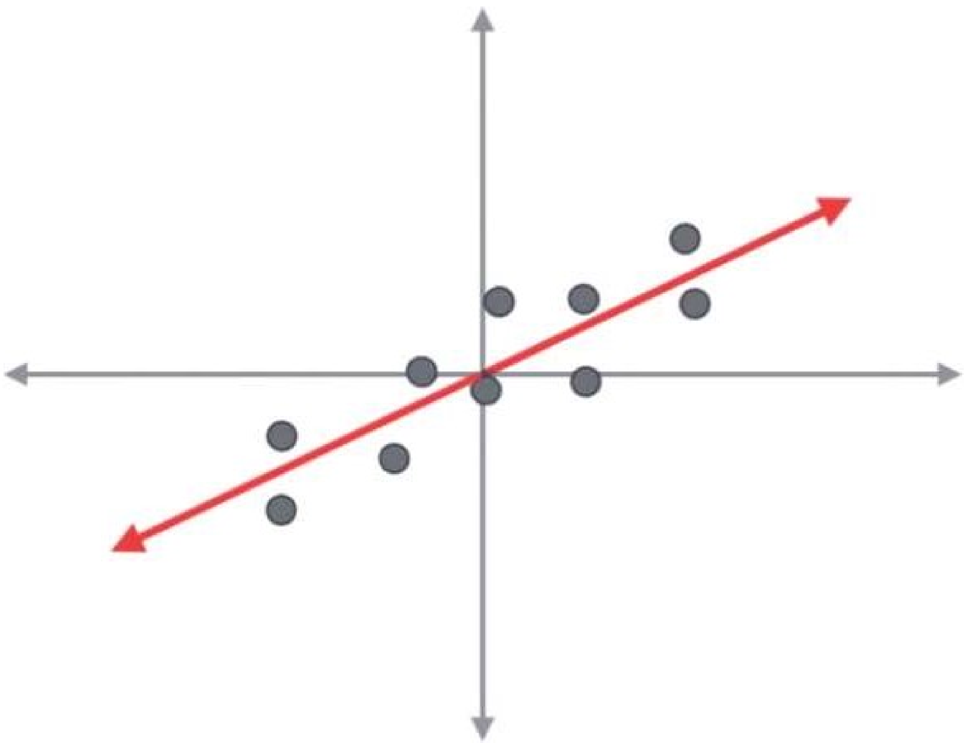
$$\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ Eigenvectors (direction)

11 1 Eigenvalues (magnitude)

For real symmetric matrix ($A=A^T$), eigenvectors are orthogonal

Principal Component Analysis (PCA)



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

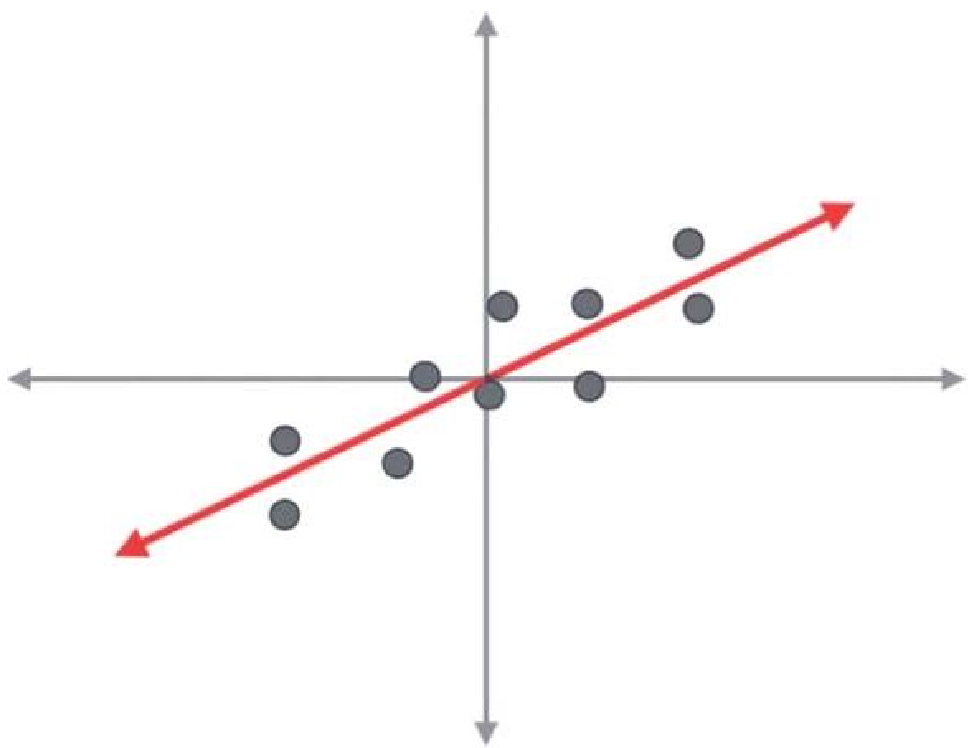
$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Eigenvectors (direction)

11

Eigenvalues (magnitude)

Principal Component Analysis (PCA)



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

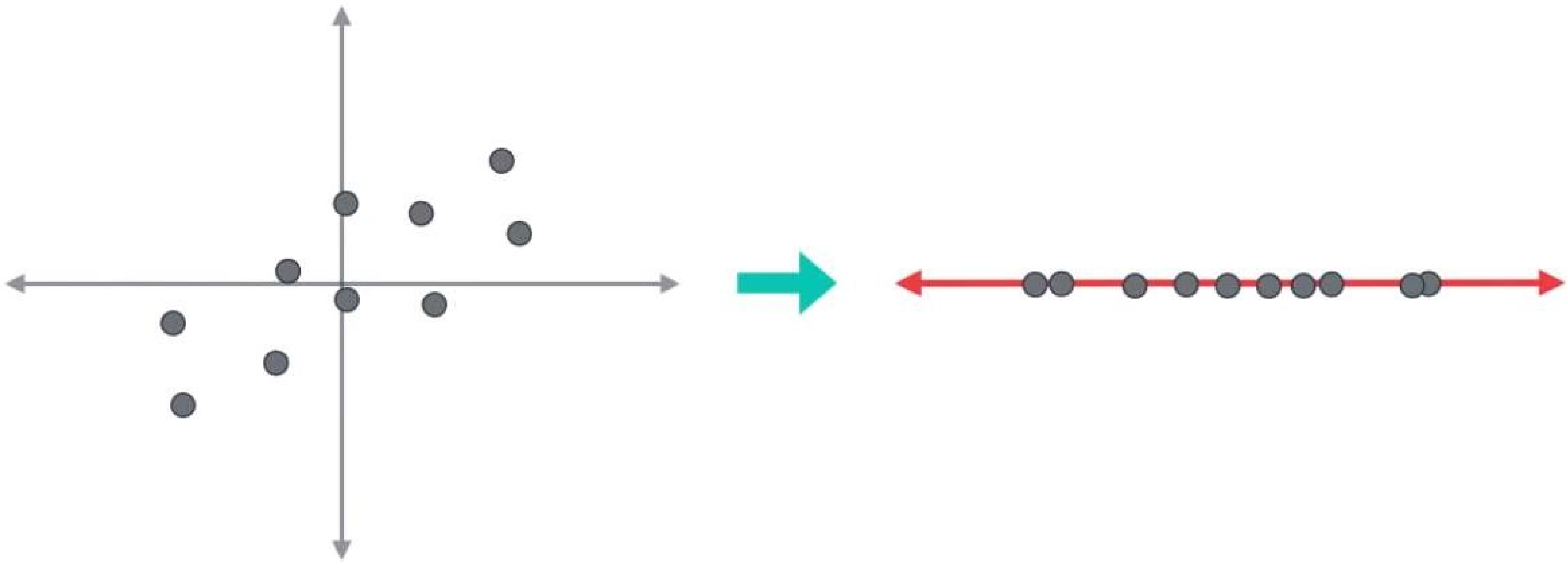
$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Eigenvectors
(direction)

$$11$$

Eigenvalues
(magnitude)

Principal Component Analysis (PCA)



PCA

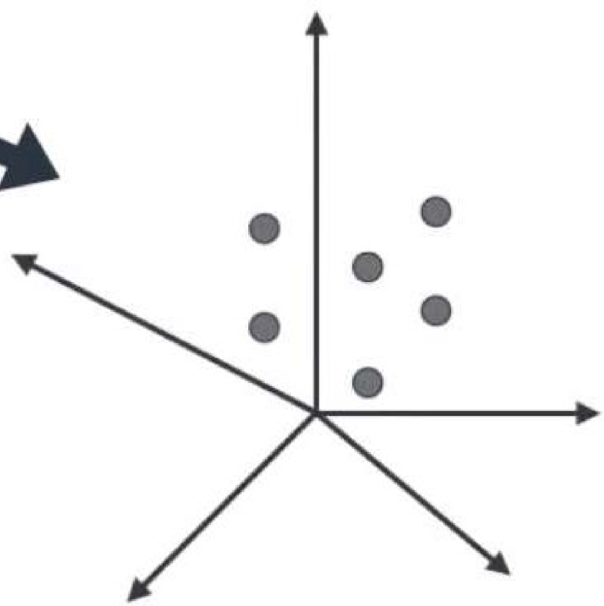
Large Table

X_1	X_2	X_3	X_4	X_5
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*

PCA

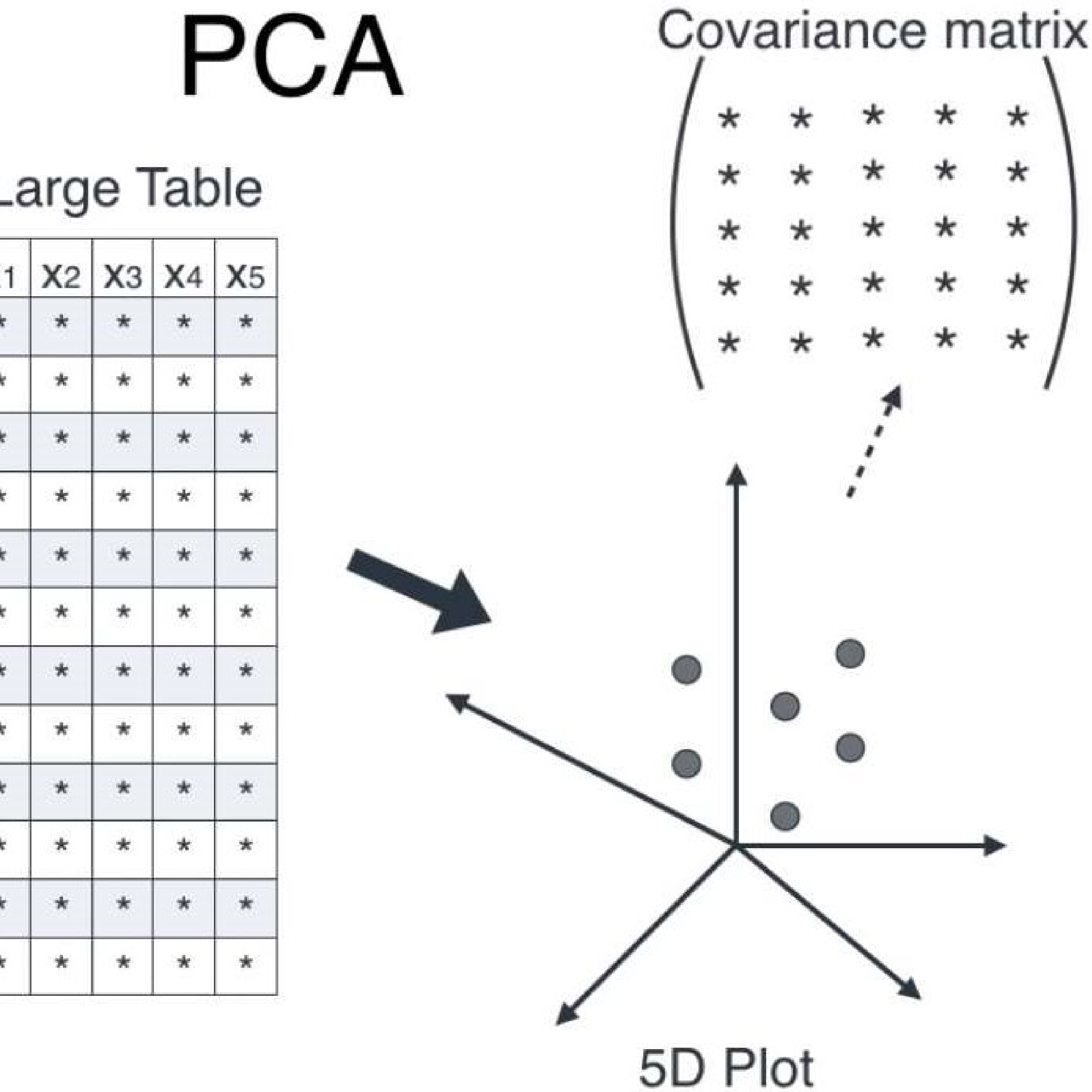
Large Table

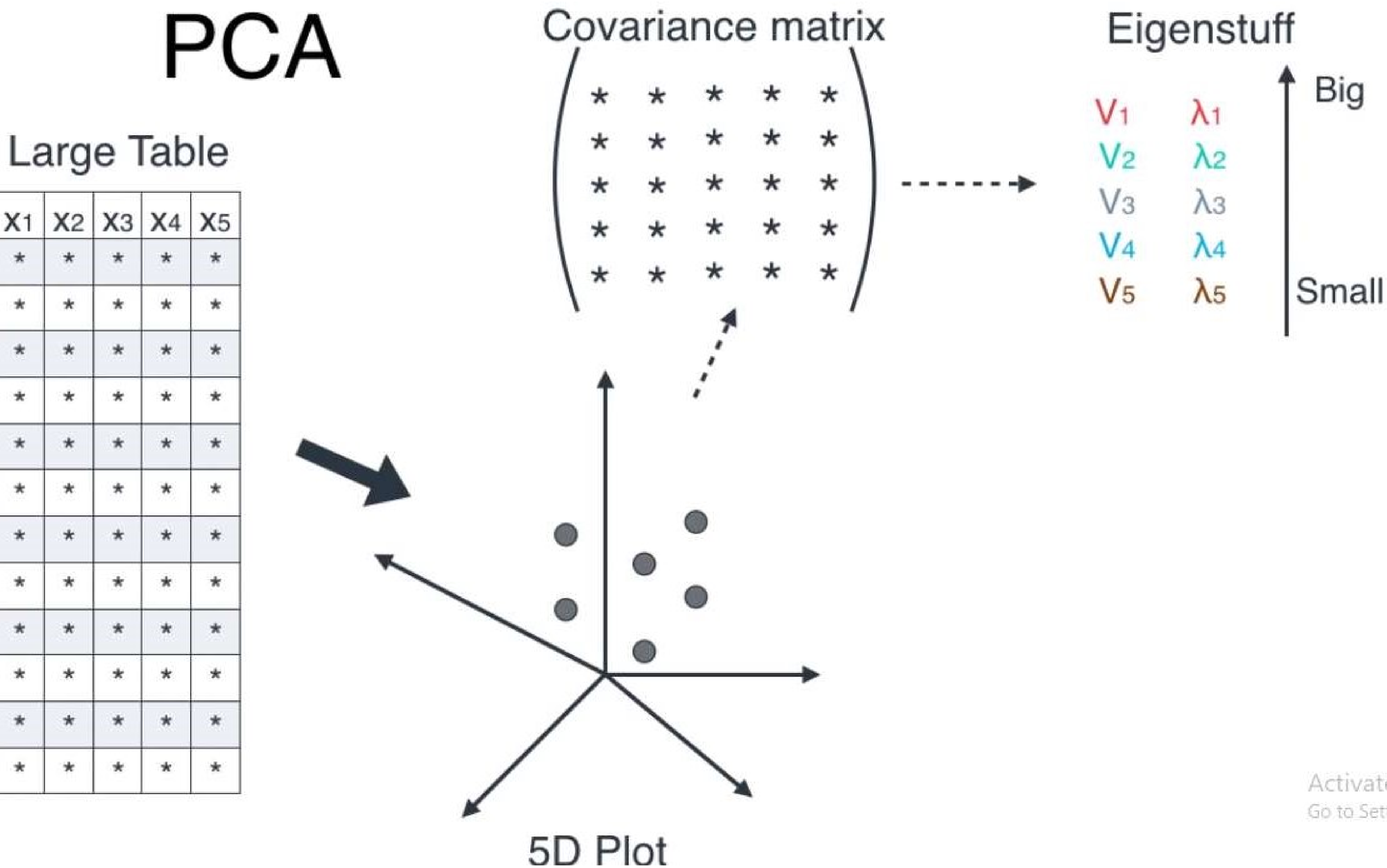
1	X2	X3	X4	X5
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*

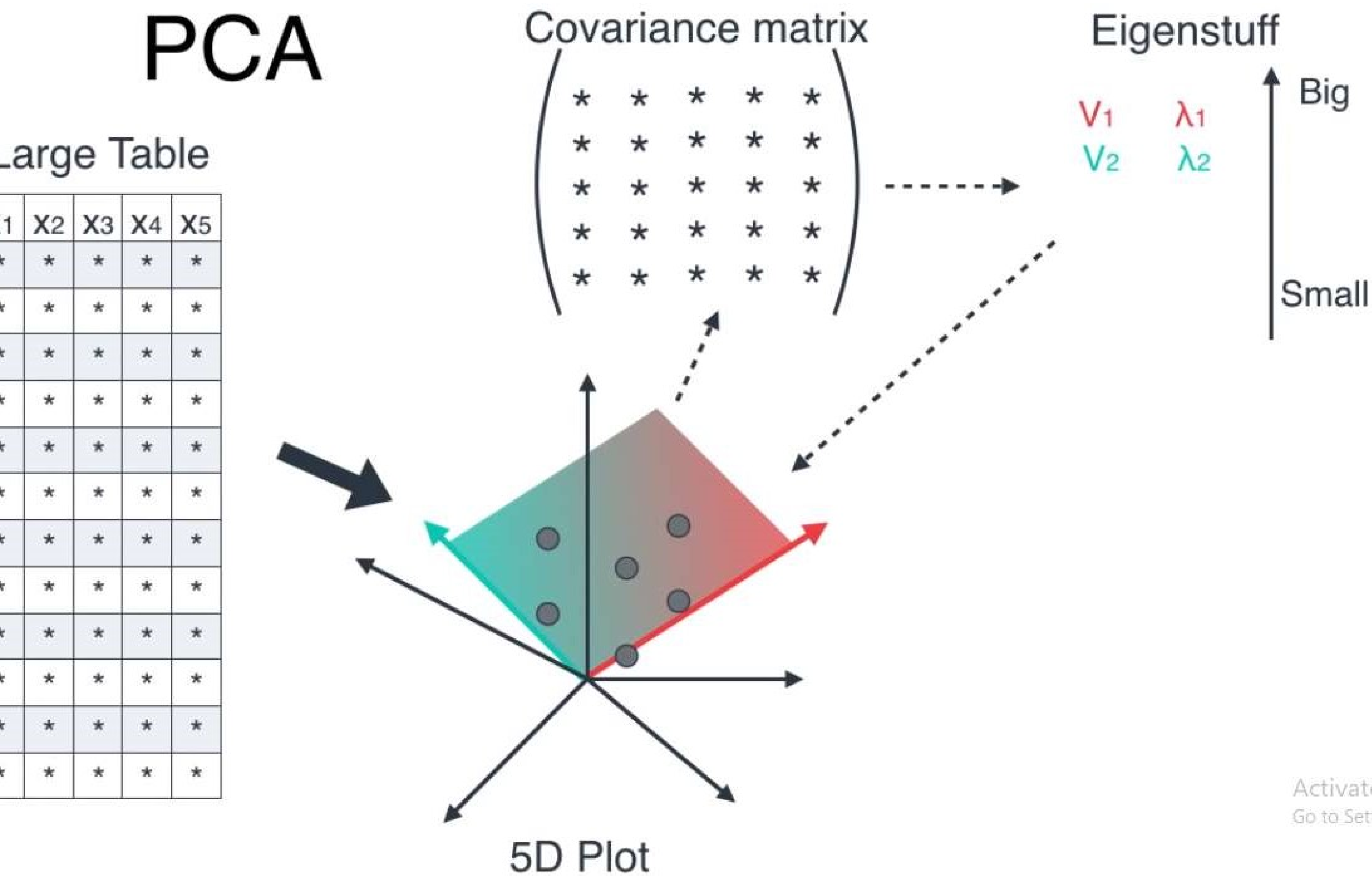


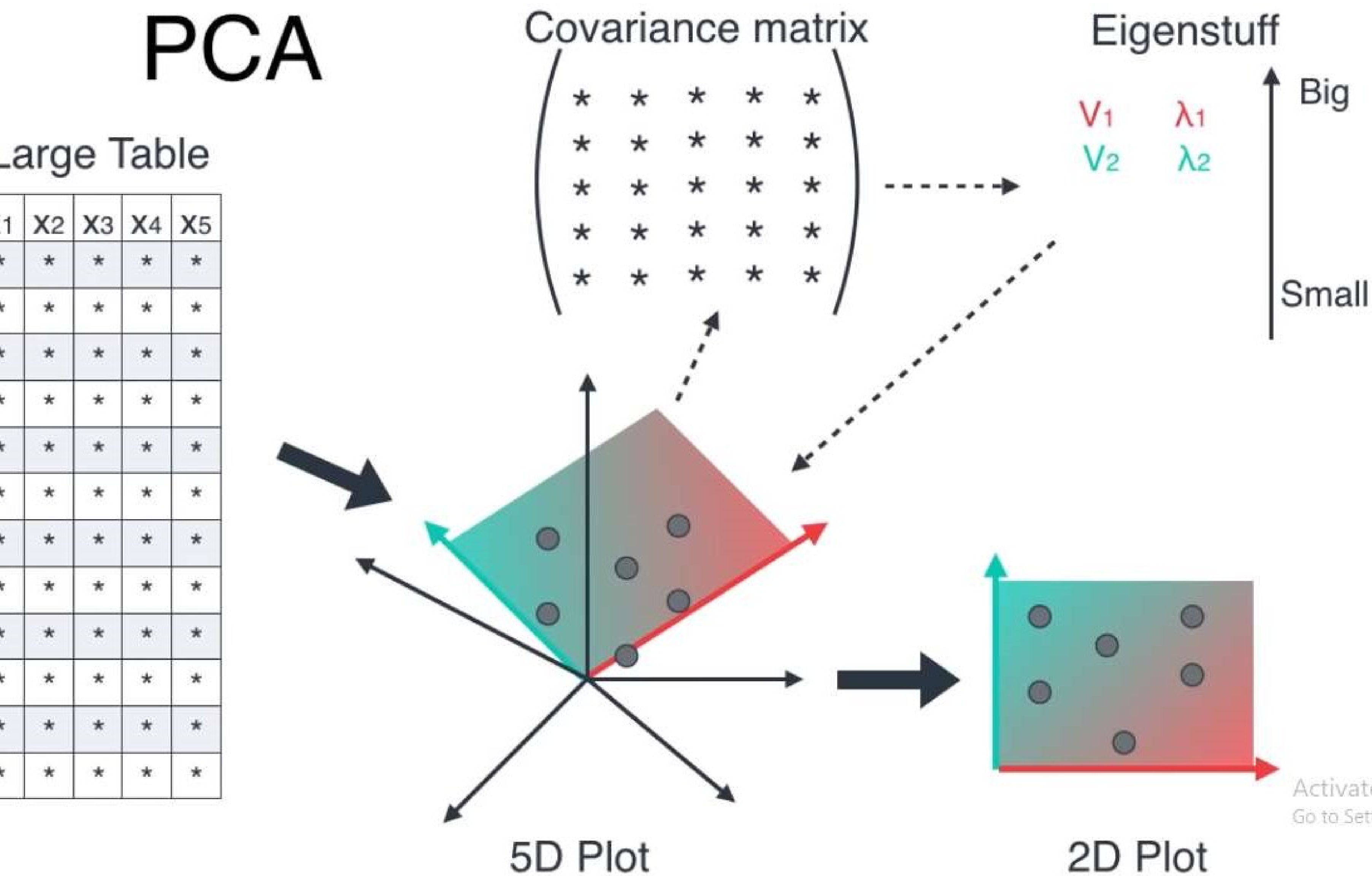
5D Plot

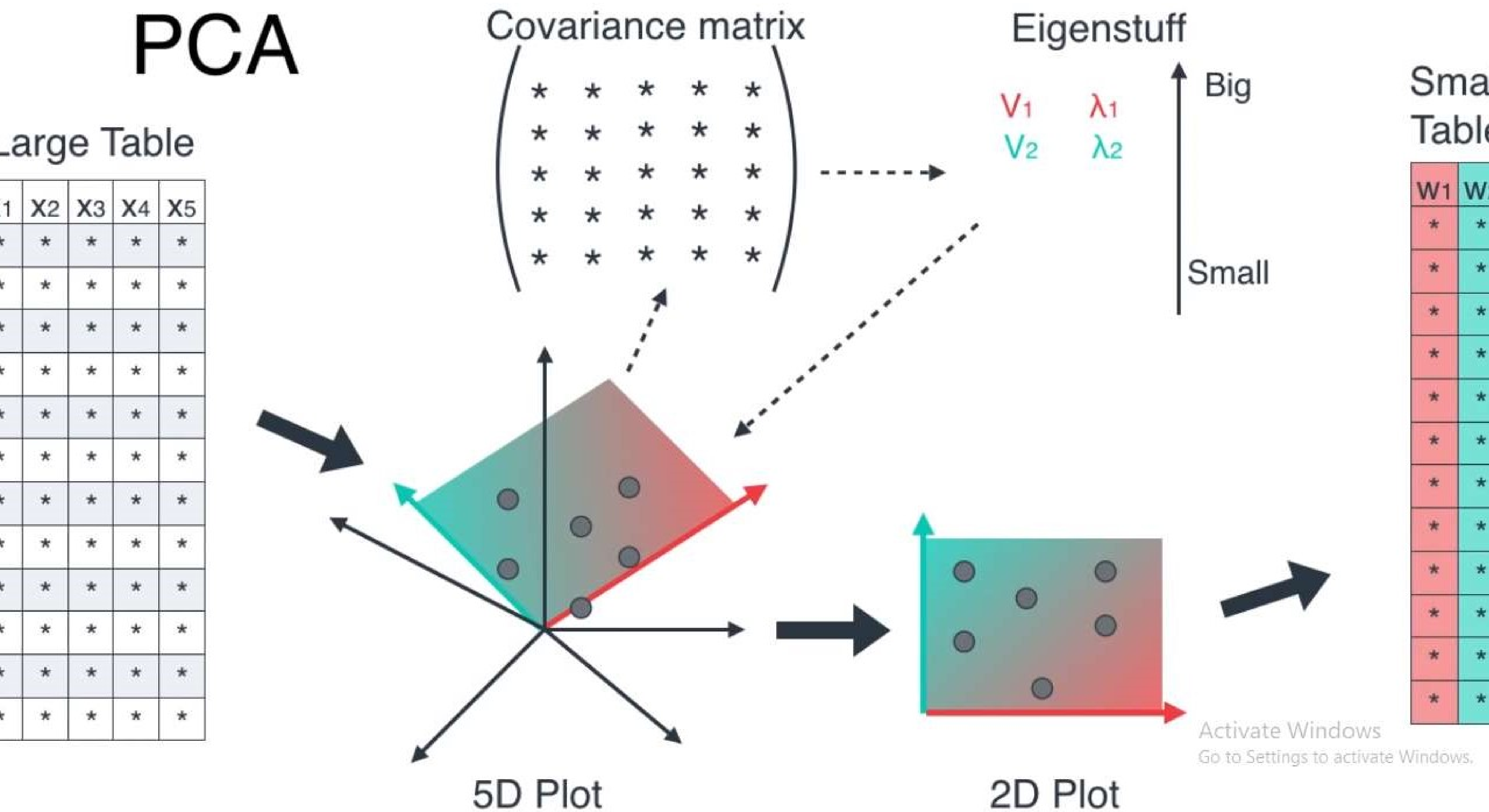
PCA




PCA
Large Table
X1 X2 X3 X4 X5
Covariance matrix
Eigenstuff
Big
V1 λ1
V2 λ2
V3 λ3
V4 λ4
V5 λ5
Small
5D Plot


PCA
Large Table
X2 X3 X4 X5
Covariance matrix
Eigenstuff
Big
V1 λ1
V2 λ2
Small
5D Plot


PCA
Large Table
X2 X3 X4 X5
Covariance matrix
Eigenstuff
Big
Small
V_1 λ_1
V_2 λ_2
5D Plot
2D Plot


PCA
Covariance matrix
Eigenstuff
Big
Small
V1 λ1
V2 λ2
X2 X3 X4 X5
W1
5D Plot
2D Plot

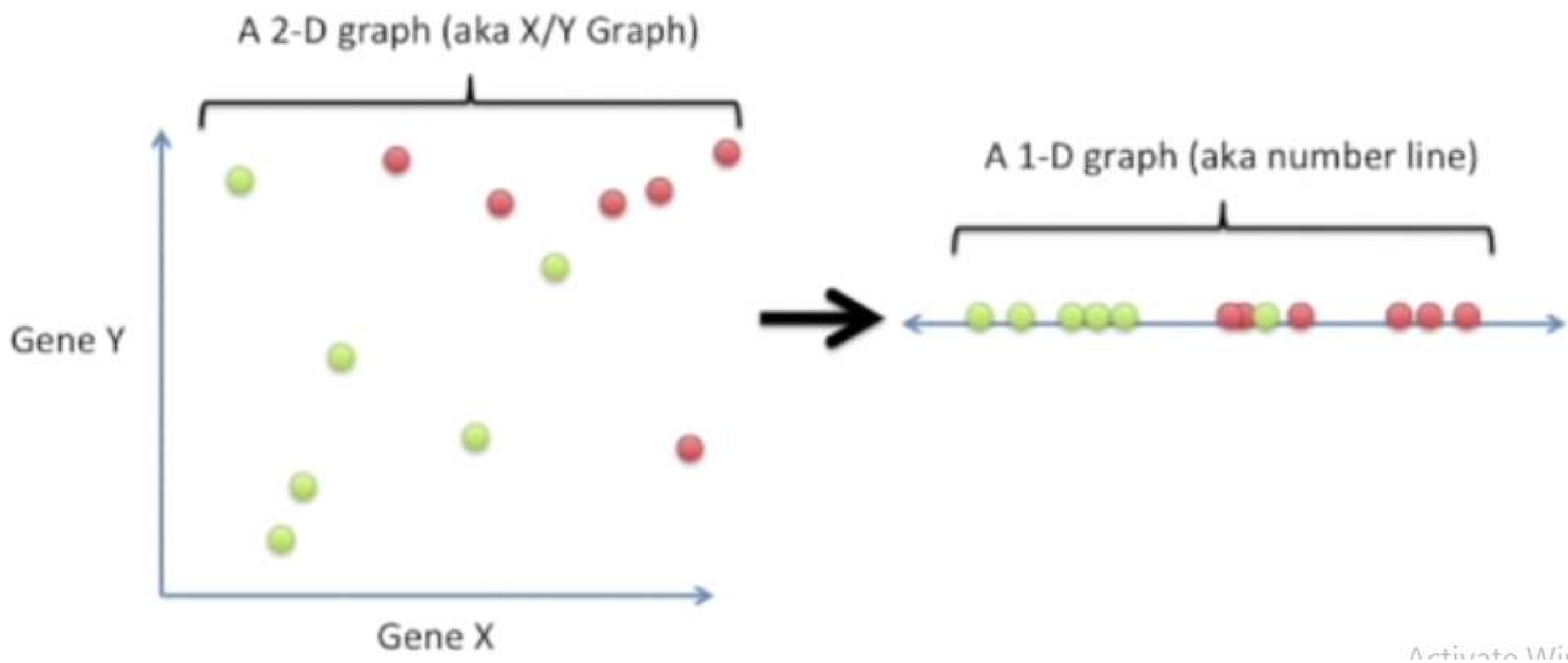
Methods

- PCA (Principal Component Analysis):
  - Find projection that maximize the variance
- ICA (Independent Component Analysis):
  - Very similar to PCA except that it assumes non-Guassian features
- Multidimensional Scaling:
  - Find projection that best preserves inter-point distances
- LDA(Linear Discriminant Analysis):
  - Maximizing the component axes for class-separation
- …
- …

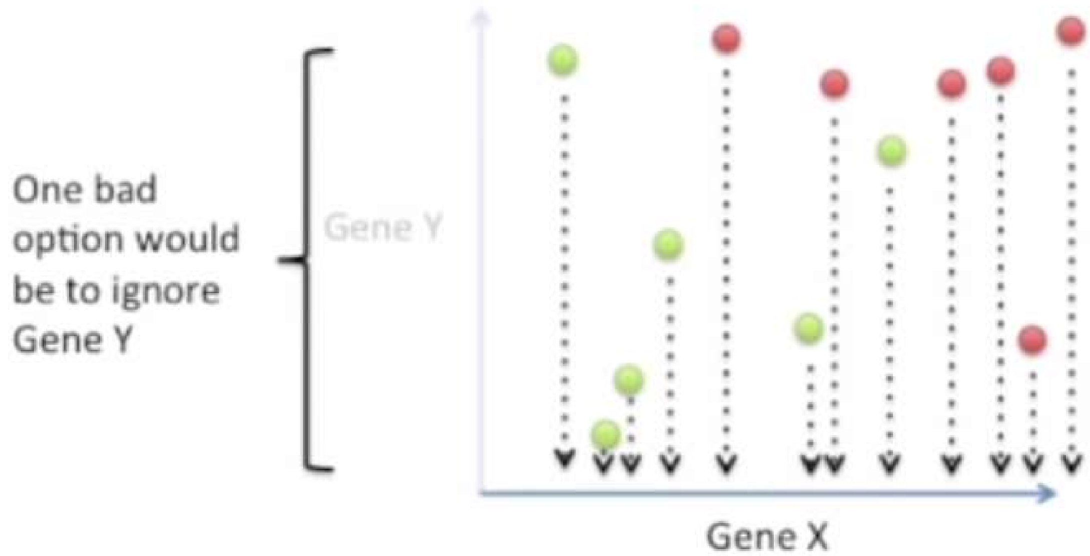
Linear Discriminant Analysis

- Linear Discriminant Analysis (LDA) is like PCA, but it focuses on maximizing the separability among known categories.

Reducing 2D to 1D

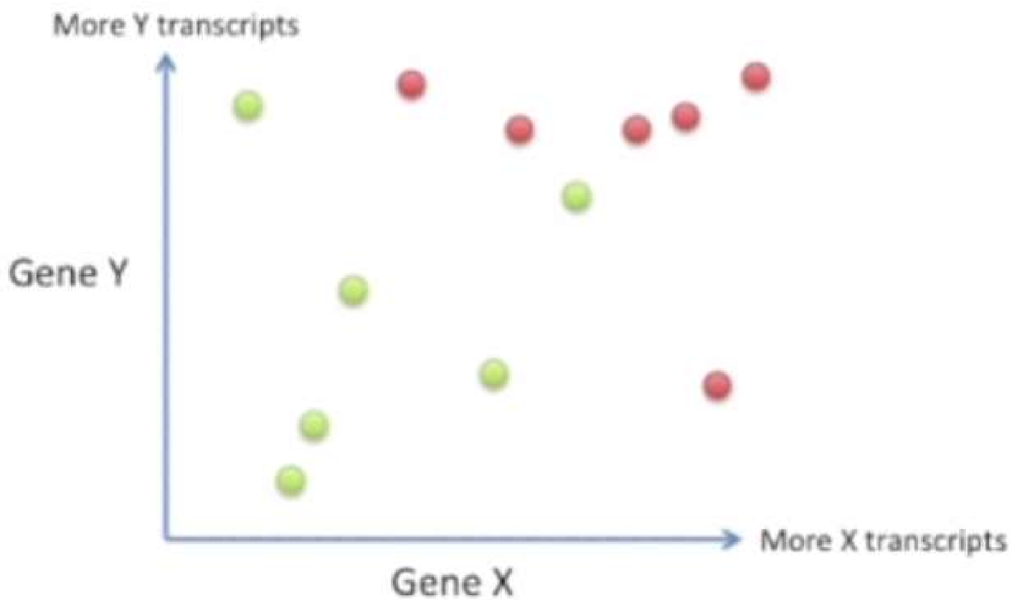


Reducing 2D to 1D

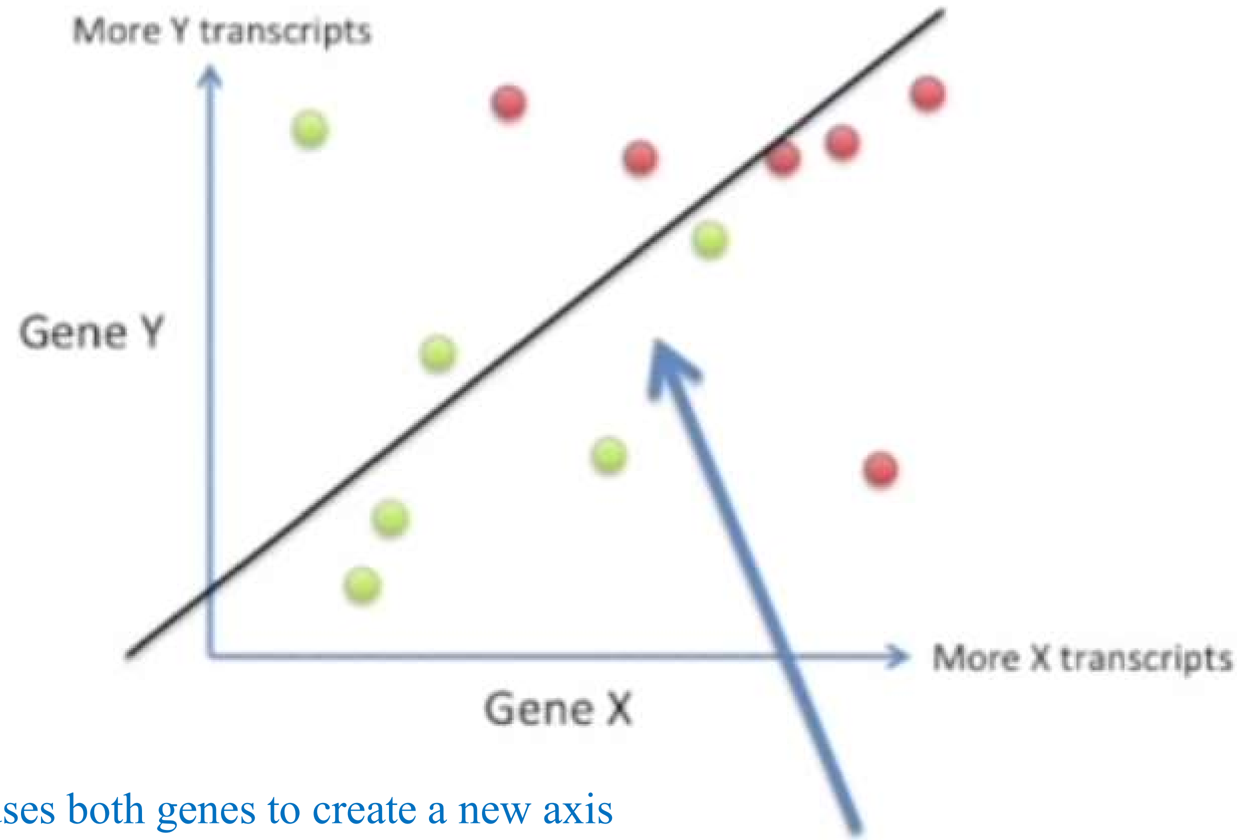


This way is bad because, it ignores the useful information that Gene Y provides…
Projecting the genes onto the Y axis.(i.e. ignoring the Gene X) isn't any better.

LDA provides a better way

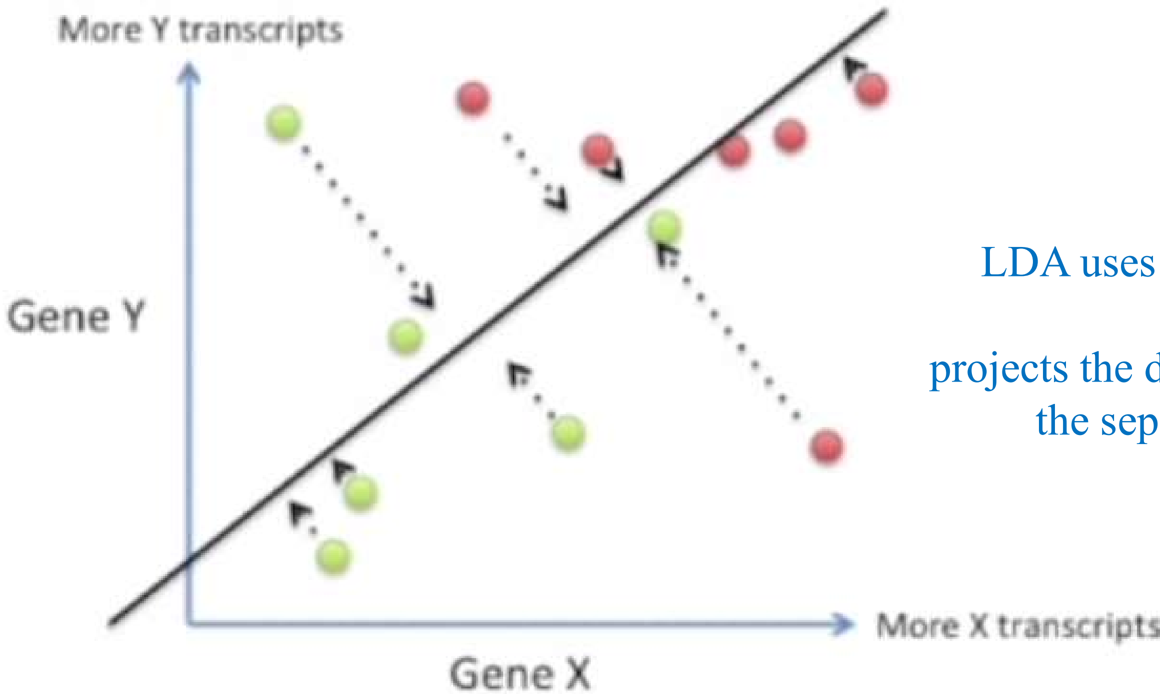


Reducing 2D to 1D using LDA



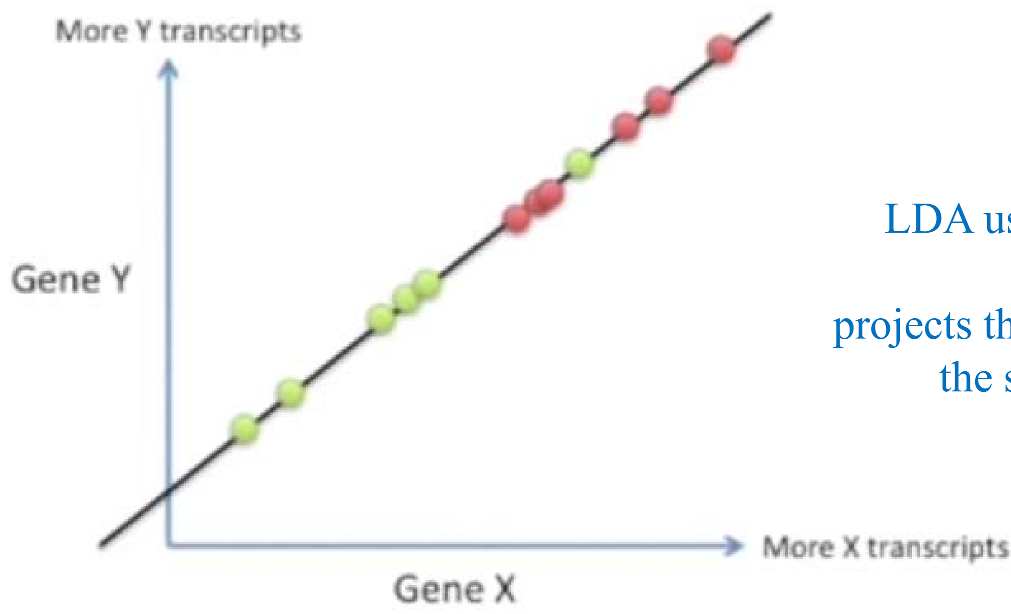
LDA uses both genes to create a new axis

Reducing 2D to 1D using LDA



LDA uses both genes to create a new axis
and
projects the data onto this new axis to maximize the separation of the two categories

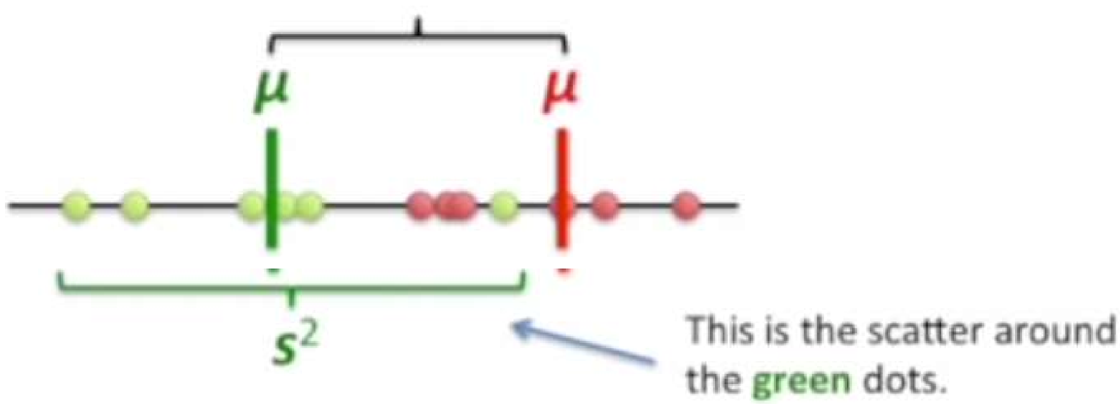
Reducing 2D to 1D using LDA



LDA uses both genes to create a new axis
and
projects the data onto this new axis to maximize
the separation of the two categories

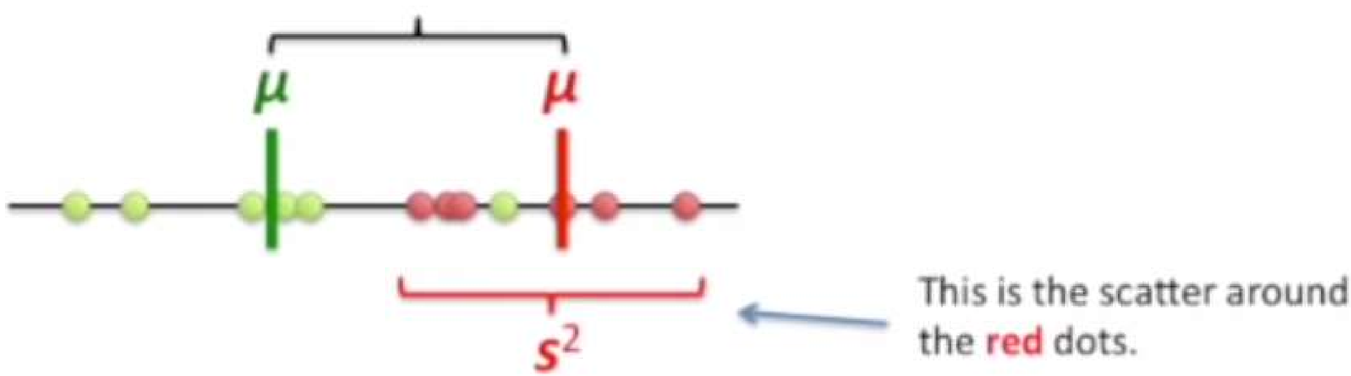
How LDA creates the new axis?

- The new axis is created according to the two criteria (considered simultaneously)
  - Maximize the distance between means.
  - Maximize the variation (“scatter (s2) as per LDA) within each category

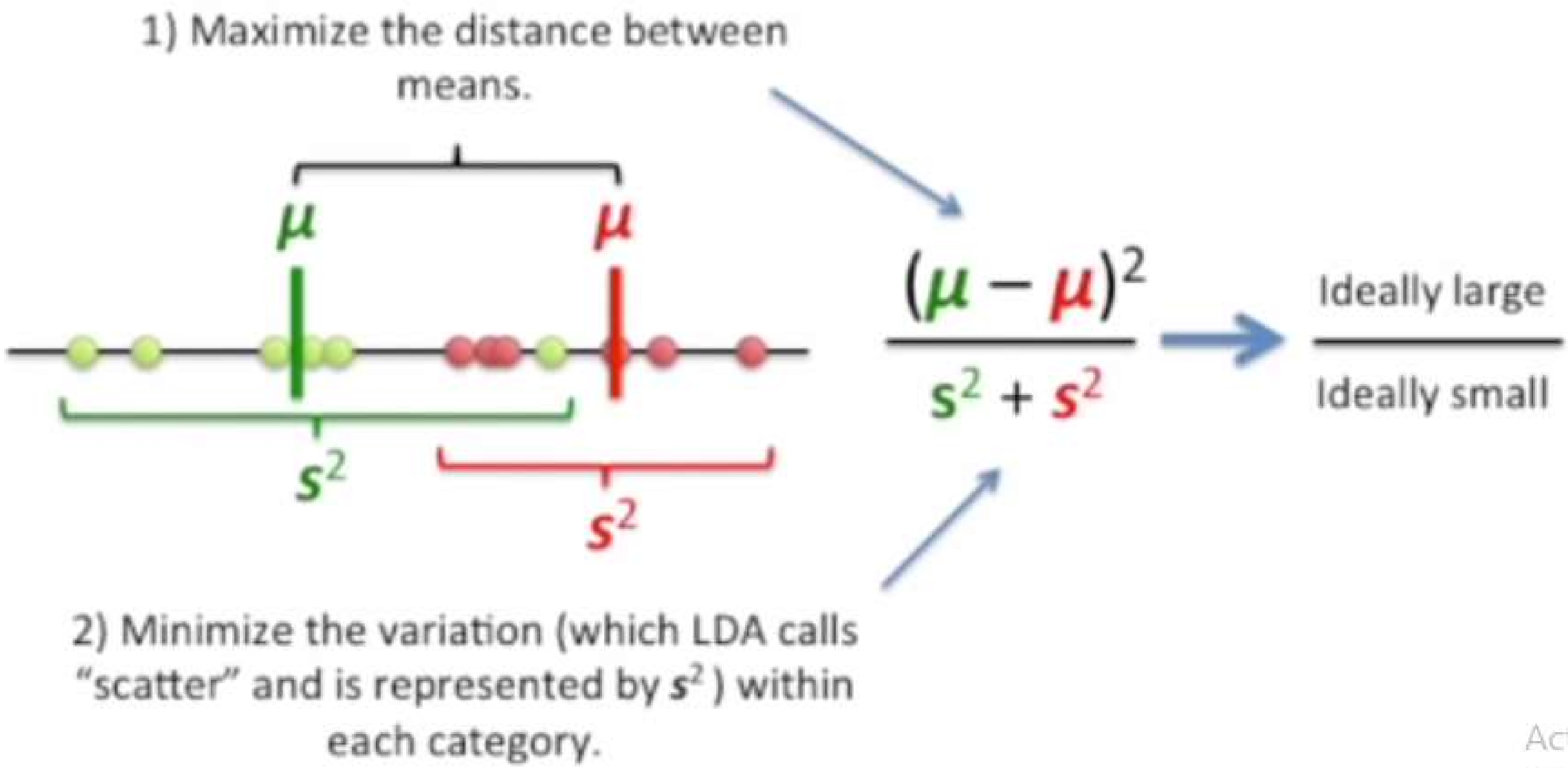


How LDA creates the new axis?

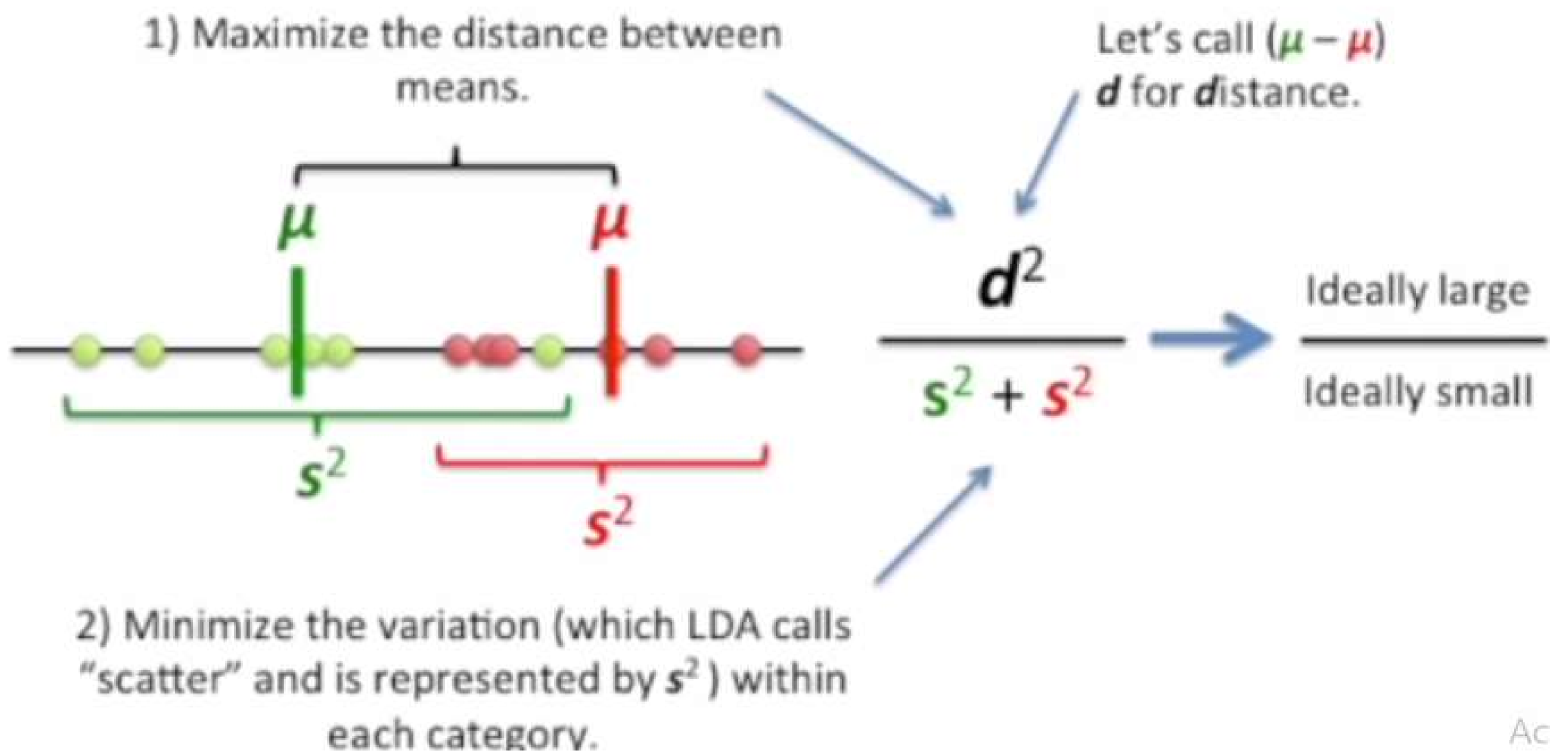
- The new axis is created according to the two criteria (considered simultaneously)
  - Maximize the distance between means.
  - Maximize the variation (“scatter (s2) as per LDA) within each category



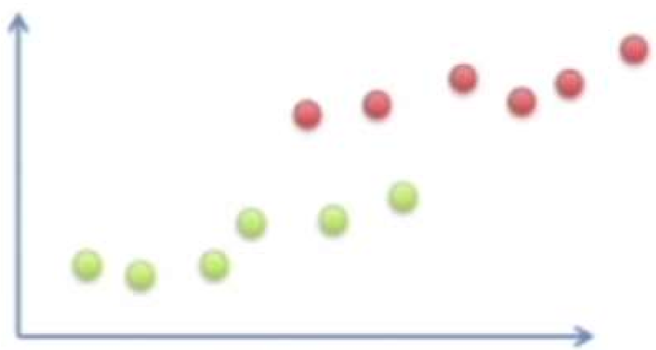
How LDA creates the new axis?



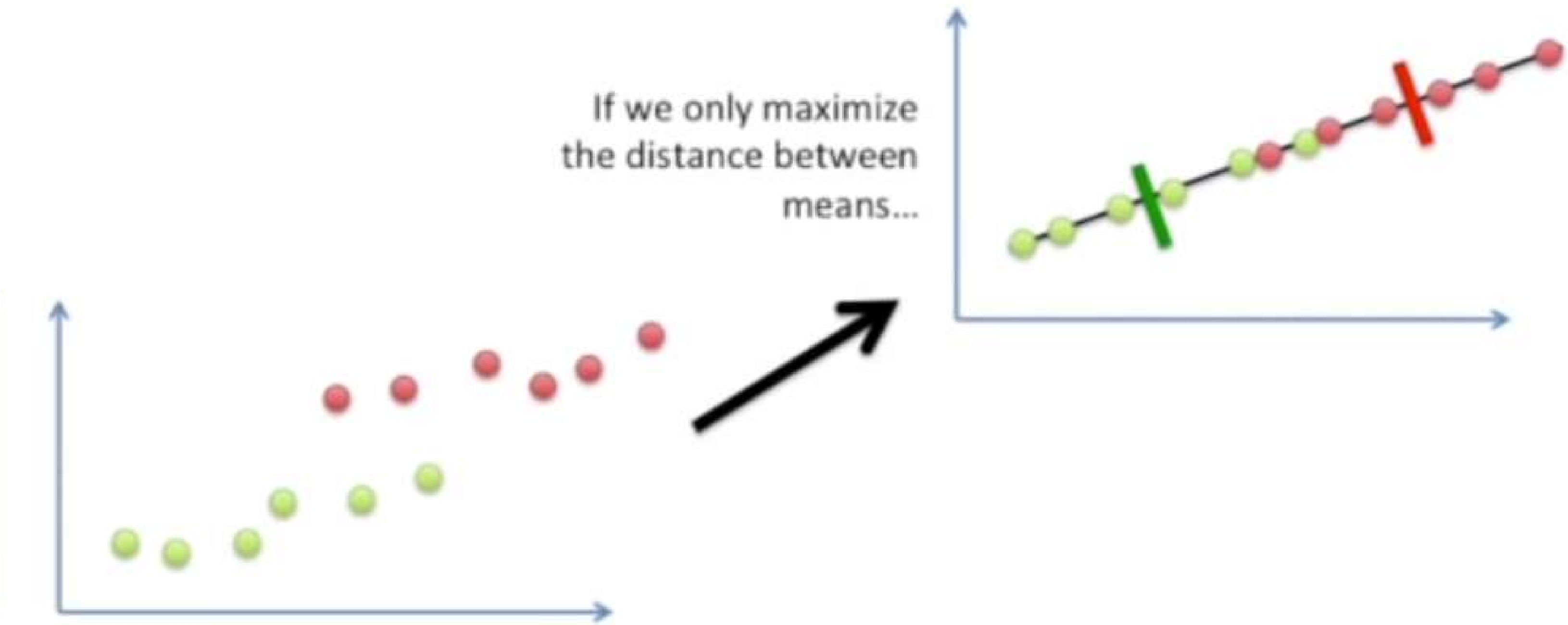
How LDA creates the new axis?



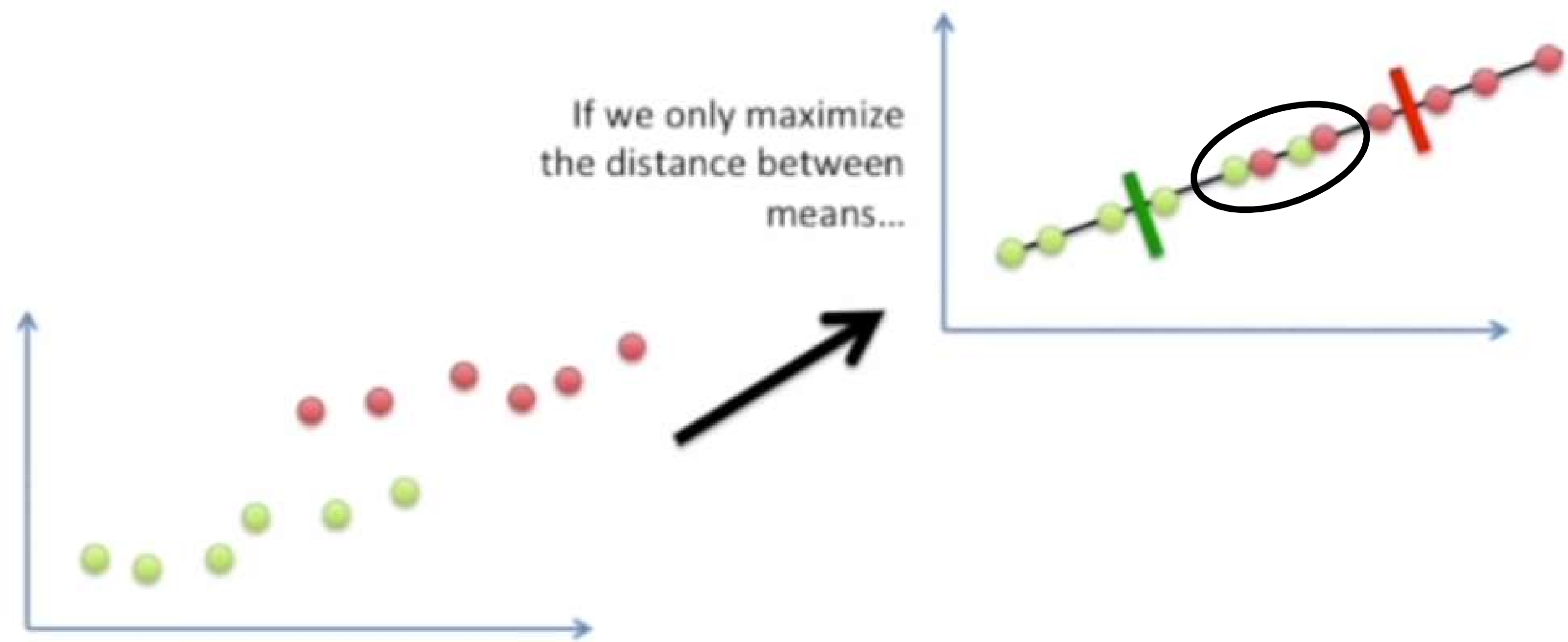
Why both distance and scatter are important?



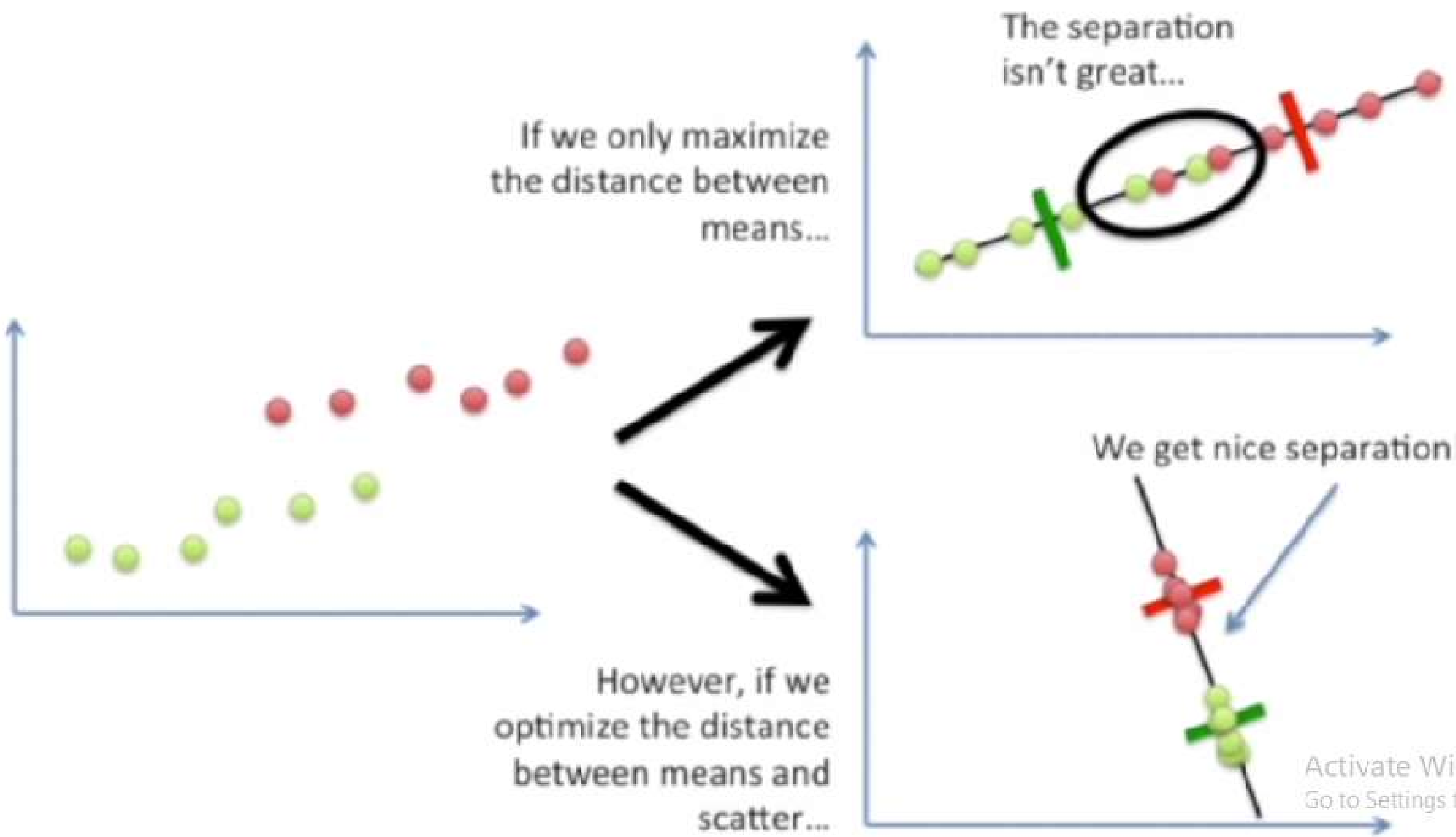
Why both distance and scatter are important?



Why both distance and scatter are important?



Why both distance and scatter are important?



There will be significant overlap between the classes in terms of their variances (scatter).
This means that even if the means are well-separated, there may be points that are close to one mean but belong to the other class due to their high variance.

we also consider the scatter or variance within each class to ensure that the classification boundary is not affected by the spread of the data points

What if we have more than 2 genes (more than 2 dimensions)?

- The process is the same.
  - Create an axis that maximizes the distance between the means for the two categories and minimizing the scatter.

?

what if more than two classes.

Questions?

Thank you