Eigenvalue and Eigenvectors, Covariance Matrix, & PCA

CE642: Social and Economic Networks
Maryam Ramezani
Sharif University of Technology
maryam.ramezani@sharif.edu

01

Linear Algebra Review

Eigenvalue & Eigenvector

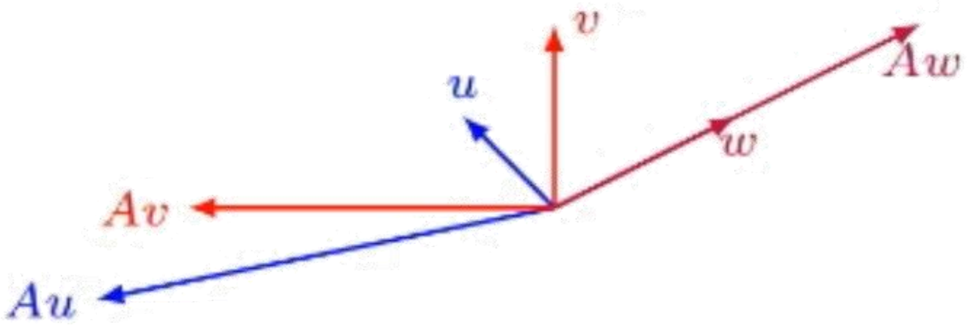
Motivation

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

  u = $\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Rightarrow$ Au = $\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$

  v = $\begin{bmatrix} 0 \\ 2 \end{bmatrix} \Rightarrow$ Av = $\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 0 \end{bmatrix}$

  w = $\begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow$ Aw = $\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$



- Vector “w” keeps the straight, but changes the scale.

Definition

Definition

An **eigenvector** of a square $n \times n$ matrix A is nonzero vector v such that $Av = \lambda v$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution v of $Av = \lambda v$; such an v is called an *eigenvector corresponding to* λ.

- **An eigenvector must be nonzero, by definition, but an eigenvalue may be zero.**

Example

- $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\lambda = 2$
- Show that 7 is an eigenvalue of matrix B, and find the corresponding eigenvectors.

$$B = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$$

Eigenspace

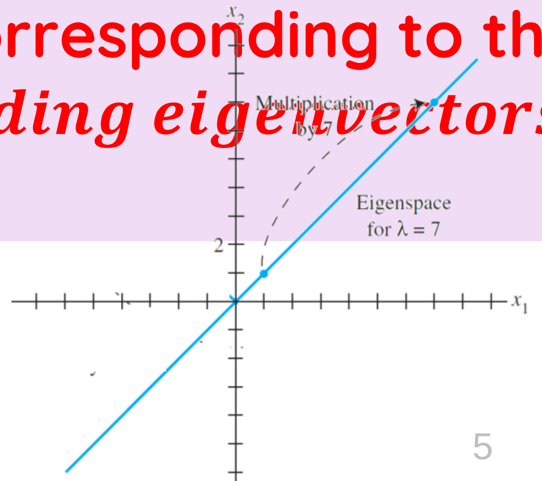
Note

λ is an eigenvalue of an $n \times n$ matrix:

$$Av = \lambda v \Rightarrow (A - \lambda I)v = 0$$

The set of all solutions of above is just the null space of the matrix $A - \lambda I$. So this set is the *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ . The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

Eigenspace: A vector space formed by eigenvectors corresponding to the same eigenvalue and the origin point. $span\{corresponding\ eigenvectors\}$



Definitions

Theorem

Let A be a $m \times n$ matrix:

$$Nullity(A) + Rank(A) = n$$

Note

- $Av = \lambda v \Rightarrow Av - \lambda vI = 0 \Rightarrow (\boldsymbol{A} - \boldsymbol{\lambda I})\boldsymbol{v} = \boldsymbol{0} \;\; \boldsymbol{v} \neq \boldsymbol{0}$
  - $\boldsymbol{v} \in \boldsymbol{N}(\boldsymbol{A} - \boldsymbol{\lambda I})$
  - $\boldsymbol{A} - \boldsymbol{\lambda I}$ **must be singular.**
  - Proof that for finding the eigenvalue we should solve the determinate zero equation. Look at nullspace, rank and nullity theorem, singular matrix, and det zero!
- **Characteristic polynomial** $\det(A - \lambda \mathrm{I})$
- **Characteristic equation** $\det(A - \lambda \mathrm{I}) = 0$
- If λ is an eigenvalue of A, then the subspace E_λ= {span{v} | Av = λv} is called the eigenspace of A associated with λ. (This subspace contains all the span of eigenvectors with eigenvalue λ, and also the zero vector.)
- **Eigenvector is basis for eigenspace.**
- Set of all eigenvalues of matrix is $\sigma(A)$ *named spectrum of a matrix*

Definitions

Note

- Instead of $\det(A - \lambda I)$, we will compute $\mathbf{det}(\boldsymbol{\lambda}\mathbf{I} - \boldsymbol{A})$**.** Why?
  - $\det(A - \lambda I) = (-1)^{\mathrm{n}} \det(\lambda I - \mathrm{A})$
  - Matrix $n \times n$ with real values has …… eigenvalues.

Finding Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix.

1. First, find the eigenvalues λ of A by solving the equation $\det(\lambda I - A) = 0$.
2. For each λ, find the basic eigenvectors $X \neq 0$ by finding the basic solutions to $(\lambda I - A) X = 0$.

To verify your work, make sure that $AX = \lambda X$ for each λ and associated eigenvector X.

Example

Find eigenvalues and eigenvectors, eigenspace (E), and *spectrum* of matrix $A = \begin{bmatrix} \mathbf{3} & \mathbf{-2} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$:

$$\det(A - \lambda \mathrm{I}) = \begin{vmatrix} 3-\lambda & -2 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 3\lambda + 2 = 0 \Rightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_2 = 2 \end{cases}$$

$$\left.\begin{matrix} \lambda_1 = 1 \\ (A - \lambda_1 I)q_1 = 0 \end{matrix}\right\} \Rightarrow q_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\left.\begin{matrix} \lambda_2 = 2 \\ (A - \lambda_2 I)q_2 = 0 \end{matrix}\right\} \Rightarrow q_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Eigenvalues={1,2}

Eigenvectors=$\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\}$

$E_1(A) = span\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$ $E_2(A) = span\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\}$

$\sigma(A)$={1,2}

$$\boldsymbol{AQ = Q\Lambda} \Rightarrow \begin{bmatrix} \mathbf{3} & \mathbf{-2} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{1} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{2} \end{bmatrix}$$

Eigenvalues Properties

- Are eigenvectors unique?
  - If $\mathbf{v}$ is an eigenvector, then $\beta\mathbf{v}$ is also an eigenvector

$$A(\beta v) = \beta(Av) = \beta(\lambda v) = \lambda(\beta v)$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

- For a 2 × 2 matrix, this is a simple quadratic equation with two solutions (maybe complex)

$$\lambda = (a_{11} + a_{22}) \pm \sqrt{\frac{(a_{11} + a_{22})^2}{4(a_{11}a_{22} - a_{12}a_{21})}}$$

Eigenvalues Properties

- If A is an n × n matrix:
  - The sum of the n eigenvalues of A is the trace of A.
  - The product of the n eigenvalues is the determinant of A.
  - $0 \in \sigma(A) \Leftrightarrow |A| = 0$
  - If A is symmetric, then any two eigenvectors from different eigenspace are orthogonal.

$$\left.\begin{array}{c} Av_1 = \lambda_1 v_1 \\ Av_2 = \lambda_2 v_2 \\ \lambda_1 \neq \lambda_2 \end{array}\right\} \Rightarrow \; {v_1}^T v_2 = 0$$

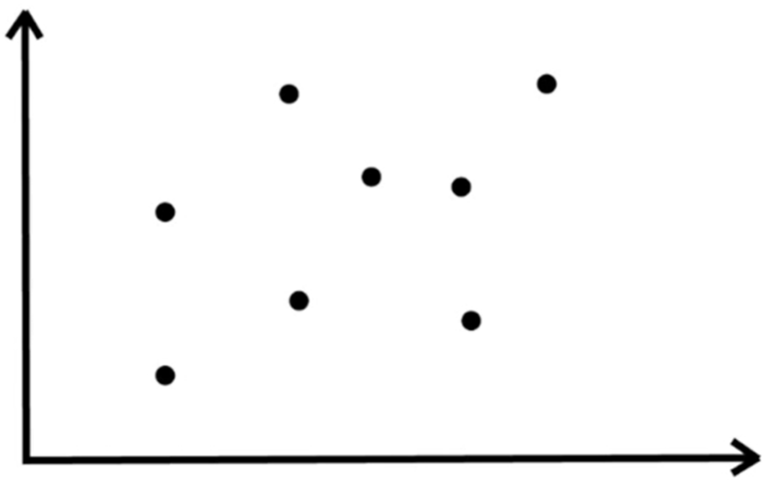
  - If A is symmetric, it has exactly n (not necessarily distinct) real eigenvalues.

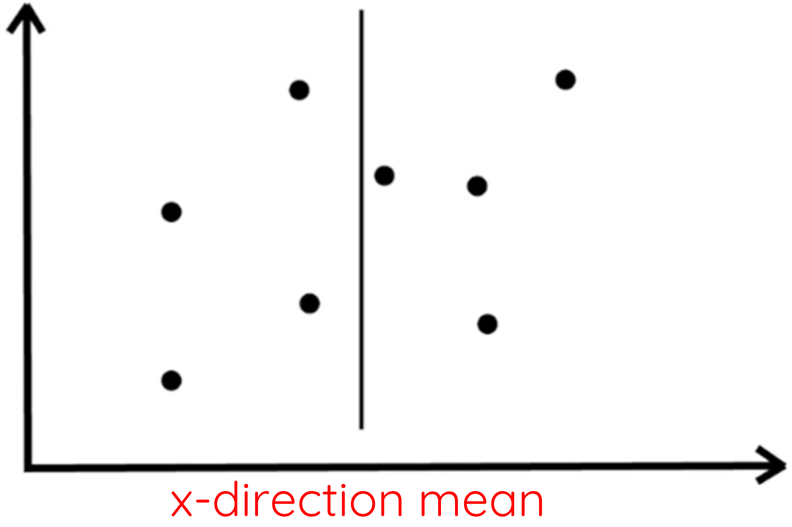
02

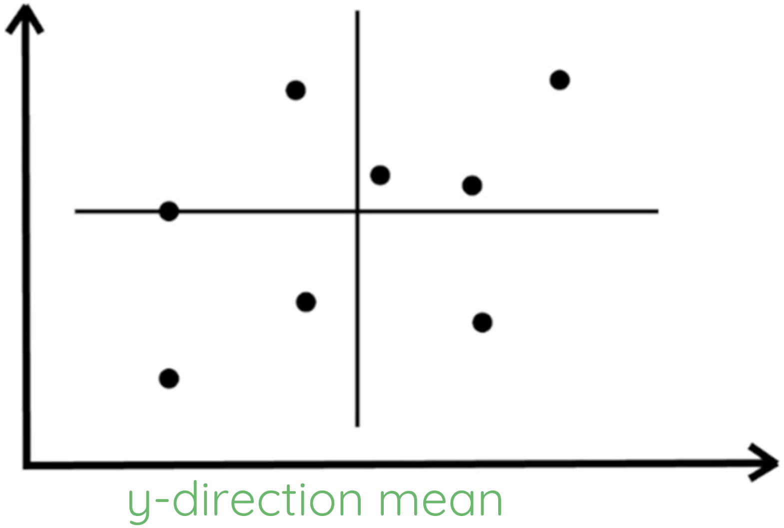
Covariance Matrix

Covariance Matrix

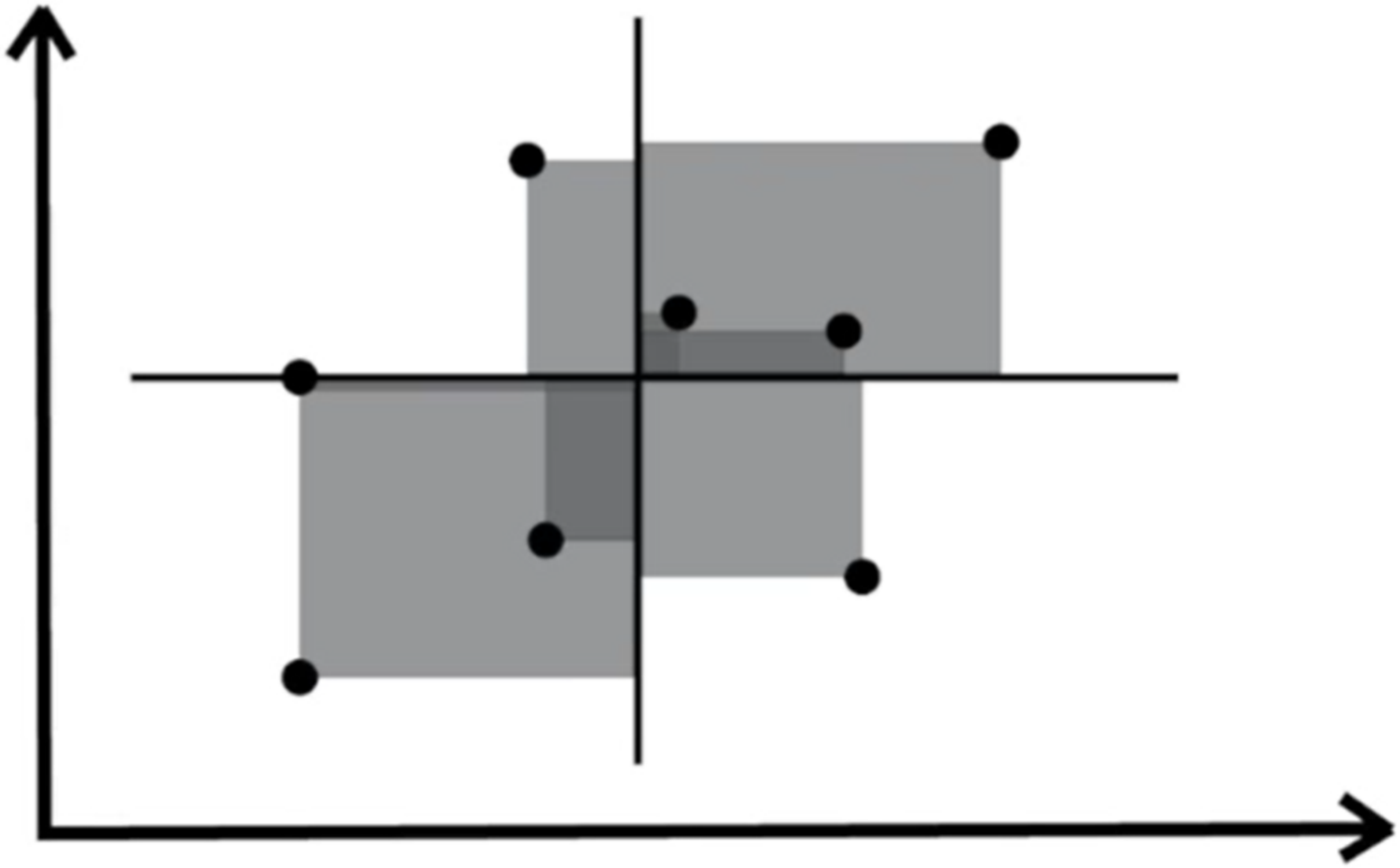
- Measures how much two variables change together.
- Look at how much is the distance of each point from the x-direction mean & y-direction mean.



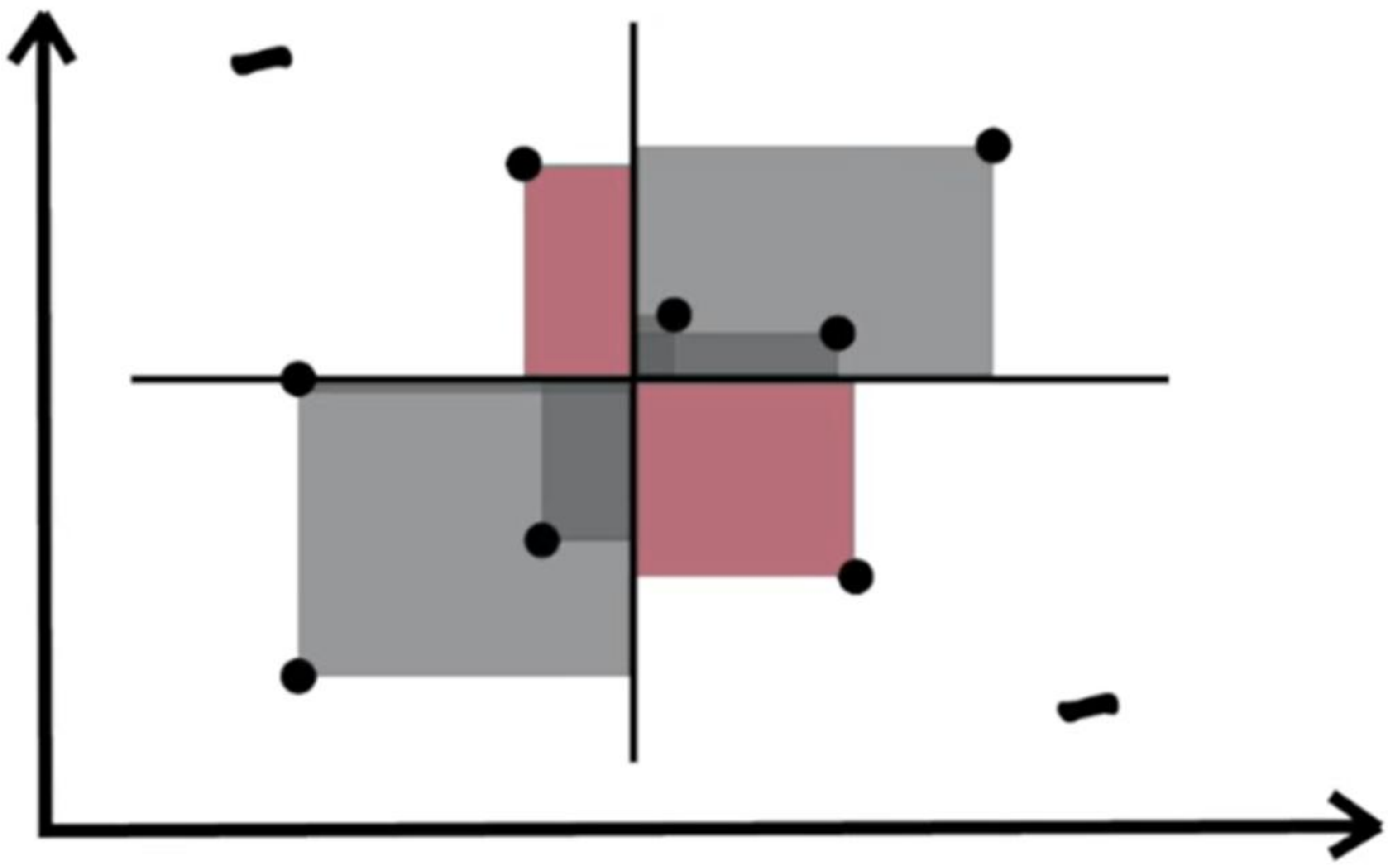




Covariance Matrix

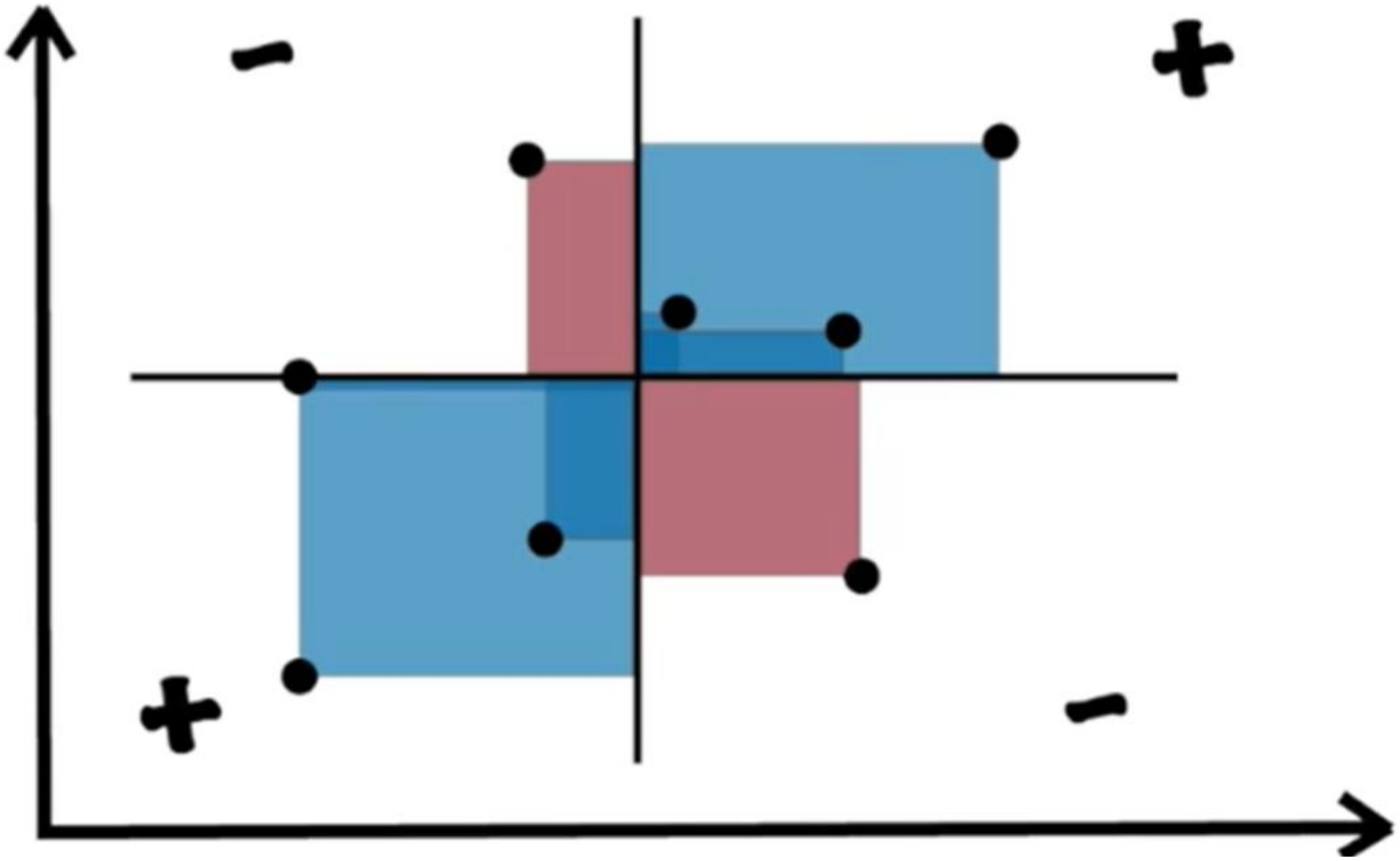


Covariance Matrix



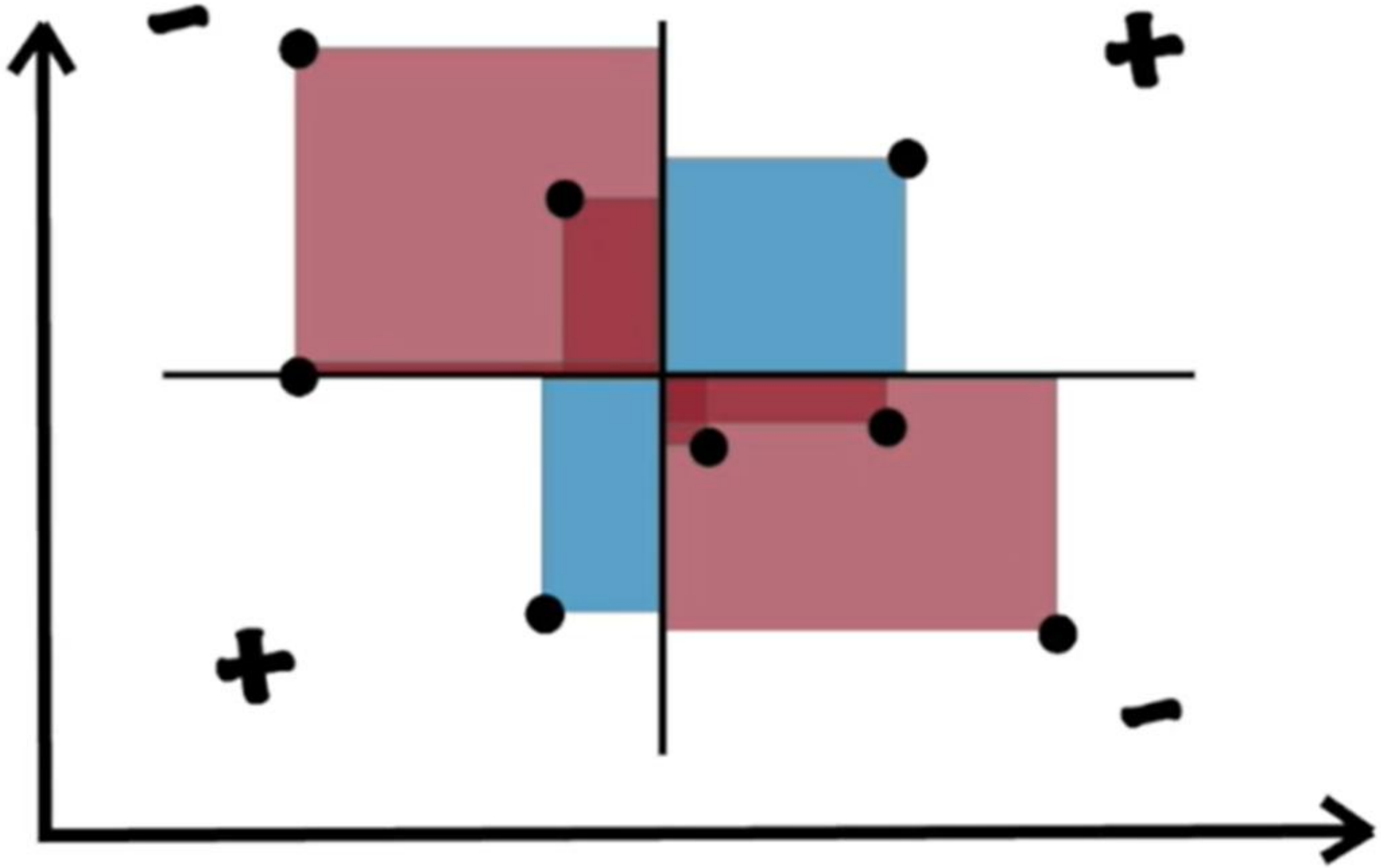
Covariance Matrix

$$\text{Cov}(X,Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$



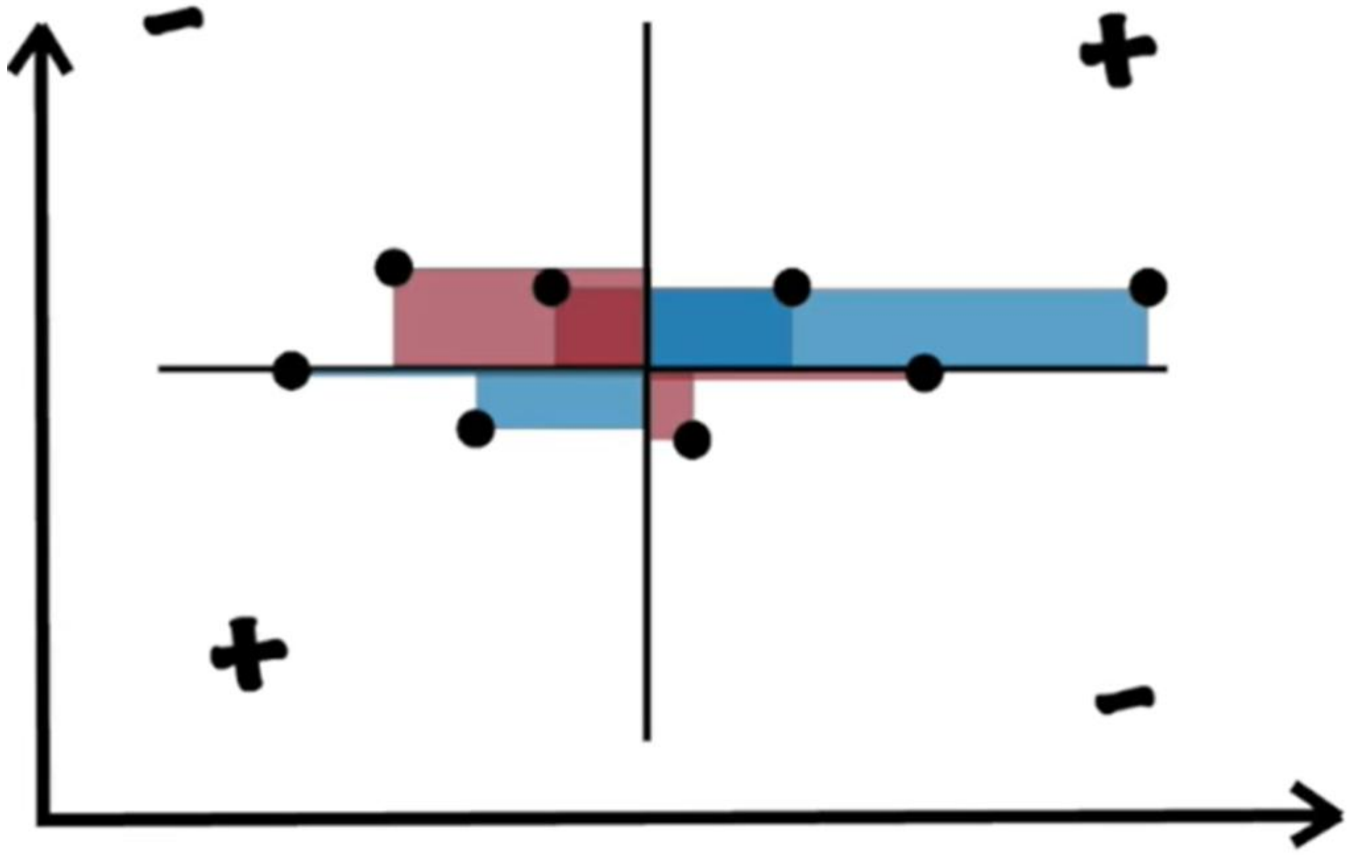
$$\sum \quad \times \frac{1}{n}$$

Negative Covariance



Low Covariance

- Dataset with spread only in one dimension will have a low covariance



Covariance Conclusion

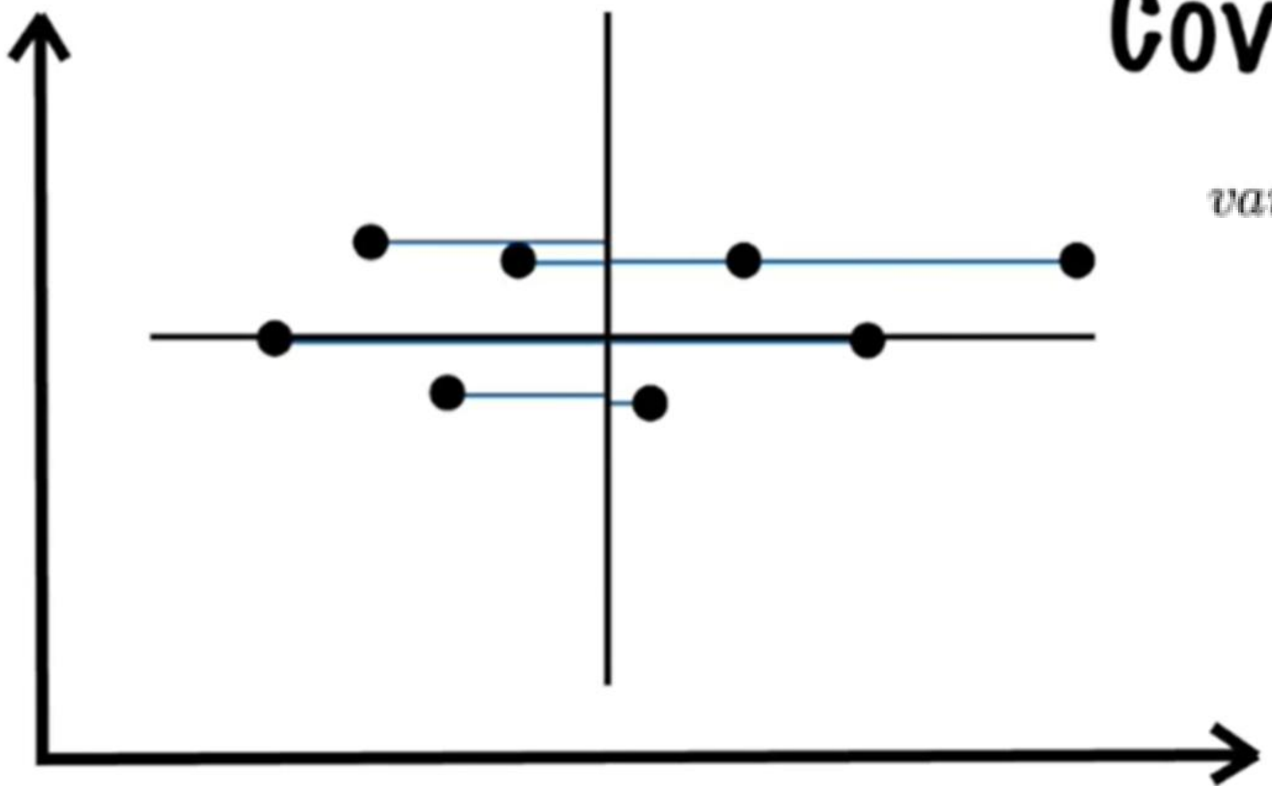
Large Negative Covariance

Near Zero Covariance

Large Positive Covariance

Variance

- Covariance of a dimension with itself.



Cov(x,x)

$$var(x) = \frac{\sum_i^n (x_i - \mu)^2}{N}$$

Covariance Matrix

- Any covariance matrix is **symmetric** and **positive semi-definite** and its main diagonal contains variances.
  - covariance is a symmetric function, i.e. Cov(X,Y)=Cov(Y,X)

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x,x) & \mathrm{Cov}(x,y) & \mathrm{Cov}(x,z) \\ \mathrm{Cov}(y,x) & \mathrm{Var}(y,y) & \mathrm{Cov}(y,z) \\ \mathrm{Cov}(z,x) & \mathrm{Cov}(z,y) & \mathrm{Var}(z,z) \end{bmatrix}$$

Covariance Matrix for Graph

- $C = \frac{1}{n}(X_{\text{centered}})^\top (X_{\text{centered}})$

1. **Data Matrix X**: Let $X = A$.

$$X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

2. **Optionally, mean-center** each column. (Calculate the mean of each column and subtract it from each entry in that column.)

3. **Covariance Matrix**:

$$C = \frac{1}{4} X_{\text{centered}}^\top X_{\text{centered}}.$$

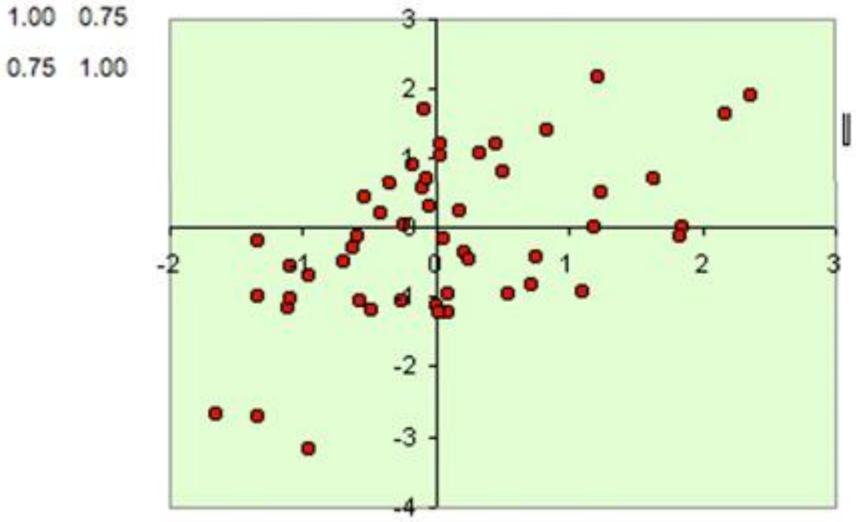
03

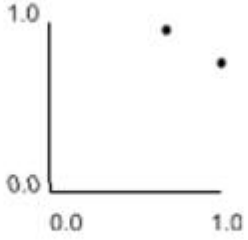
Physical interpretation

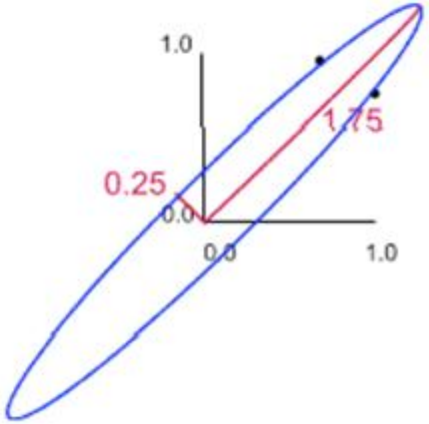
Physical interpretation

- Consider a covariance matrix, **A**, i.e., A = 1/n $S^T S$ for some S

$$\mathbf{A} = \begin{bmatrix} 1 & .75 \\ .75 & 1 \end{bmatrix} \Rightarrow \lambda_1 = 1.75, \lambda_2 = 0.25$$







- Error ellipse with the major axis as the larger eigenvalue and the minor axis as the smaller eigenvalue

Eigenvalues and Eigenvectors

- The value λ is an eigenvalue of matrix A if there exists a non-zero vector x, such that $Ax=\lambda x$. Vector x is an eigenvector of matrix A
  - The largest eigenvalue is called the principal eigenvalue
  - The corresponding eigenvector is the principal eigenvector
  - Corresponds to the direction of maximum change

04

Principal Components

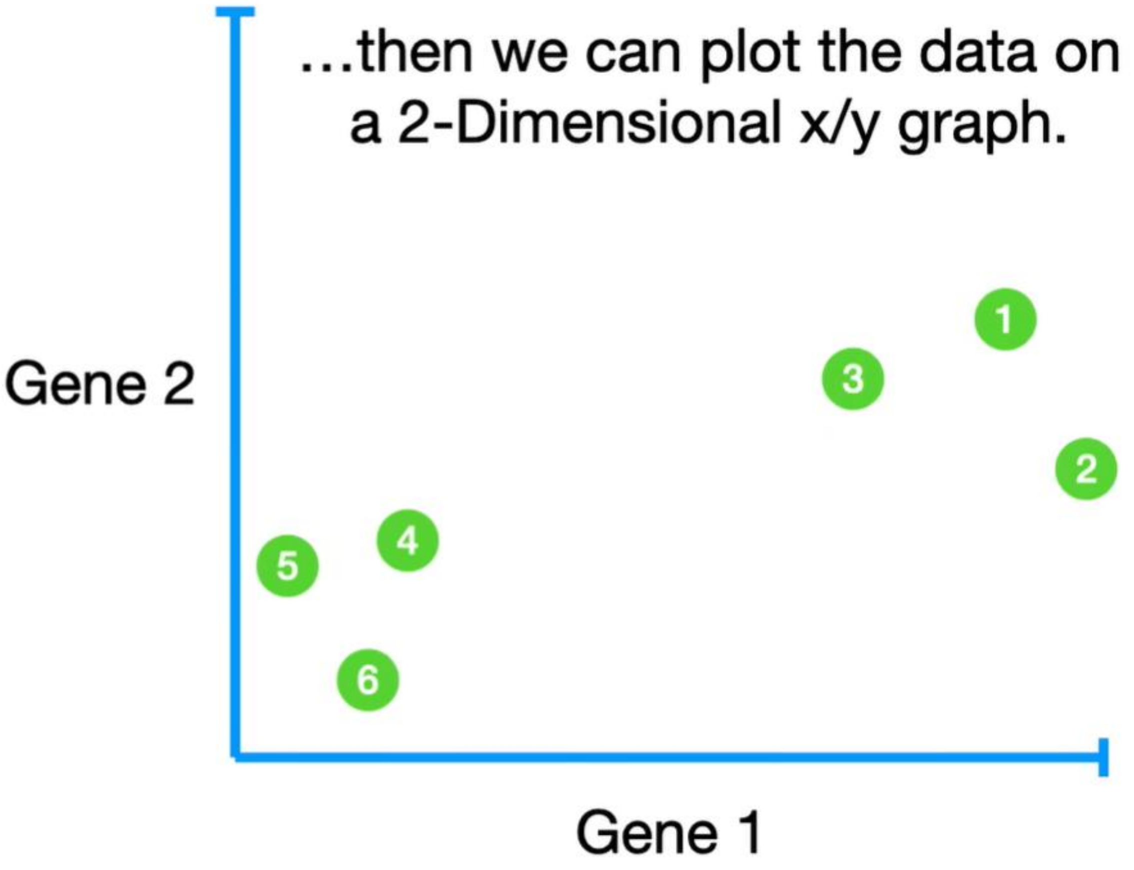
Introduction

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	1	2

Even though it's a simple graph, it shows us that mice 1, 2 and 3 are more similar to each other than they are to mice 4, 5 6.

Introduction

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	1	2
Gene 2	6	4	5	3	2.8	1



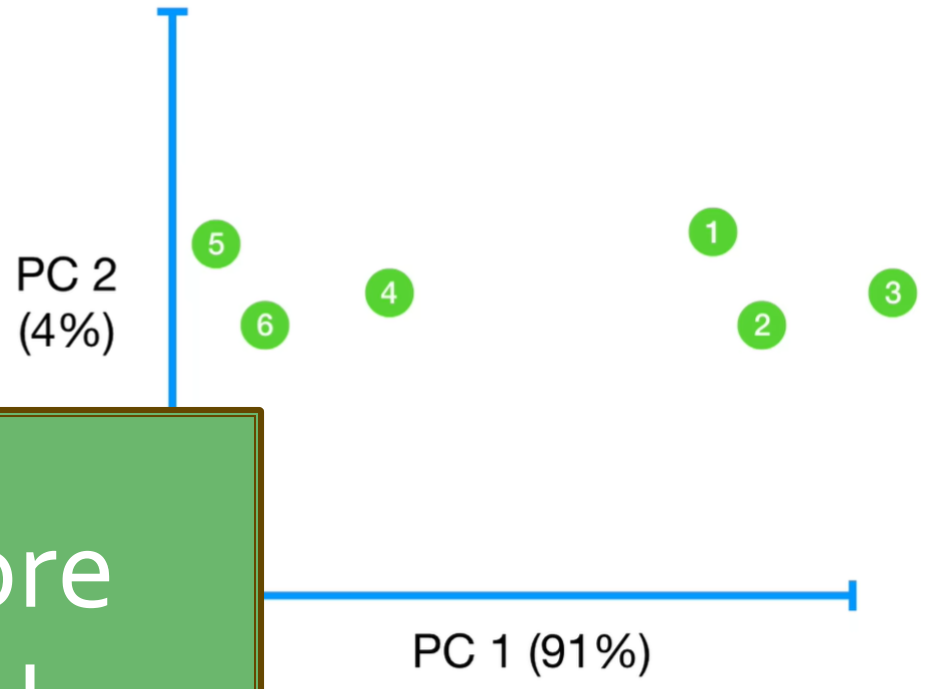
Introduction

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1
Gene 3	12	9	10	2.5	1.3	2
Gene 4	5	7	6	2	4	7

If we measured 4 genes, however, we can no longer plot the data - 4 genes require 4 dimensions.

Introduction

PCA might tell us that Gene 3 is responsible for separating samples along the x-axis.

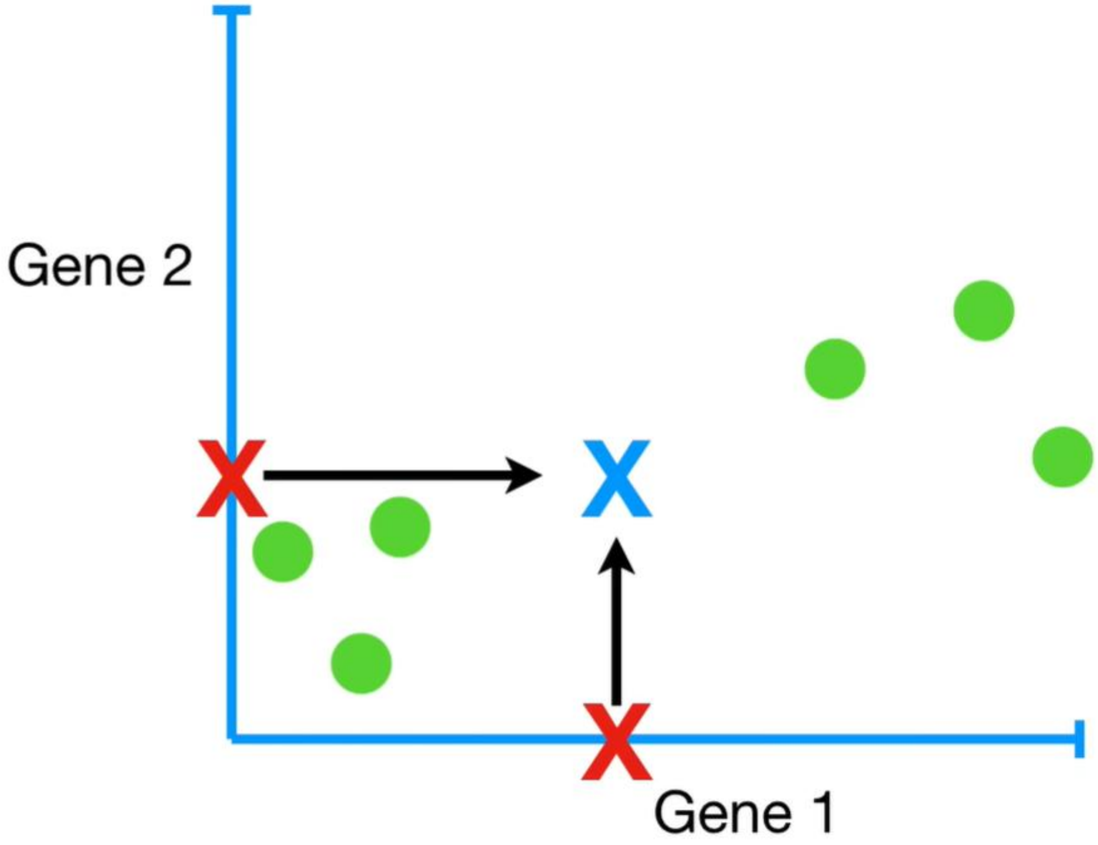


How PCA can take 4 or more gene measurements and make a 2-D PCA Plot?

What PCA Does

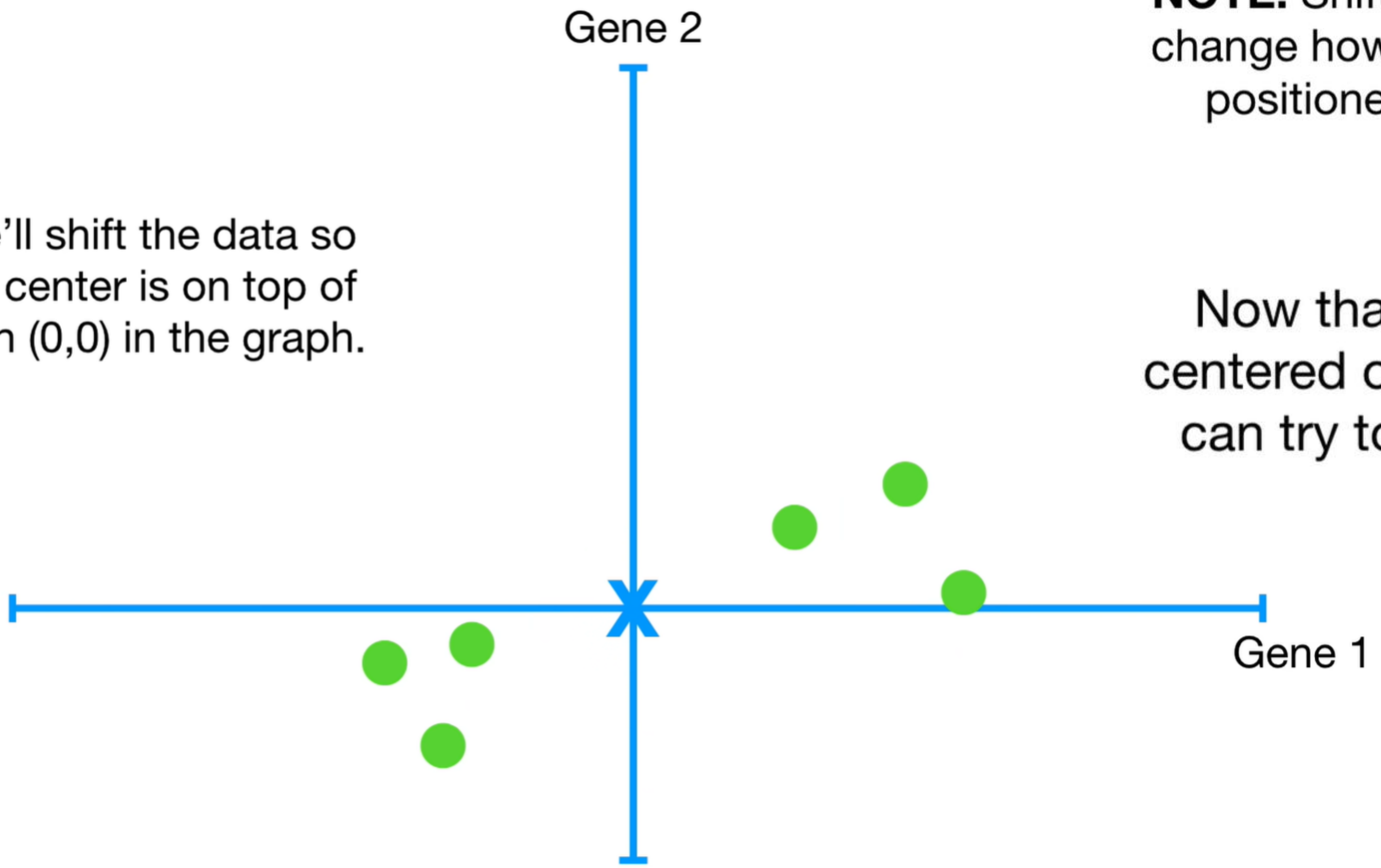
	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1

From this point on, we'll focus on what happens in the graph; we no longer need the original data…



What PCA Does

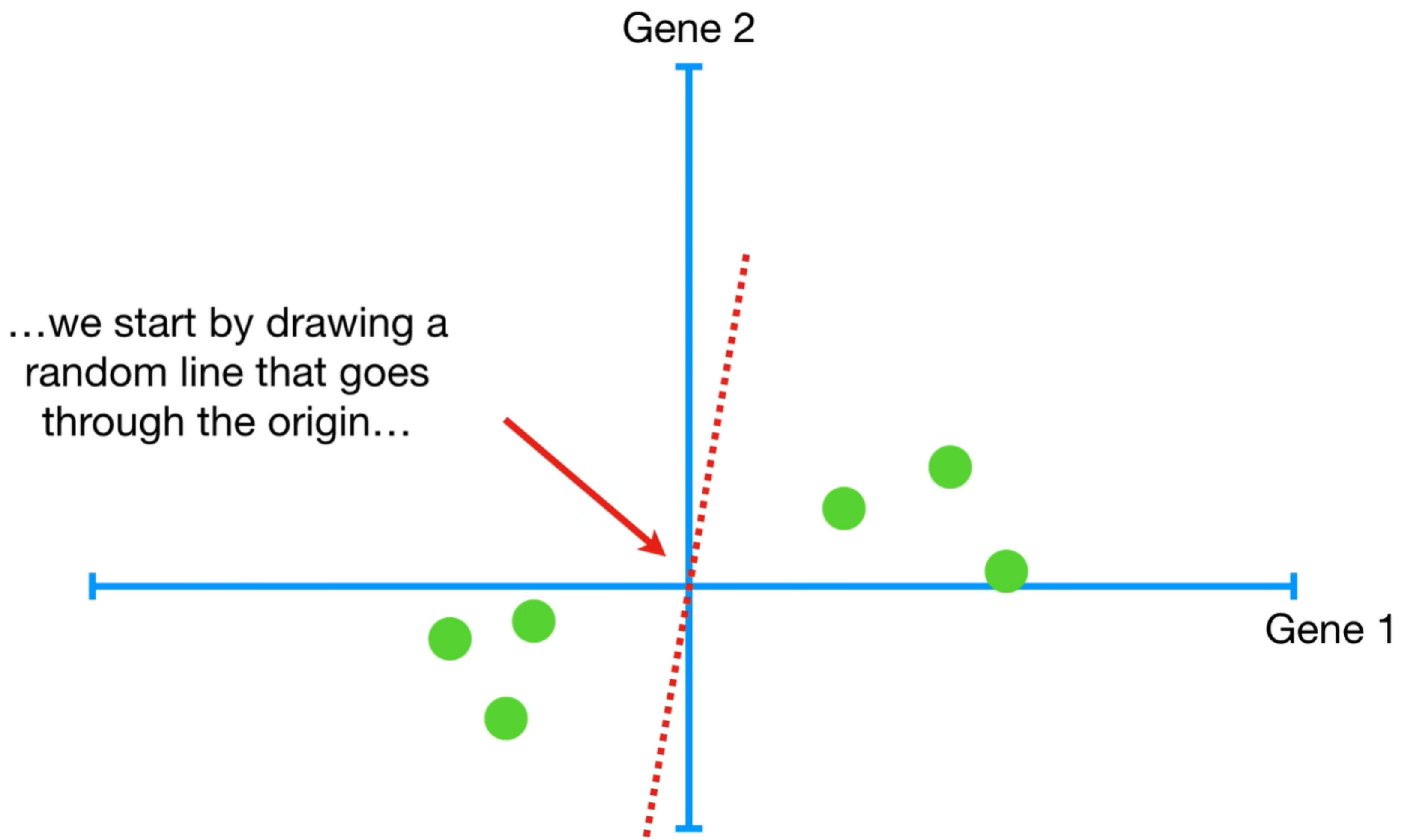
Now we'll shift the data so that the center is on top of the origin (0,0) in the graph.



NOTE: Shifting the data did not change how the data points are positioned *relative to each other*.

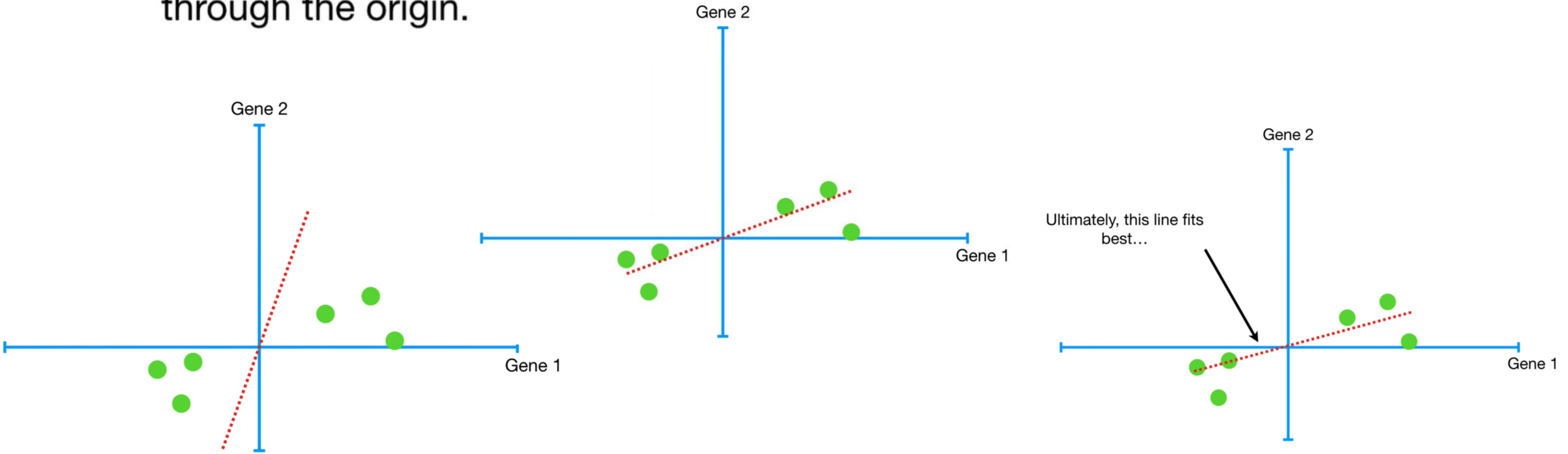
Now that the data are centered on the origin, we can try to fit a line to it.

What PCA Does

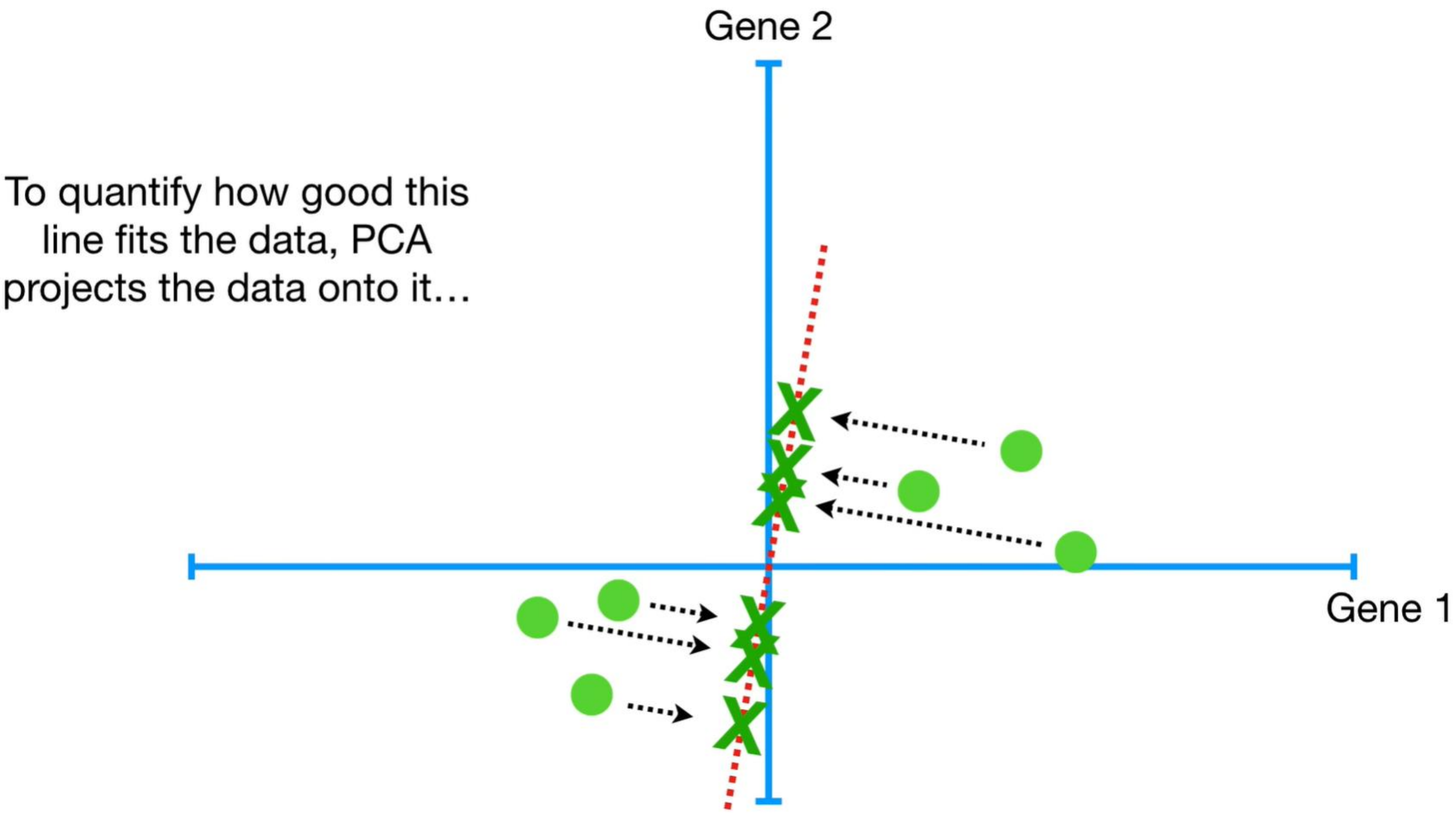


What PCA Does

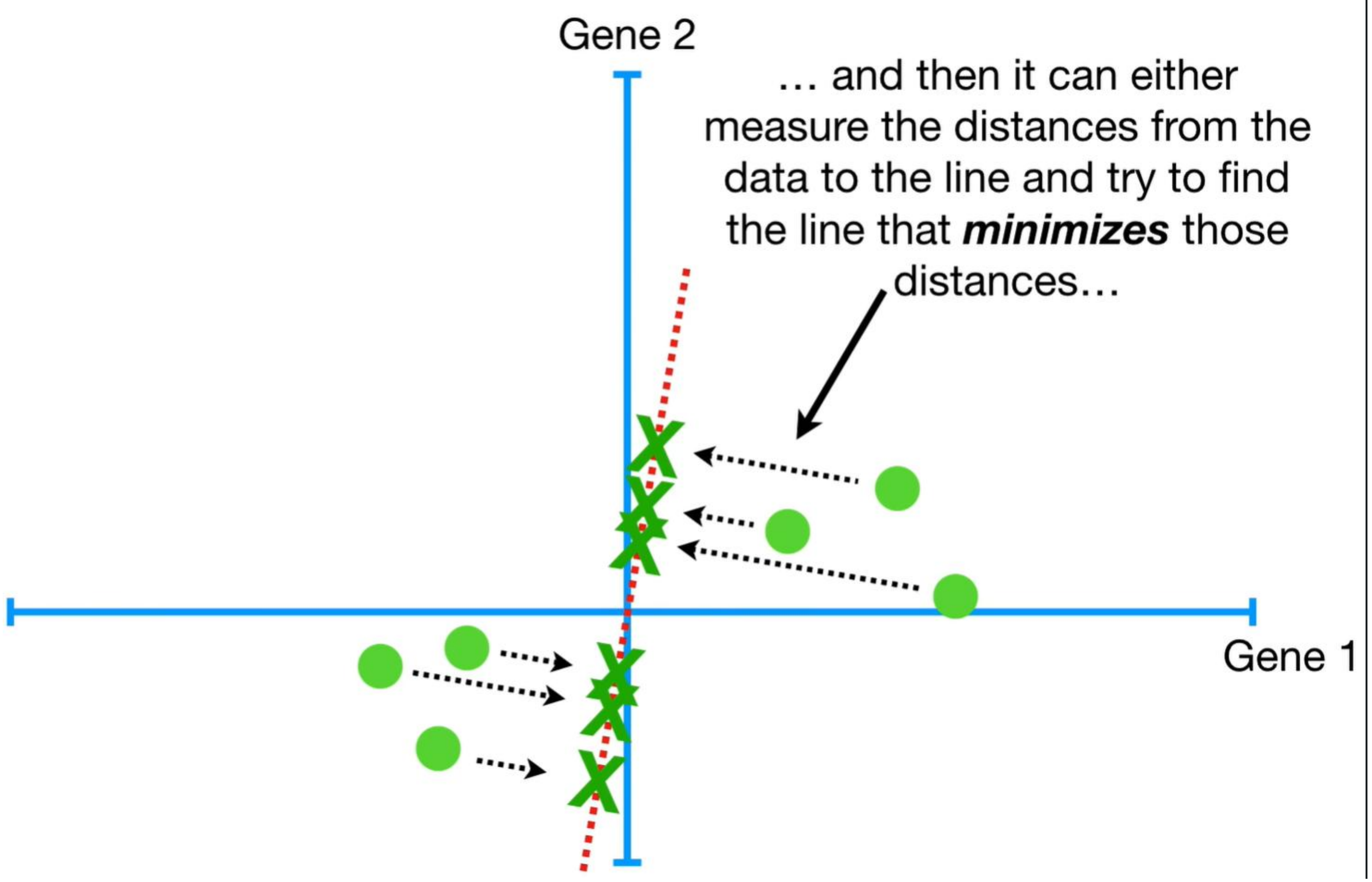
…then we rotate the line until it fits the data as well is it can, given that it has to go through the origin.



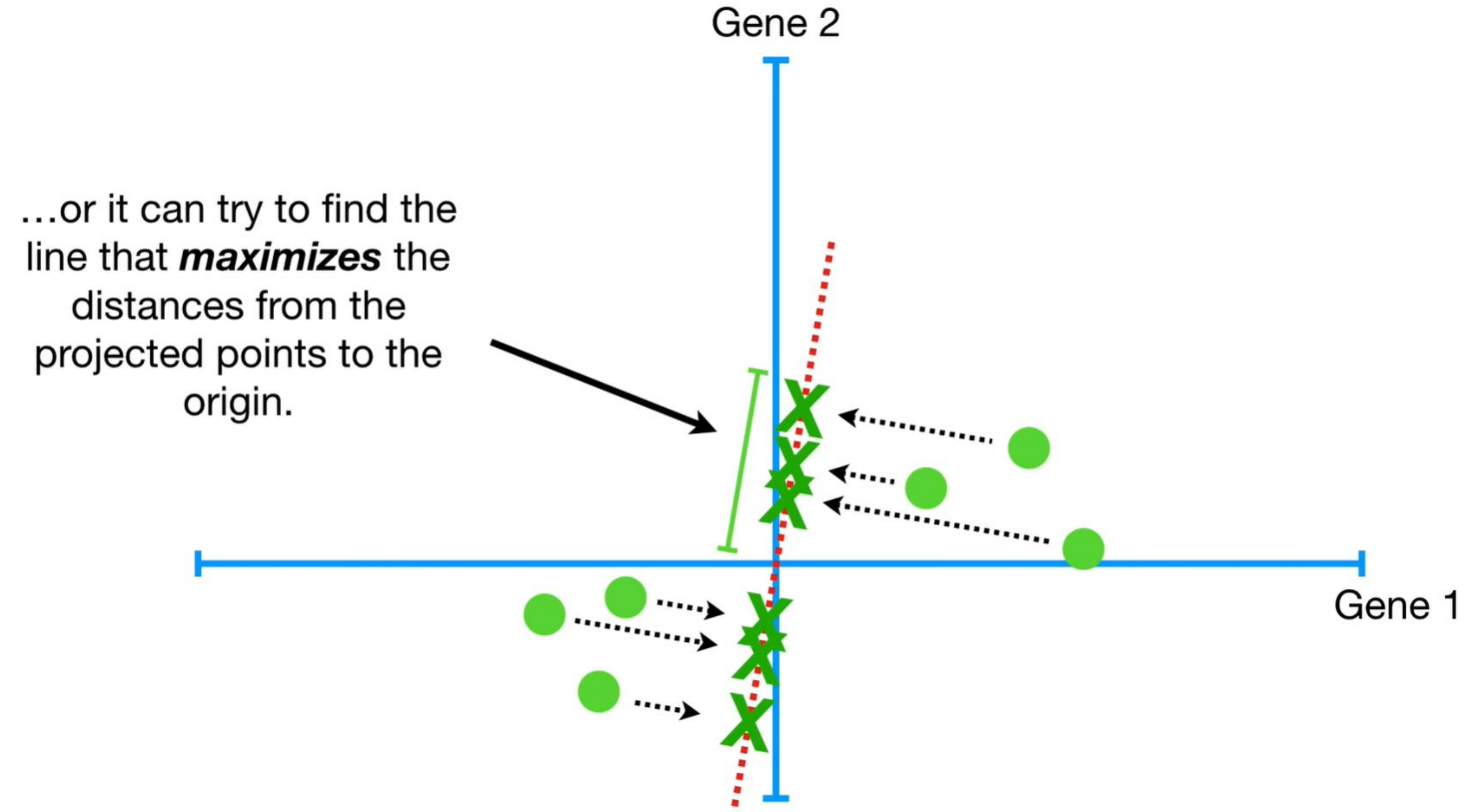
How PCA Decides the Best Line?



How PCA Decides the Best Line?

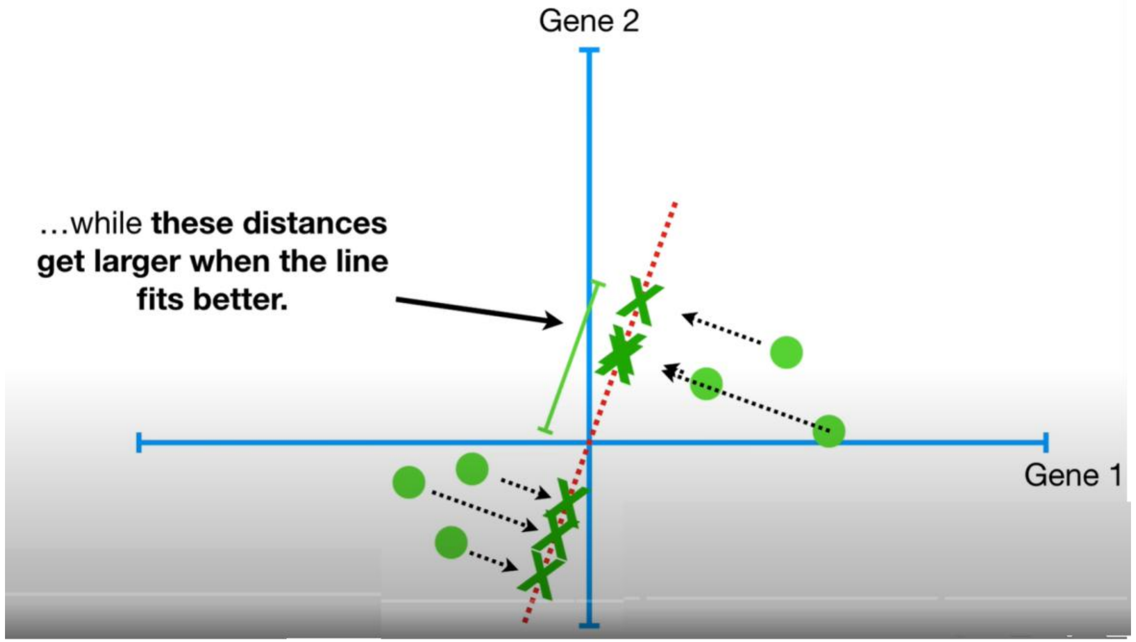


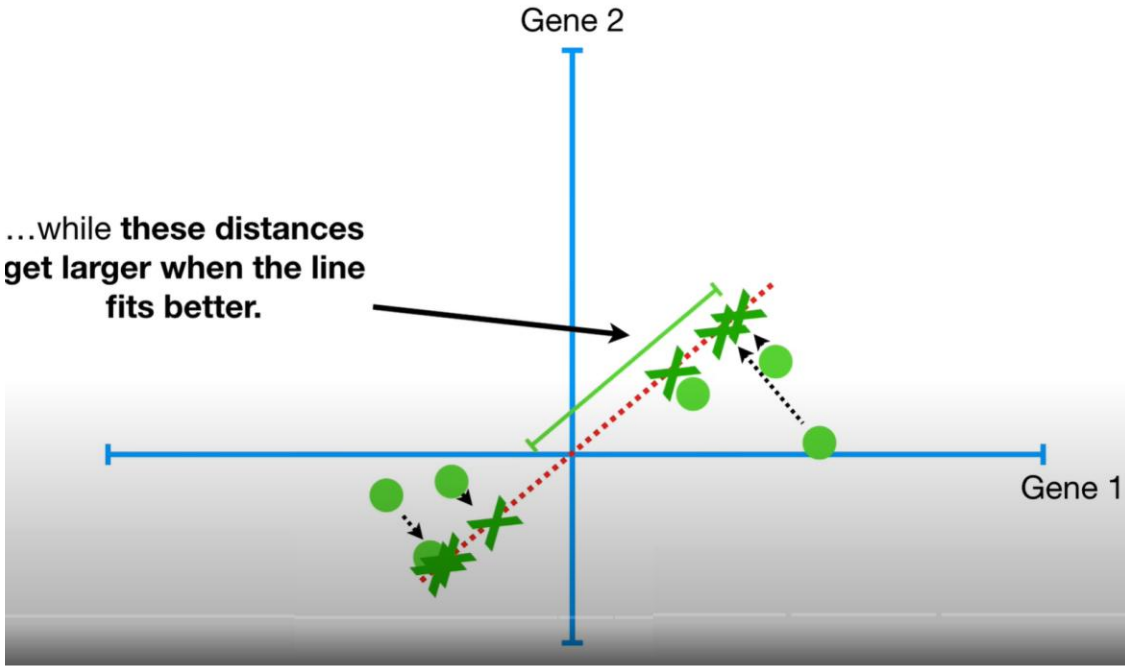
How PCA Decides the Best Line?



Let's Think

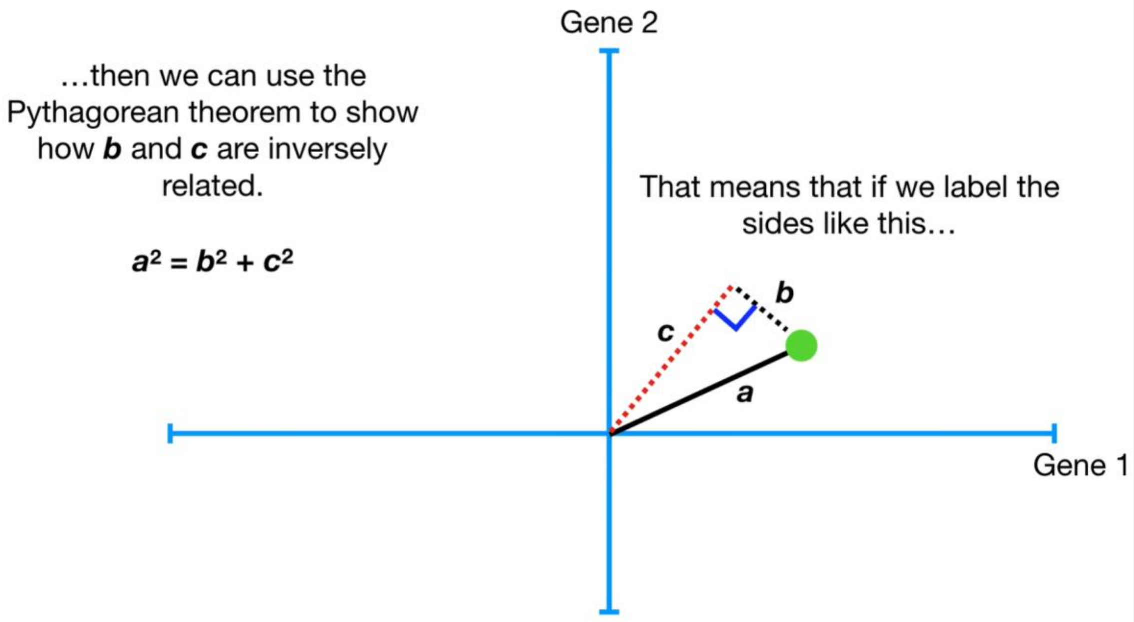
- min(distances of points from line) = max(distances of projected points to the origin)

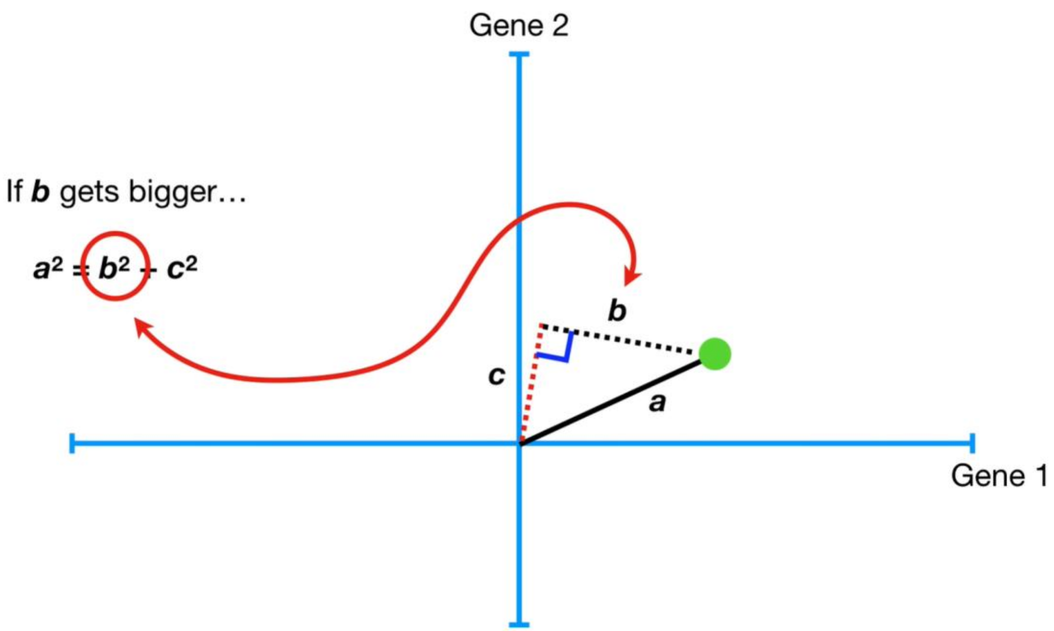




Let’s Think

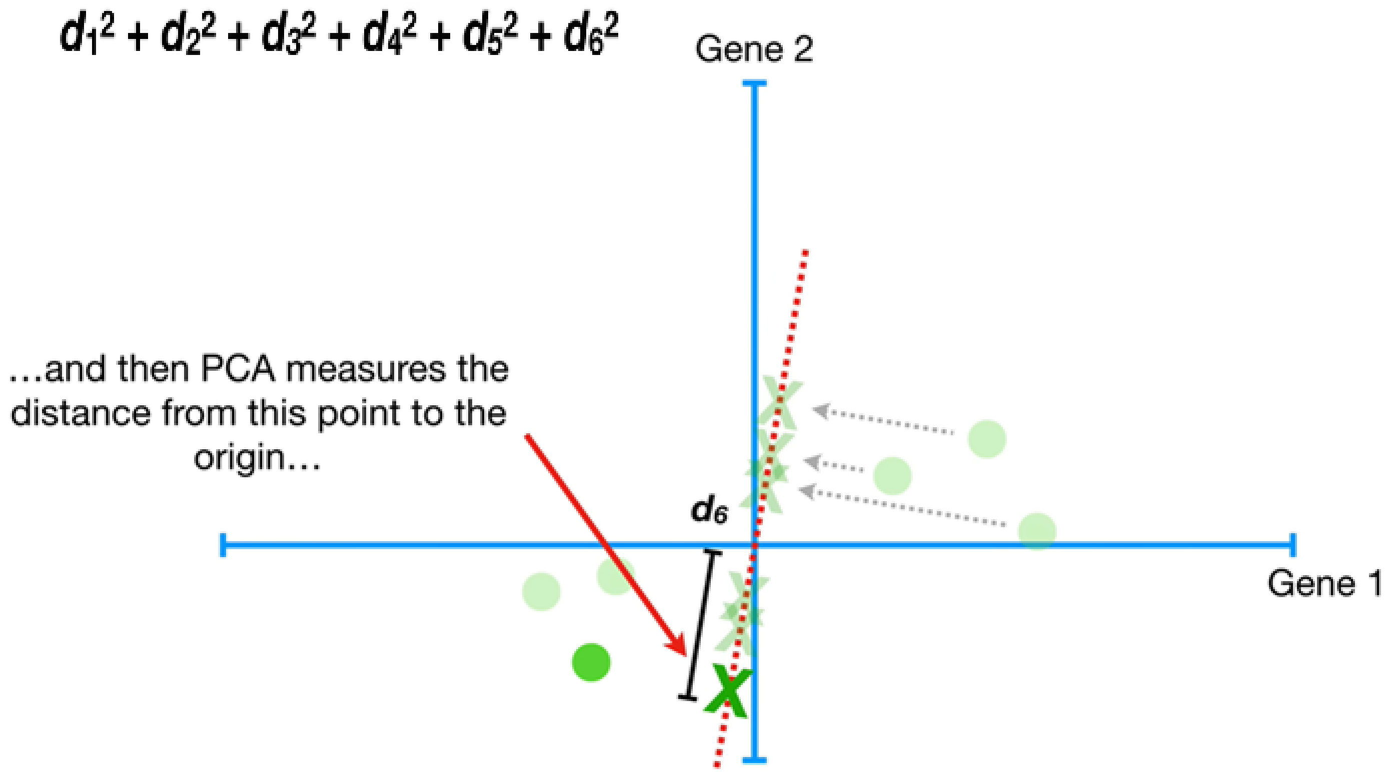
- Consider one data point.
- The distance from the point to the origin doesn’t change when the red dotted line rotates.
- Project the point onto the line
- It is usually easier to calculate “c”.





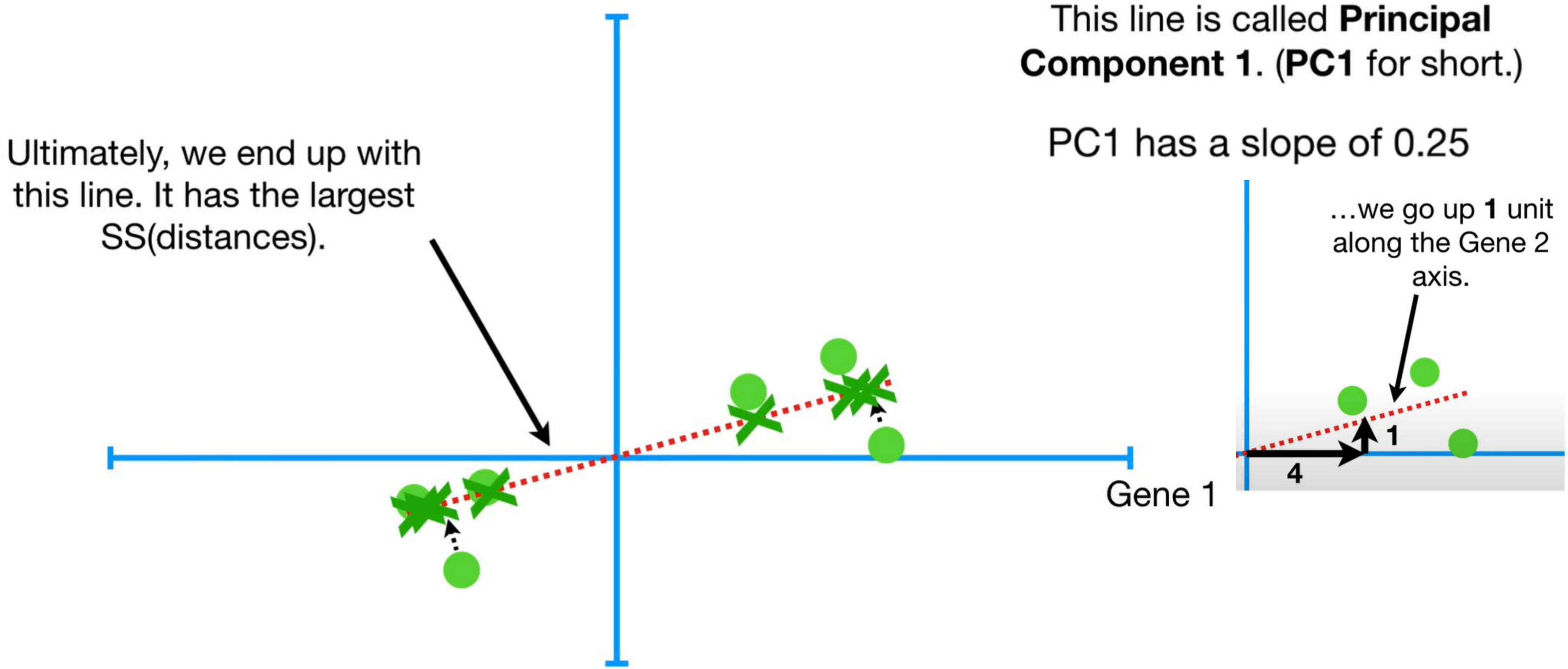
PCA

- PCA finds the best fitting line by maximizing the sum of the squared distances from the projected points to the origin.

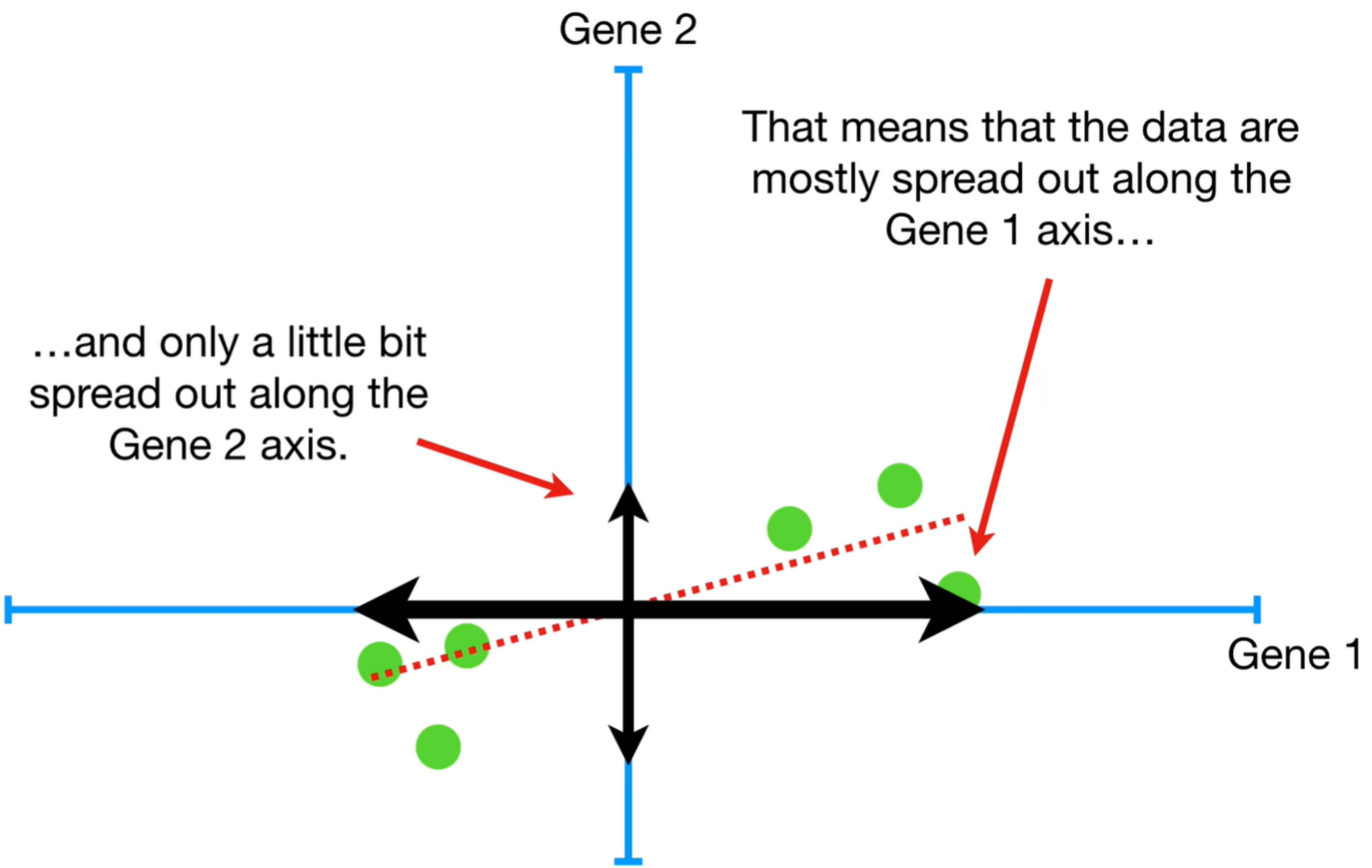


PC1

$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$ = sum of squared distances = SS(distances)



PC1

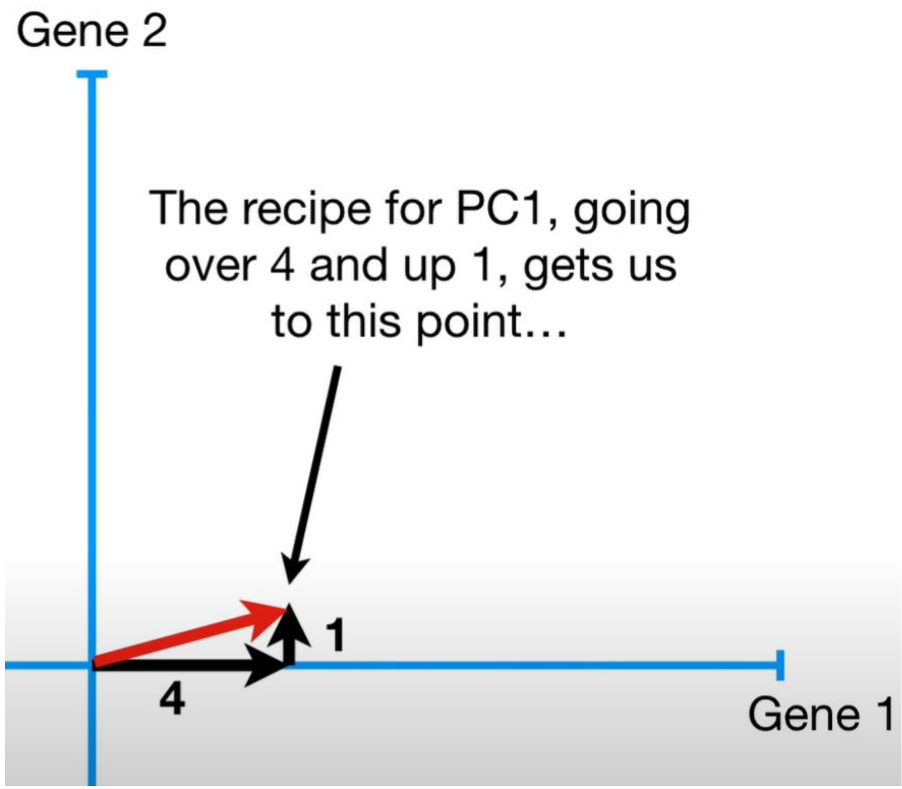


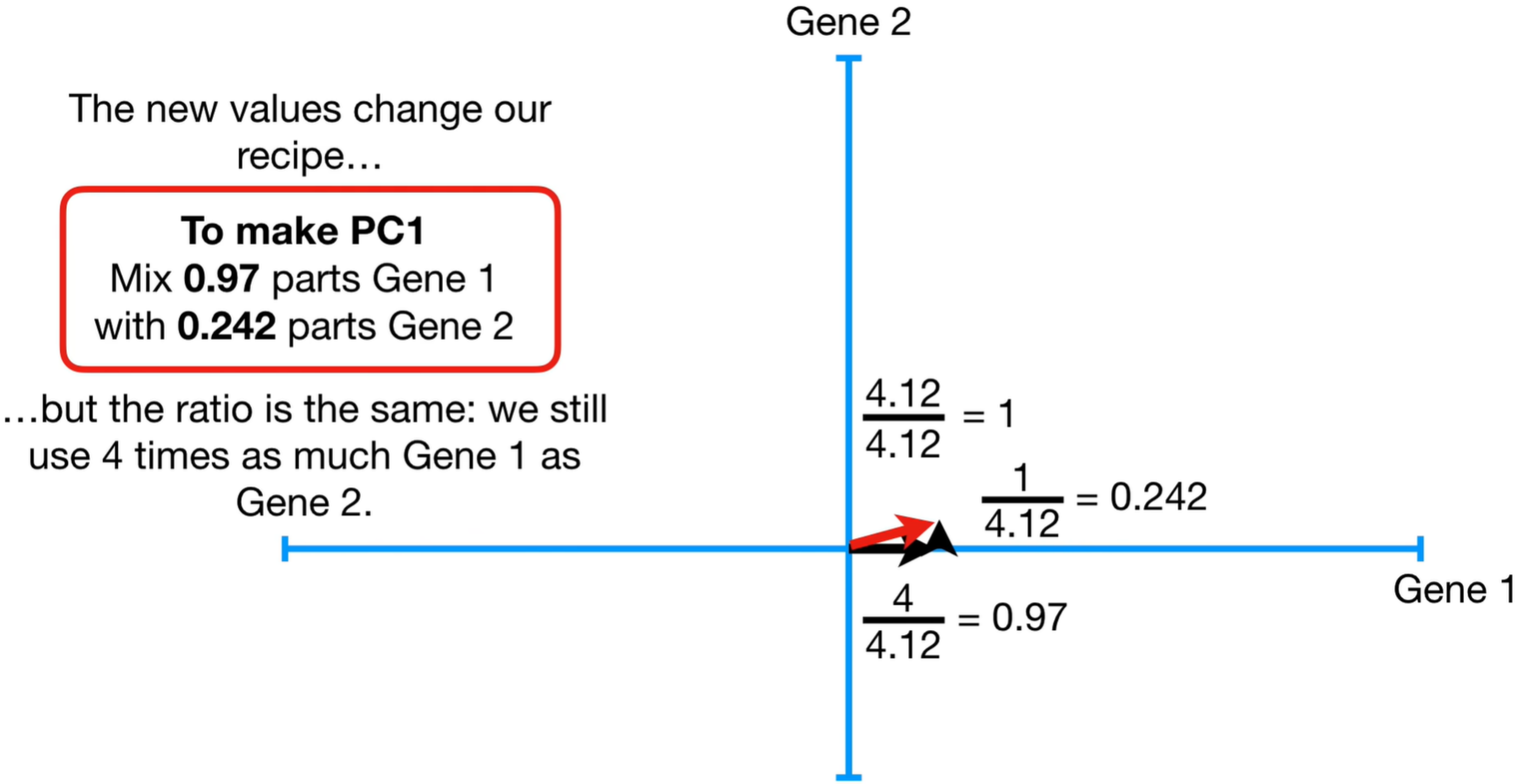
To make PC1
Mix **4** parts Gene 1 with **1** part Gene 2

The ratio of Gene 1 to Gene 2 tells you that Gene 1 is more important when it comes to describing how the data are spread out..

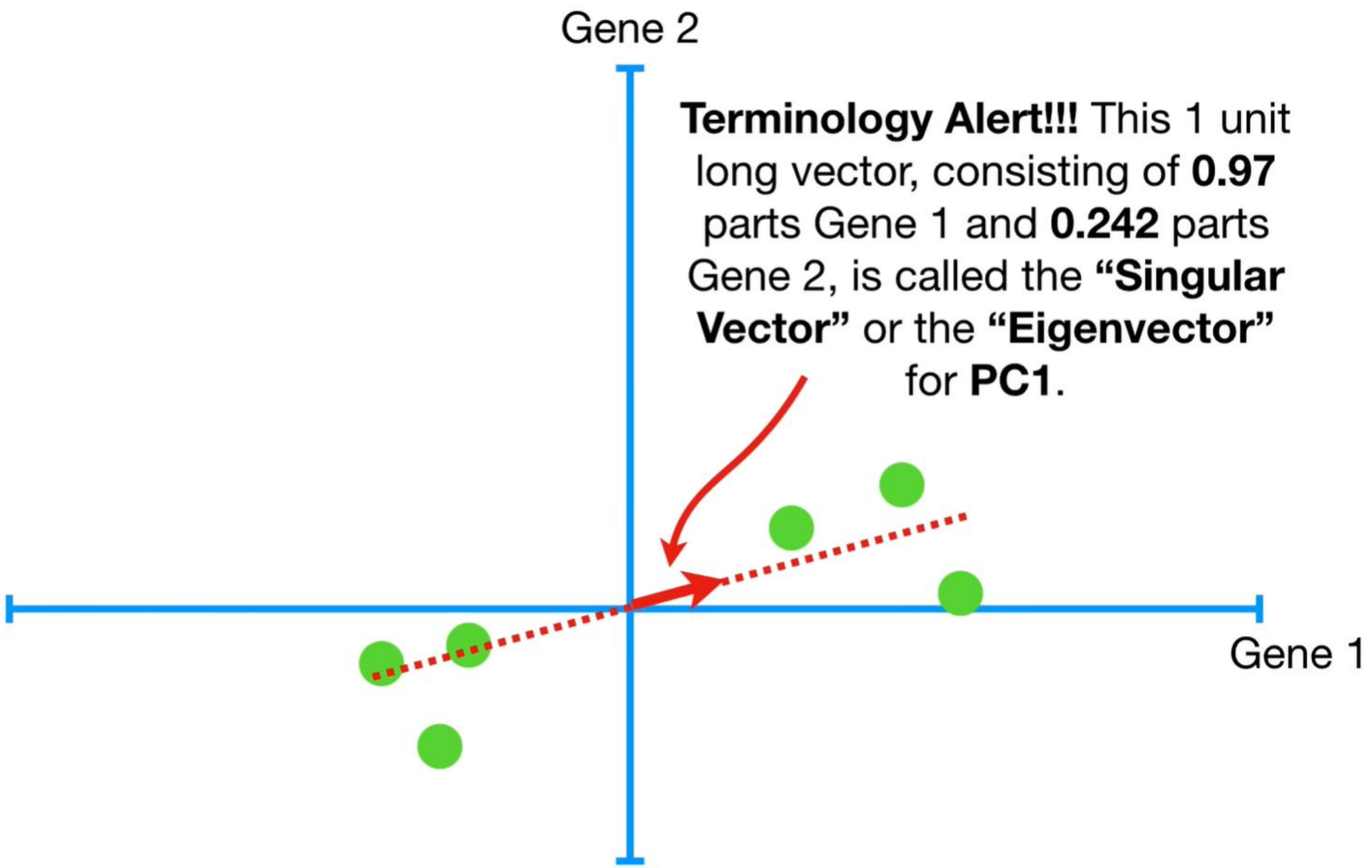
a "*linear combination*" of Genes 1 and 2.

PC1


Gene 2
The recipe for PC1, going over 4 and up 1, gets us to this point…
1
4
Gene 1


The new values change our recipe…
To make PC1
Mix **0.97** parts Gene 1
with **0.242** parts Gene 2
…but the ratio is the same: we still use 4 times as much Gene 1 as Gene 2.
Gene 2
4.12 / 4.12 = 1
1 / 4.12 = 0.242
Gene 1
4 / 4.12 = 0.97

PC1


Gene 2
Terminology Alert!!! This 1 unit long vector, consisting of **0.97** parts Gene 1 and **0.242** parts Gene 2, is called the **“Singular Vector”** or the **“Eigenvector”** for **PC1**.
Gene 1

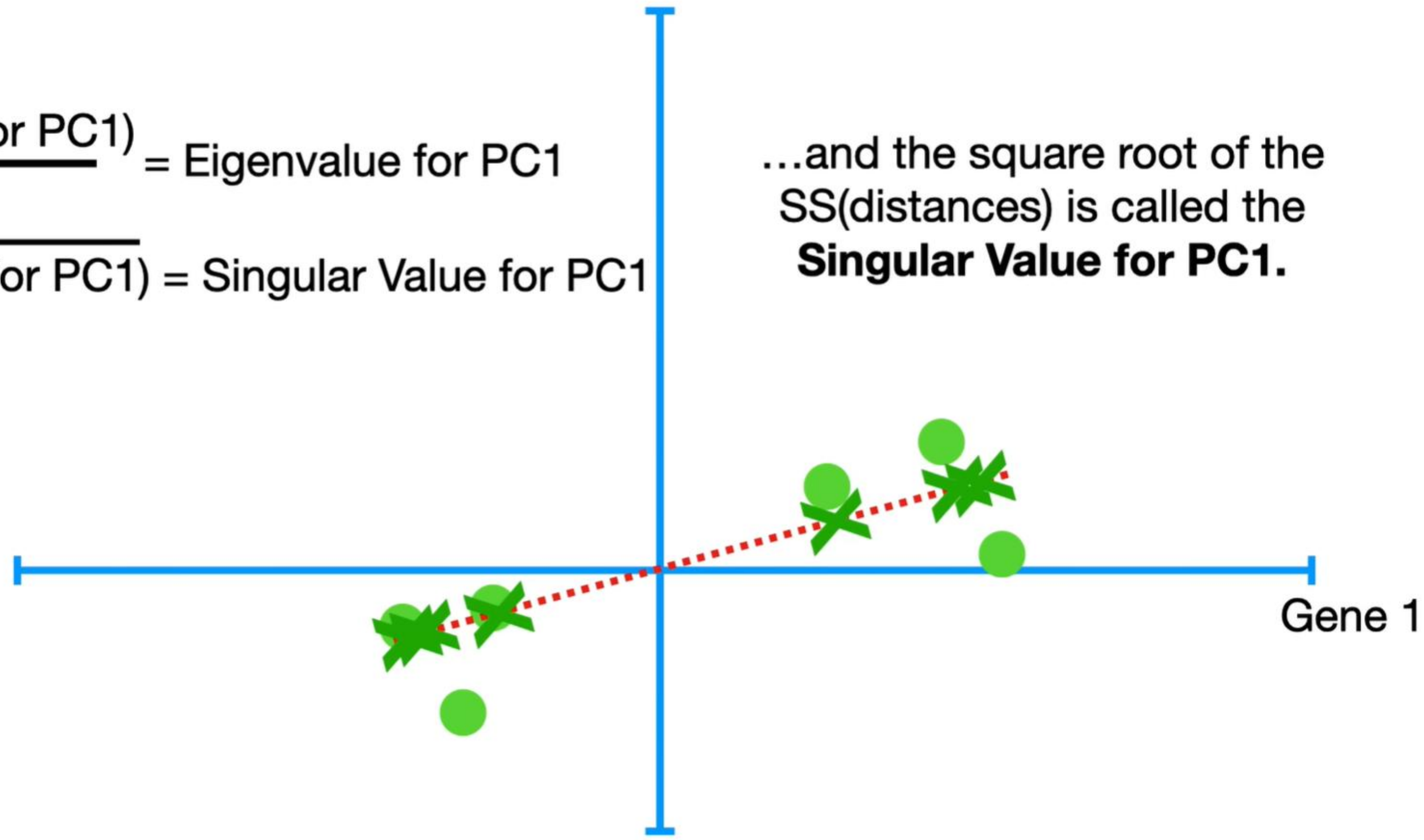
PC1

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS(distances)}$$

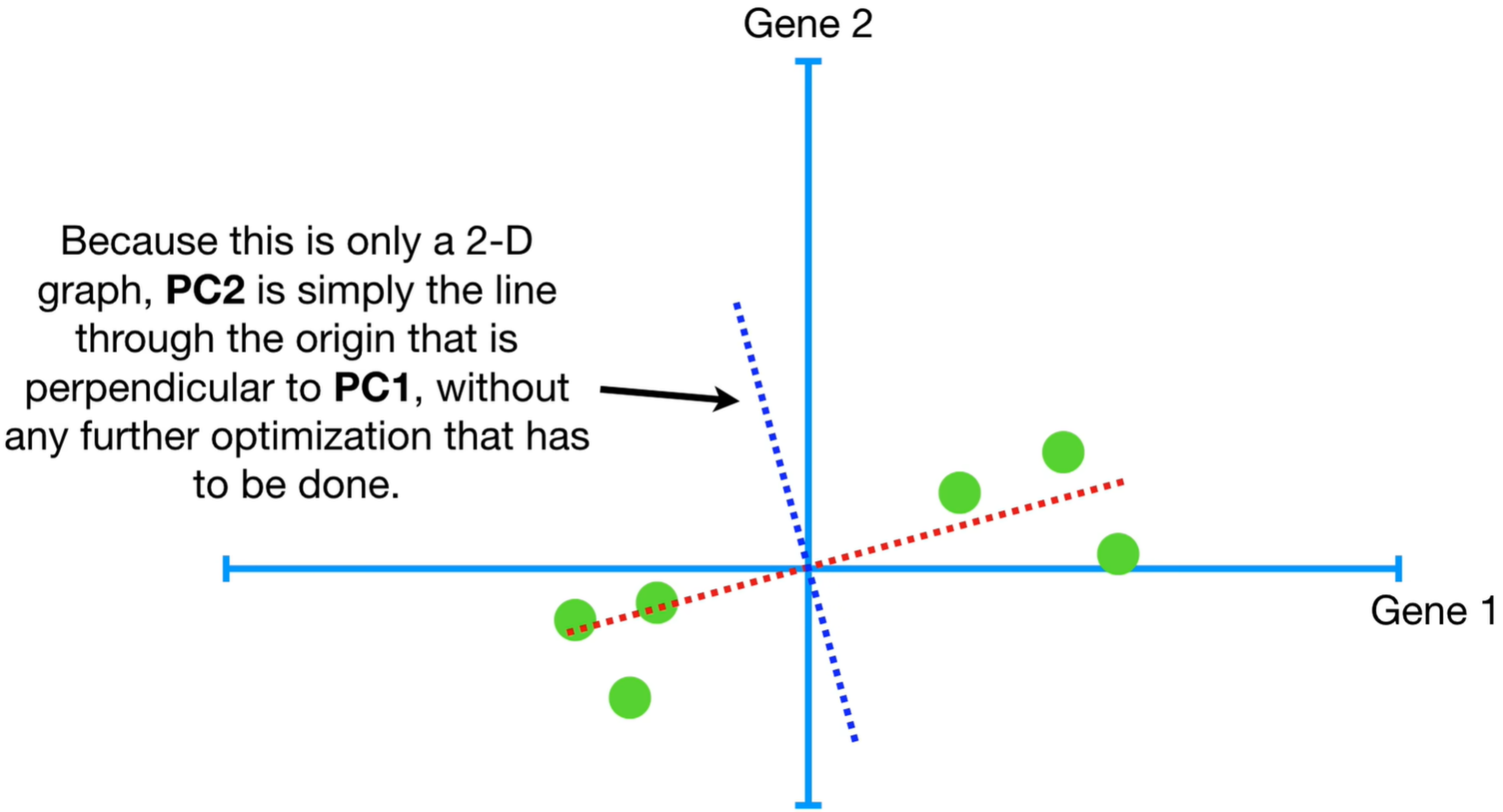
$$\frac{\text{SS(distances for PC1)}}{n} = \text{Eigenvalue for PC1}$$

$$\sqrt{\text{SS(distances for PC1)}} = \text{Singular Value for PC1}$$

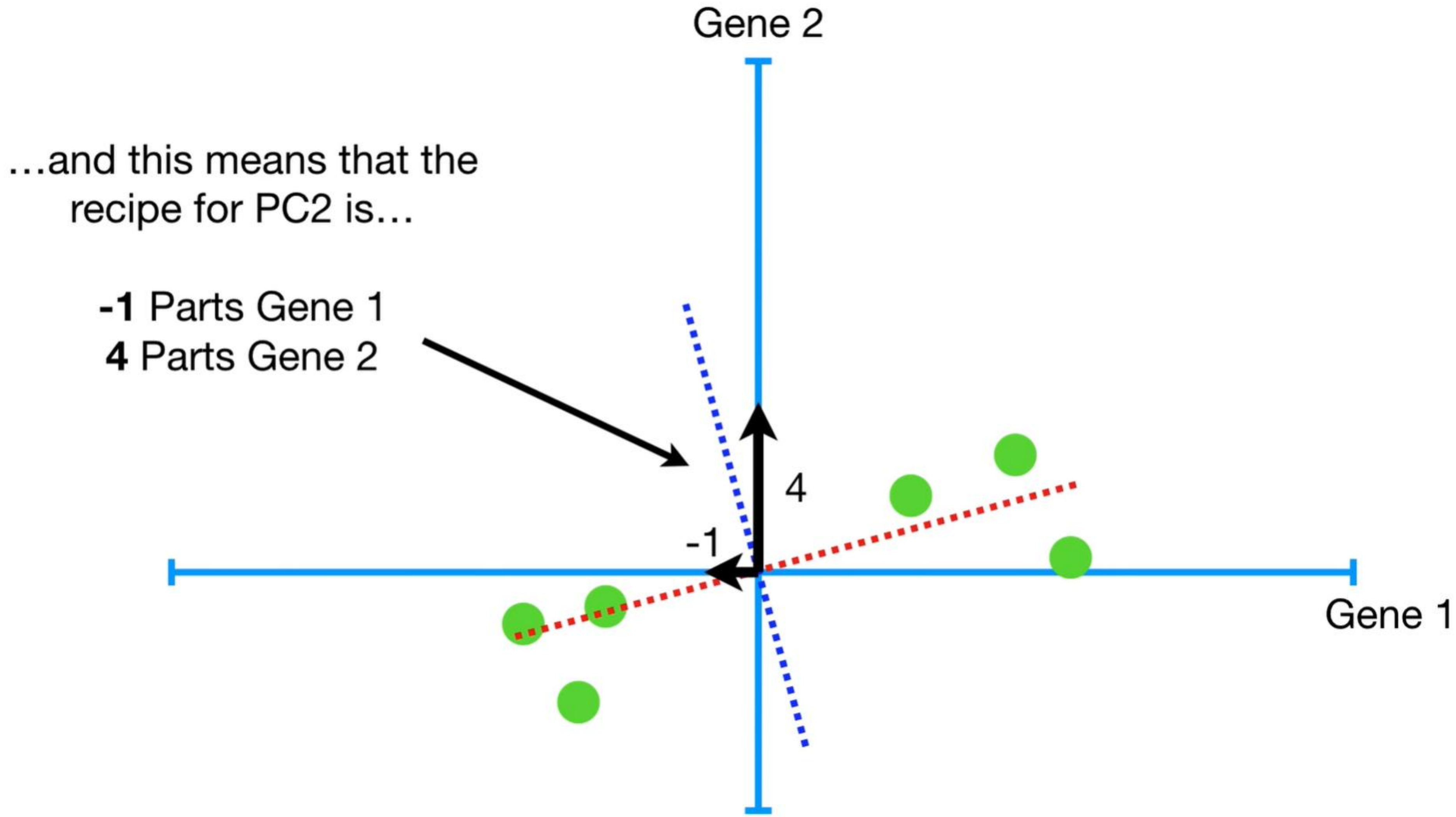
…and the square root of the SS(distances) is called the **Singular Value for PC1.**



PC2


Gene 2
Because this is only a 2-D graph, **PC2** is simply the line through the origin that is perpendicular to **PC1**, without any further optimization that has to be done.
Gene 1

PC2



Gene 2
...and this means that the recipe for PC2 is...
-1 Parts Gene 1
4 Parts Gene 2
4
-1
Gene 1

Conclusion

- Why? $$\underbrace{\text{Eigen-decomposition}(C)}_{C=X_{\text{centered}}^\top X_{\text{centered}}} \longleftrightarrow \underbrace{\text{SVD}(X_{\text{centered}})}_{X_{\text{centered}}=U\,\Sigma\,V^\top}.$$
- PCA is the eigen decomposition of covariance matrix.
- PCA is the SVD decomposition of matrix.

- PCA on $C = X^\top X$:
  - Eigenvalues of $C \rightarrow$ variances in principal directions.
  - Eigenvectors of $C \rightarrow$ principal axes (PCs).
- SVD on X:
  - Right singular vectors of $X \rightarrow$ same directions as eigenvectors of C.
  - Singular values are $\sqrt{\text{eigenvalues of } C}$.

Example

- $X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

- Eigen vector of covariance matrix

Covariance Matrix $C = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$

PCA on C means we look for eigen-decomposition:

$$C\,\mathbf{v} = \lambda\,\mathbf{v}.$$

- The eigenvalues λ indicate the variance captured.
- The eigenvectors $\mathbf{v}$ are the principal components.

Approximate Largest Eigenvalue & Eigenvector

- Largest eigenvalue $\lambda_{\max} \approx 4.70.$
- Corresponding eigenvector $\mathbf{v}_{\max} \approx (0.529,\ 0.597,\ 0.529,\ 0.291).$

This eigenvector is **PC1** (the first principal component) of C.

Example

- $X = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

- SVD of X

Performing **Singular Value Decomposition** (SVD) on the same matrix X:

$$X = U \Sigma V^{\top},$$

where

- U is 4×4,
- Σ is 4×4 diagonal (singular values),
- V is 4×4.

From our eigen-decomposition of C, the largest eigenvalue was ≈ 4.70. Its square root is $\sqrt{4.70} \approx 2.17$. In the SVD of X:

- The **largest singular value** $\sigma_{\max}$ is about 2.17.
- The **corresponding right singular vector** is $\mathbf{v}_{\max} \approx (0.529,\ 0.597,\ 0.529,\ 0.291)$.

This is exactly the same vector (up to a possible sign) as the principal component from the eigen-decomposition of C.

Any Question?