Dimensionality Reduction

Principal Component Analysis (PCA)

Mr.Gangadhar Immadi

immadi.gangadhar@gmail.com

9986789040

Why PCA?

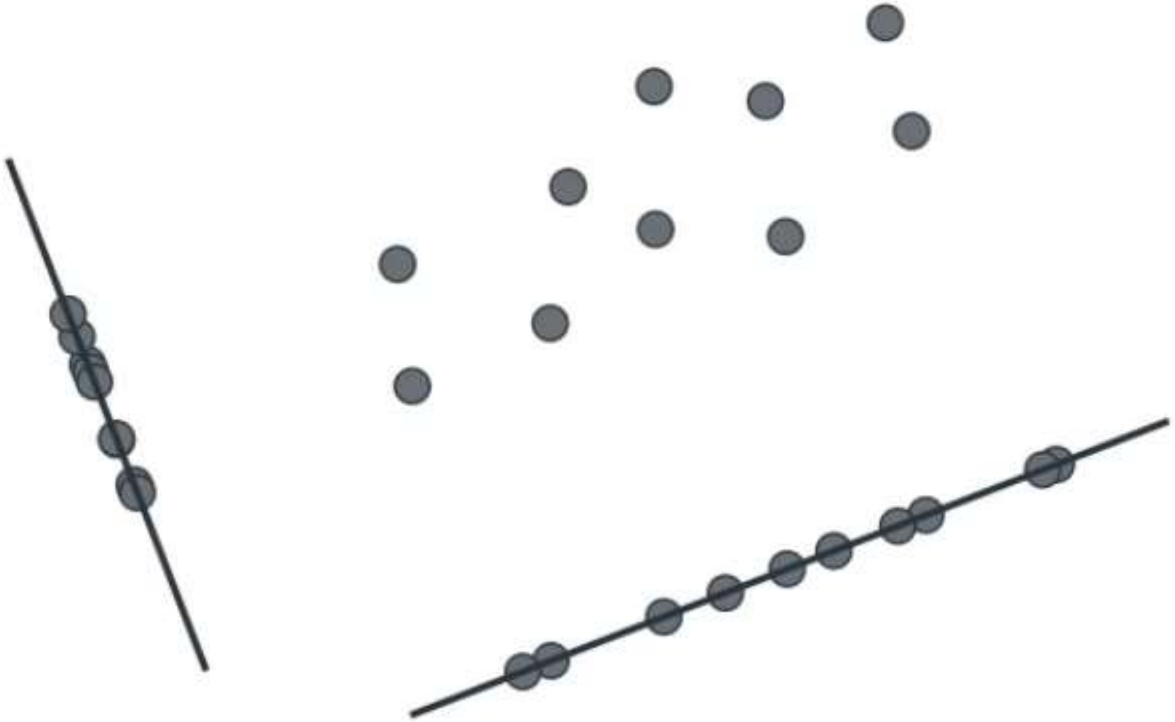
- Try to keep the maximum information with less number of features
- It’s a Dimensionality reduction technique, Not a feature selection


Nikon


Canon
EOS
R


Canon
EOS
R

Dimensionality Reduction



Housing Data

Size
Number of rooms
Number of Bathrooms
Schools around
Crime rate

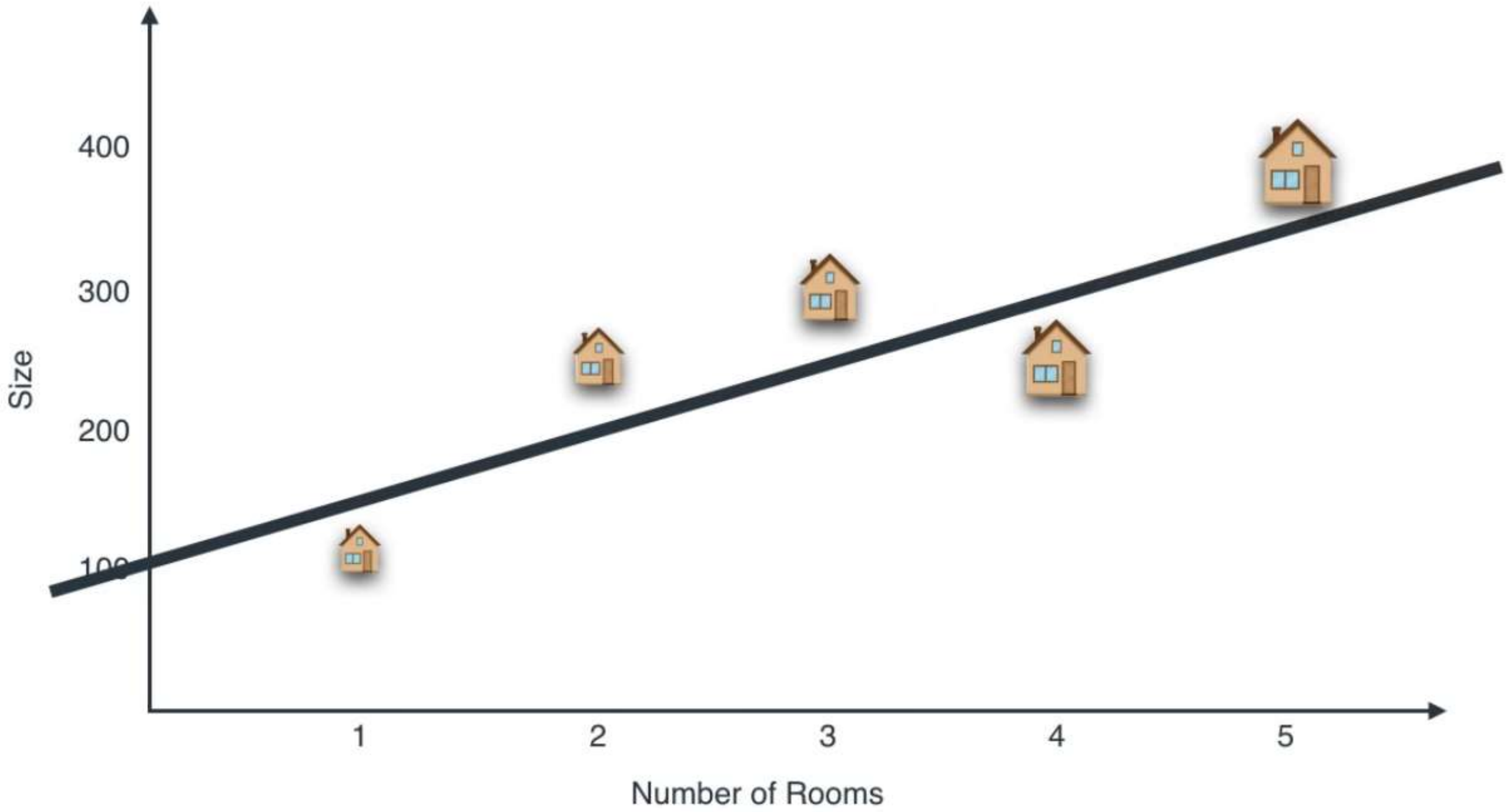
Size
Number of rooms
Number of Bathrooms

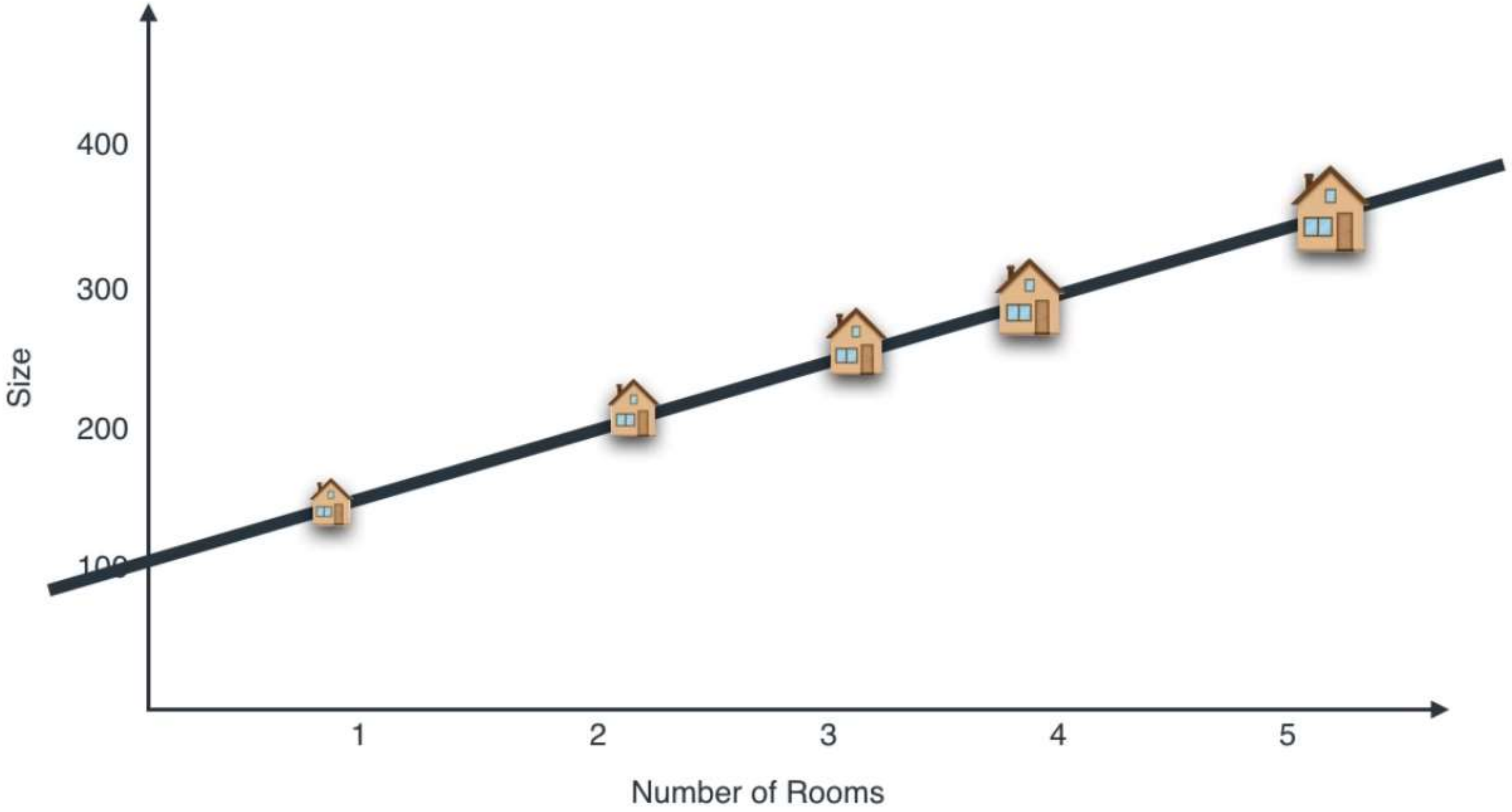
→ Size feature

Schools around
Crime rate

→ Location feature

Housing Data




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

Size feature

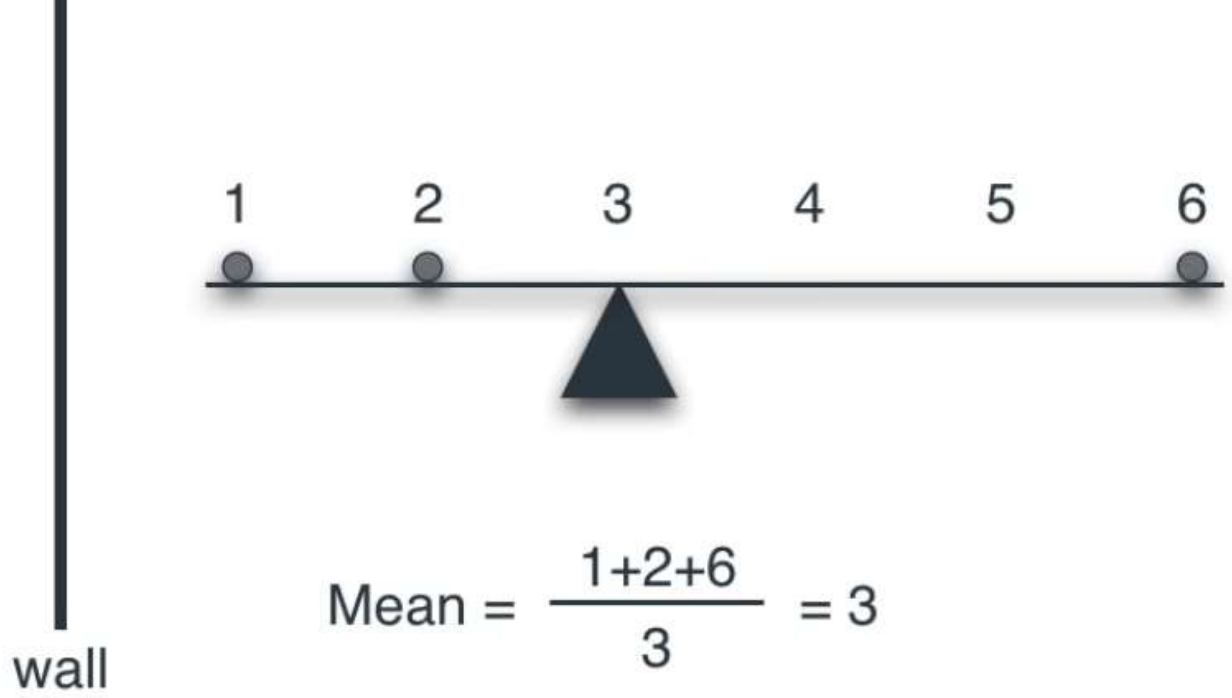
2 dimensions

size
number of rooms

1 dimension

size feature

Mean



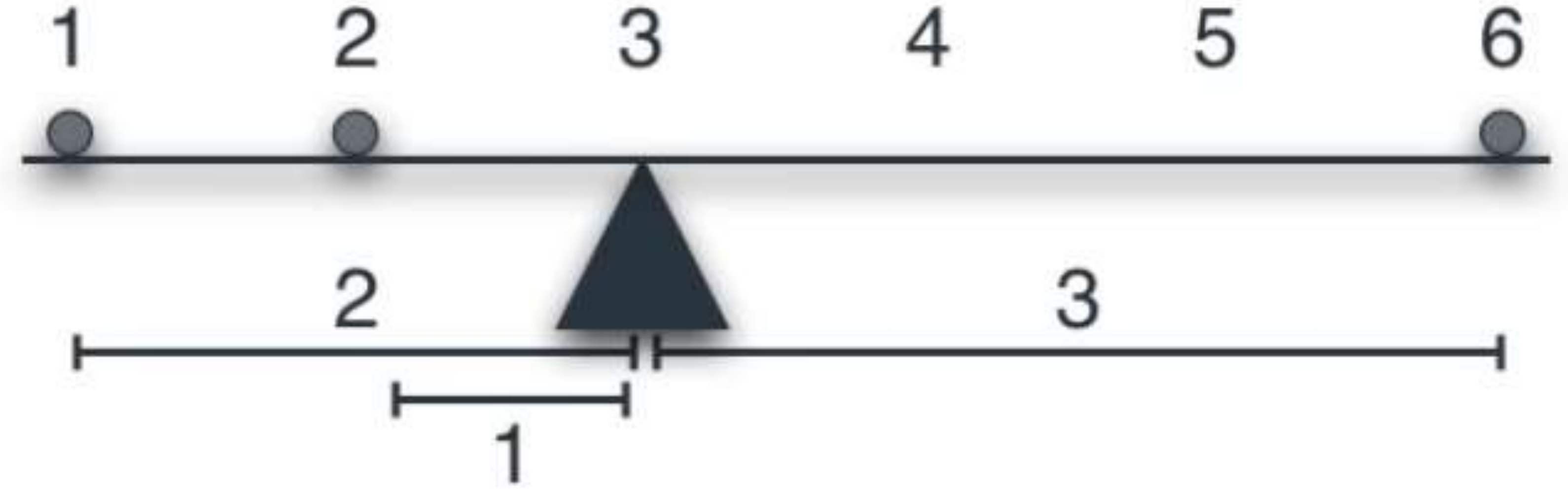
$$\text{Mean} = \frac{1+2+6}{3} = 3$$

Variance

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

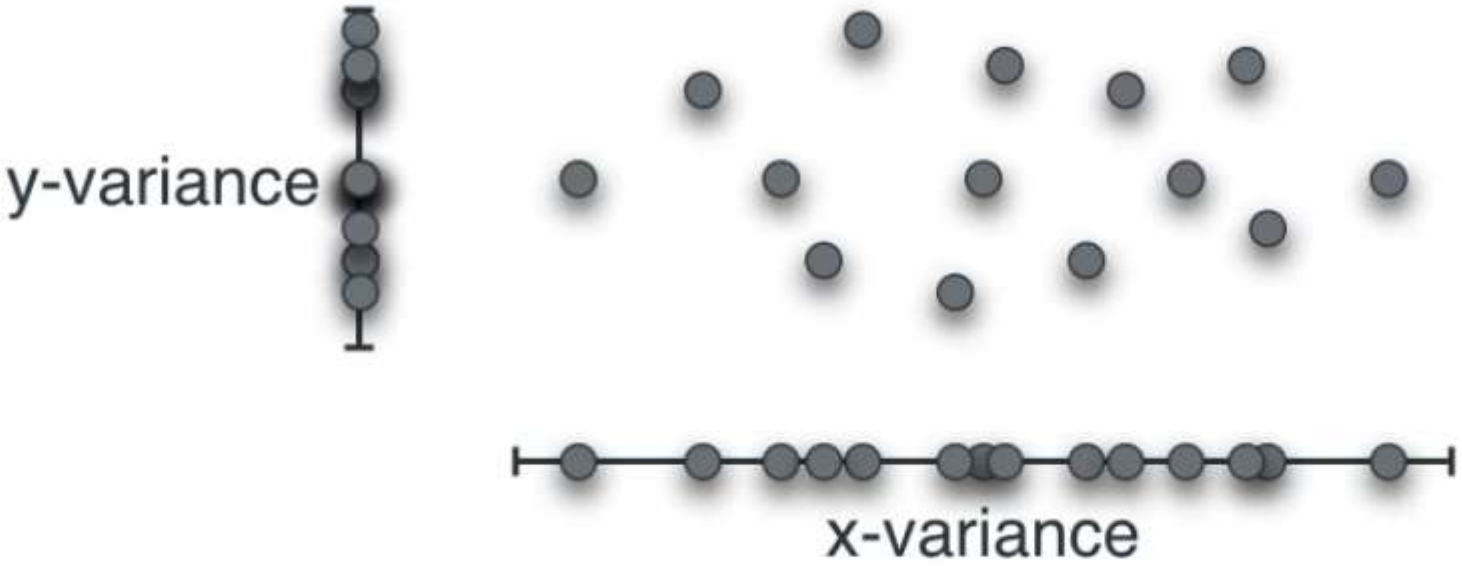
$$\text{Variance} = \frac{5^2+0^2+5^2}{3} = 50/3$$

Variance



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

Find Variance



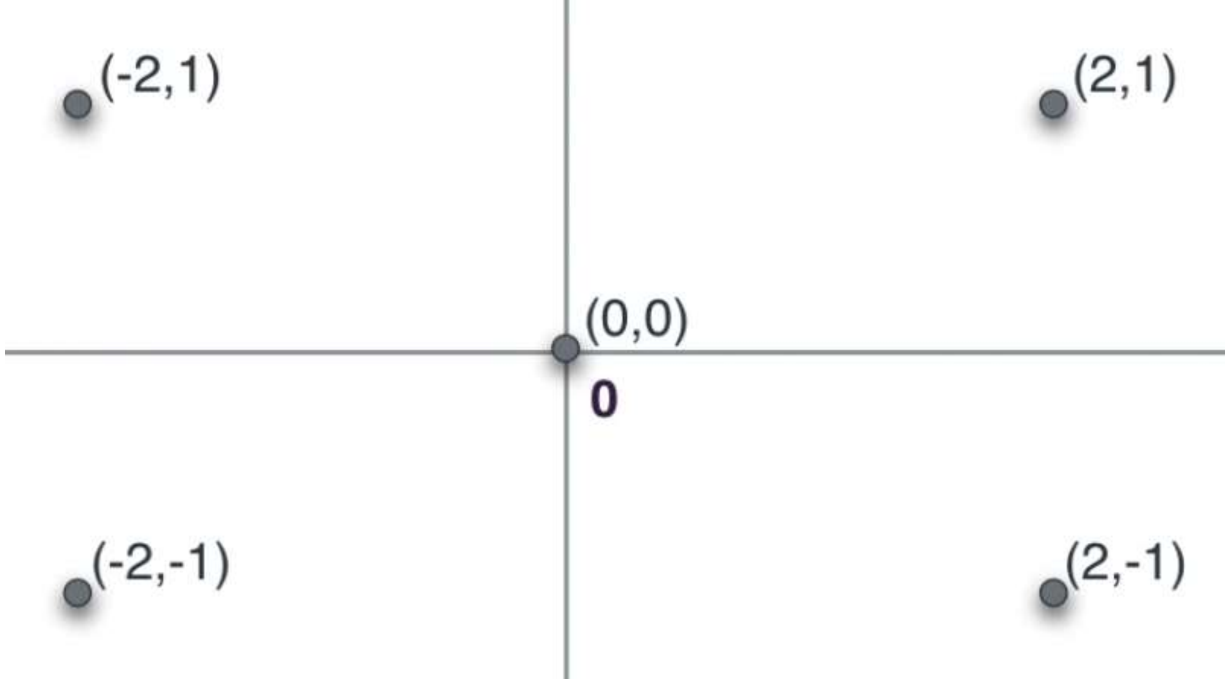
Find Variance & How to Differentiate

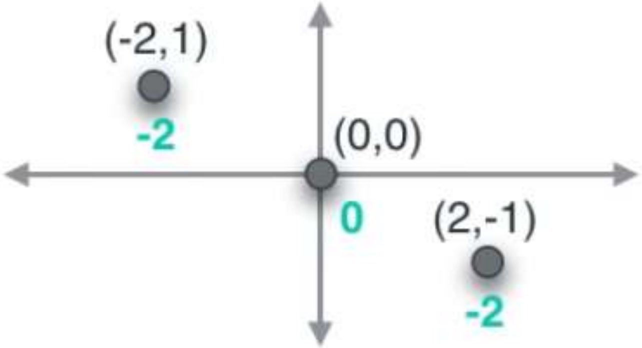
$$\text{x-variance} = \frac{2^2+0^2+2^2}{3} = 8/3$$

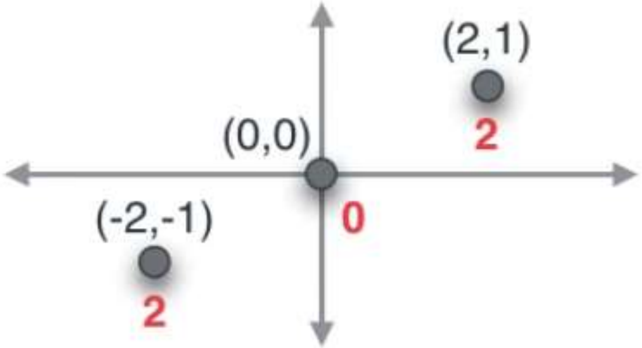
$$\text{y-variance} = \frac{1^2+0^2+1^2}{3} = 2/3$$

Covariance

Sum of Product of coordinates







$$\text{covariance} = \frac{(-2) + 0 + (-2)}{3} = -4/3$$

$$\text{covariance} = \frac{2 + 0 + 2}{3} = 4/3$$

$$cov_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

$cov_{x,y}$ = covariance between variable a and y

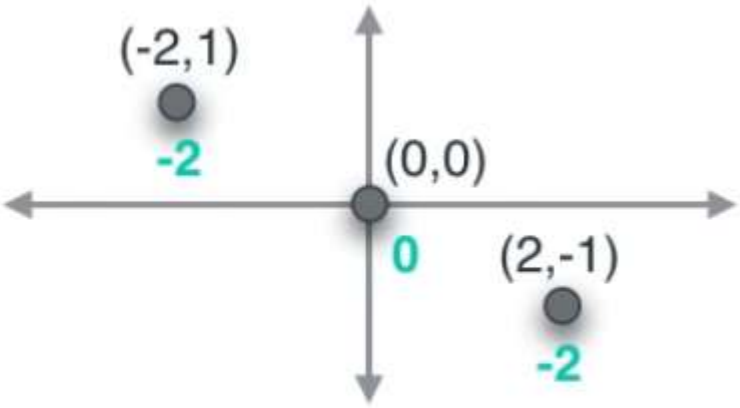
x_i = data value of x

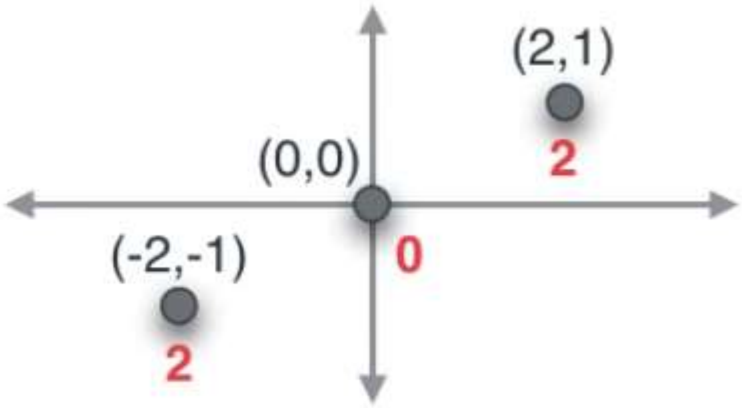
y_i = data value of y

$\bar{x}$ = mean of x

$\bar{y}$ = mean of y

N = number of data values





$$\text{covariance} = \frac{(-2) + 0 + (-2)}{3} = -4/3$$

$$\text{covariance} = \frac{2 + 0 + 2}{3} = 4/3$$

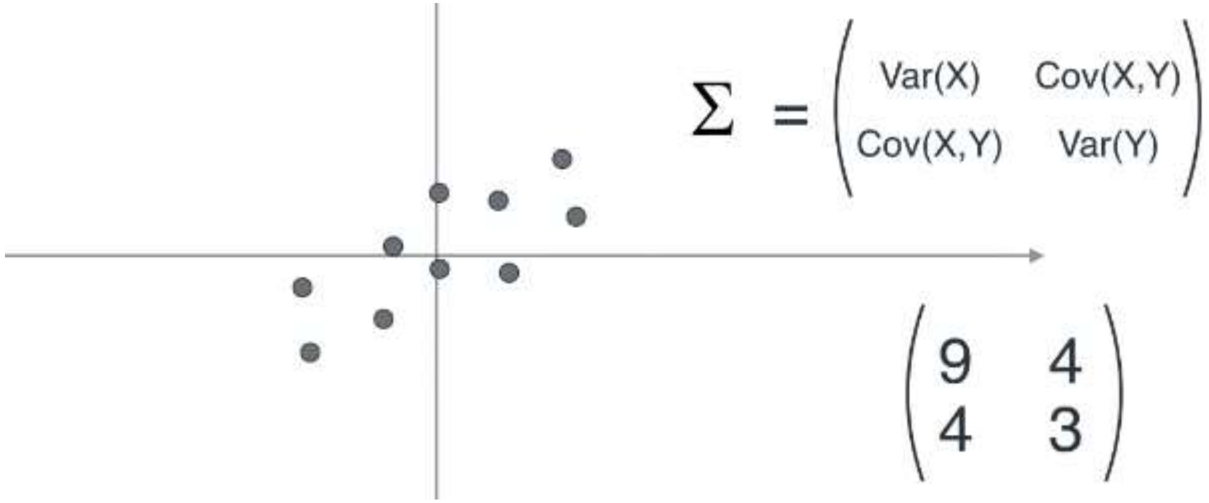
$$\text{covariance} = \frac{-2 + 0 + 2 + 0 + 0 + 0 + 2 + 0 + -2}{9} = 0$$

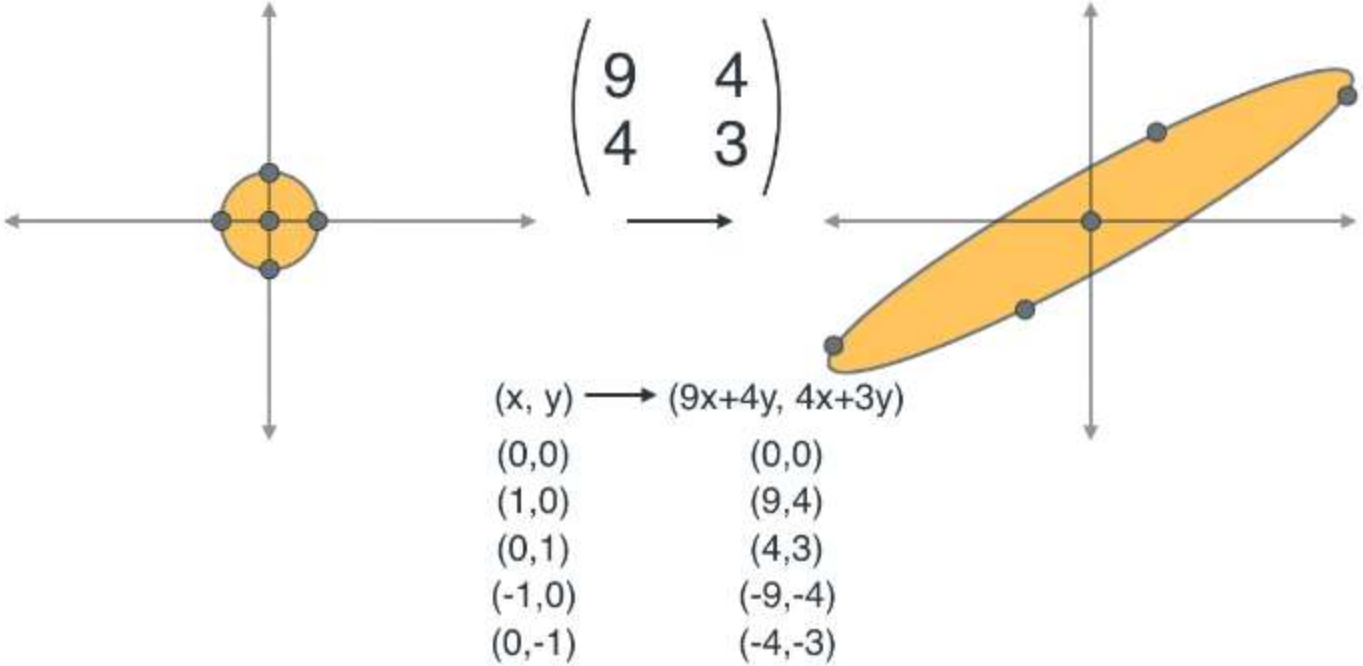
negative covariance

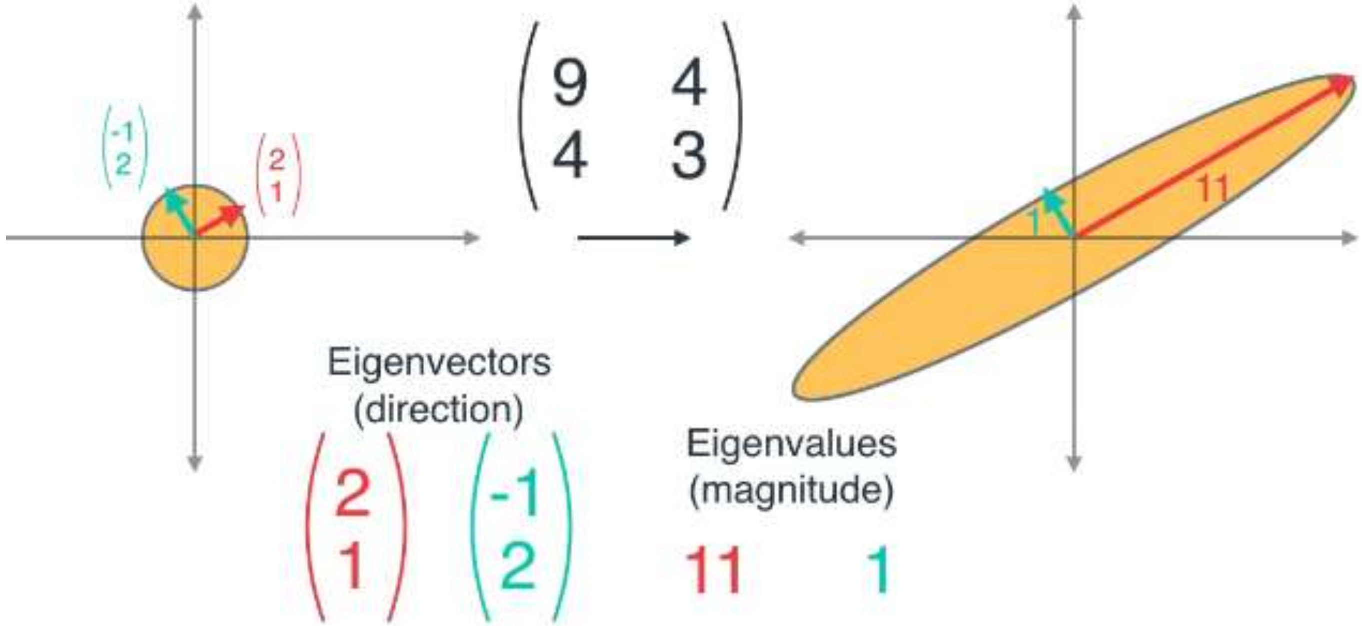
covariance zero (or very small)

positive covariance

Covariance Matrix

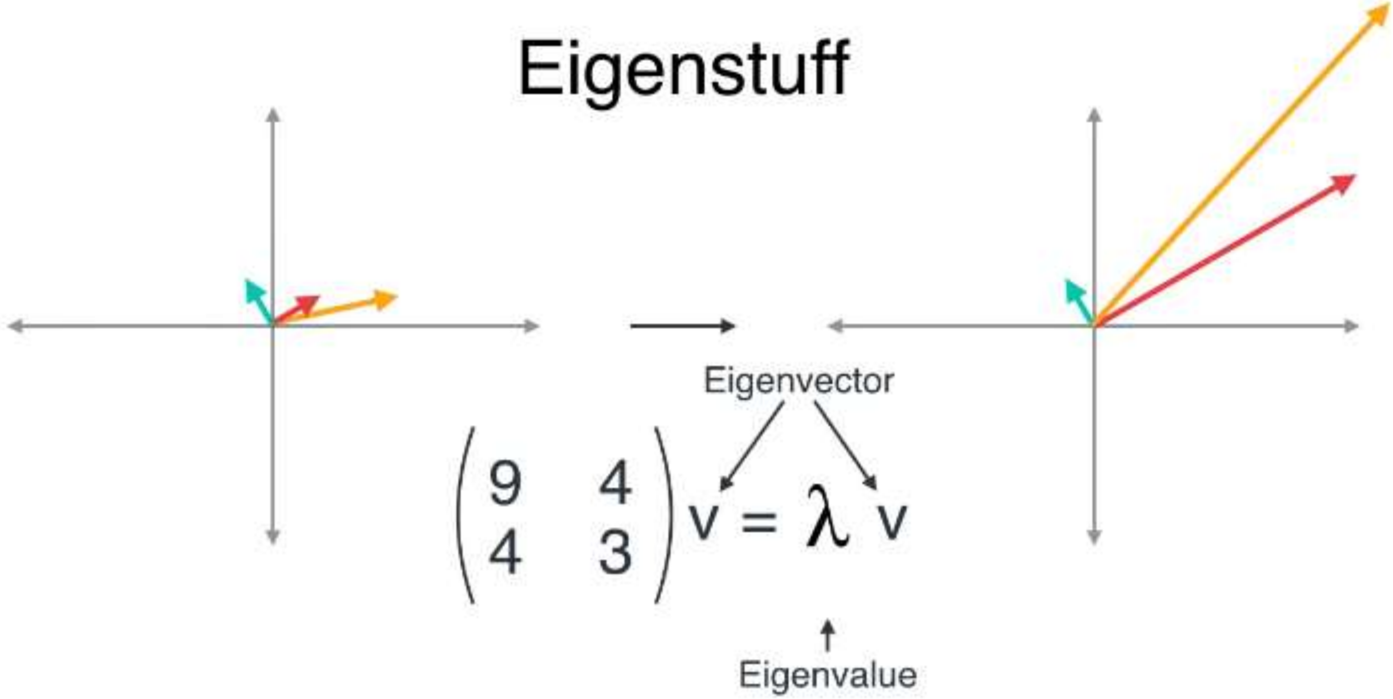






-1
2
2
1
9 4
4 3
11
1
Eigenvectors
(direction)
2
1
-1
2
Eigenvalues
(magnitude)
11
1

Eigenstuff



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

Eigen Values

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

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

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

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

Principal Component Analysis

STEP 1: STANDARDIZATION

$$z = \frac{value - mean}{standard\ deviation}$$

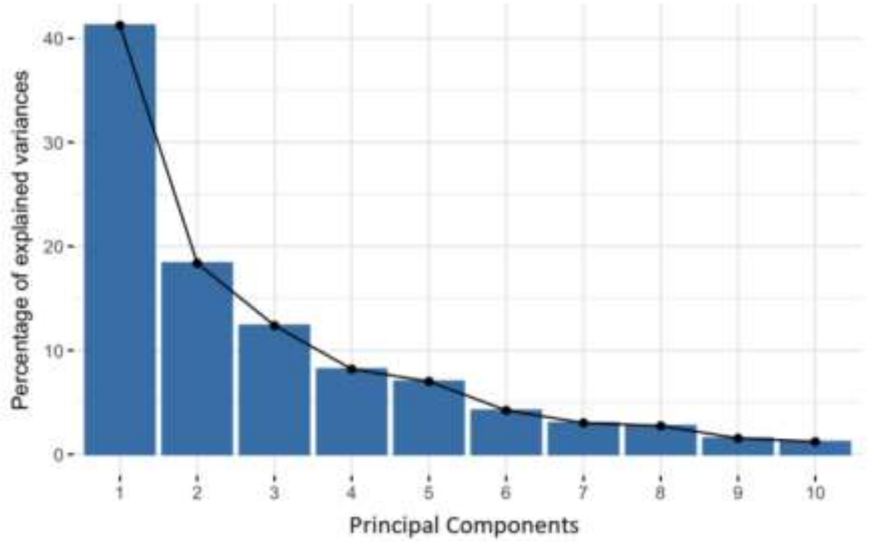
STEP 2: COVARIANCE MATRIX COMPUTATION

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

STEP 3: COMPUTE THE EIGENVECTORS AND EIGENVALUES OF THE COVARIANCE MATRIX TO IDENTIFY THE PRINCIPAL COMPONENTS

$$v1 = \begin{bmatrix} 0.6778736 \\ 0.7351785 \end{bmatrix} \qquad \lambda_1 = 1.284028$$

$$v2 = \begin{bmatrix} -0.7351785 \\ 0.6778736 \end{bmatrix} \qquad \lambda_2 = 0.04908323$$



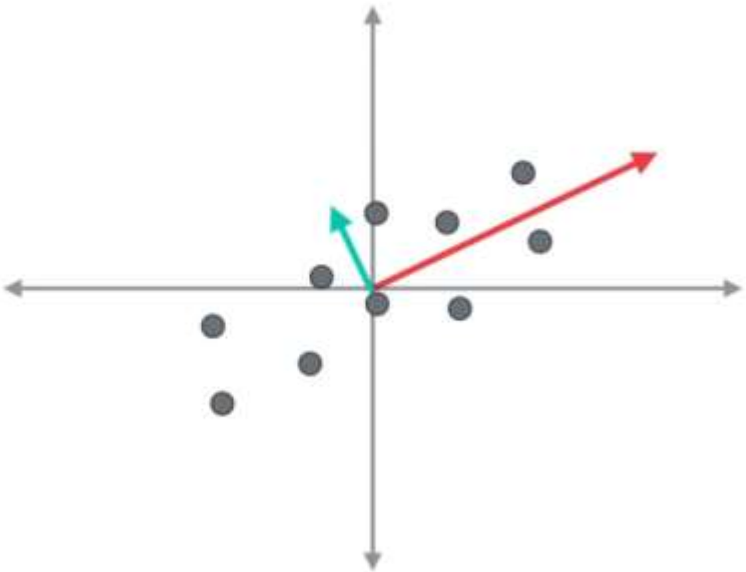
STEP 4: FEATURE VECTOR

$$\begin{bmatrix} 0.6778736 & -0.7351785 \\ 0.7351785 & 0.6778736 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.6778736 \\ 0.7351785 \end{bmatrix}$$

LAST STEP: RECAST THE DATA ALONG THE PRINCIPAL COMPONENTS AXES

$$FinalDataSet = FeatureVector^T * StandardizedOriginalDataSet^T$$

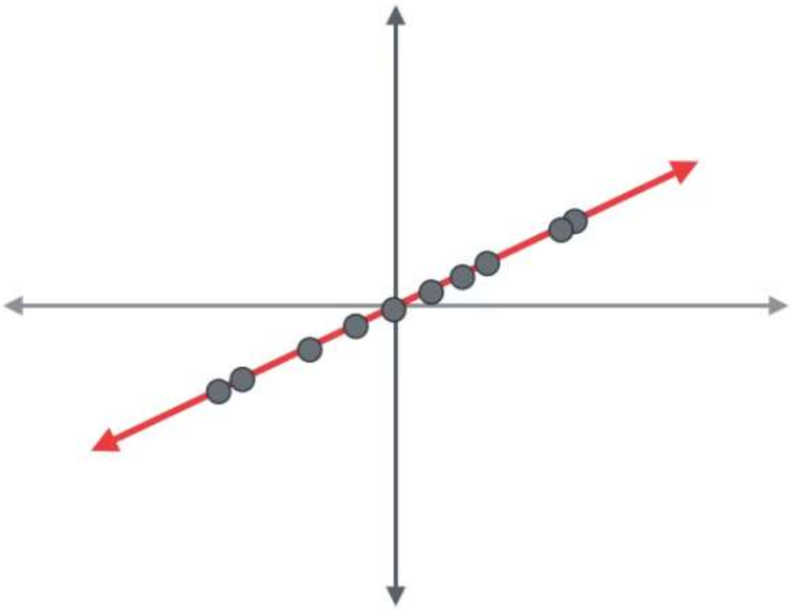
Principal Component Analysis



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

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

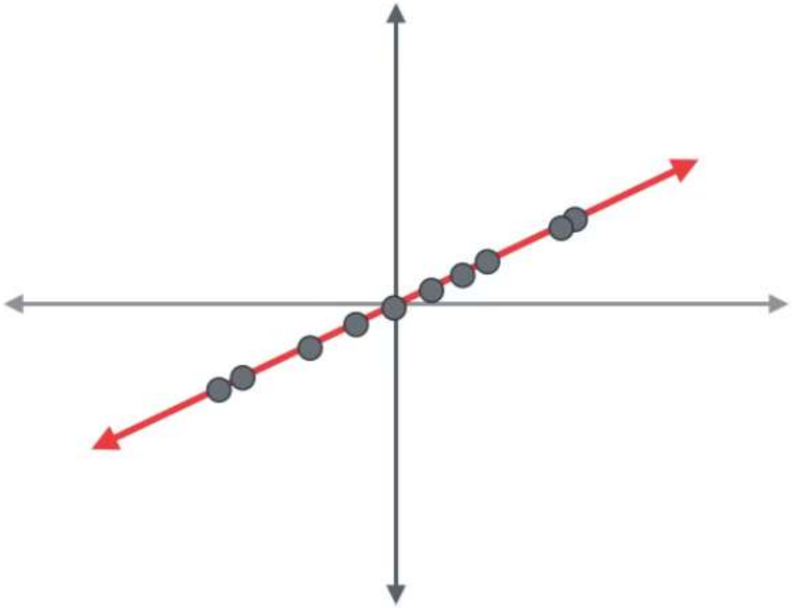
11 1 Eigenvalues (magnitude)



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

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

11 Eigenvalues (magnitude)

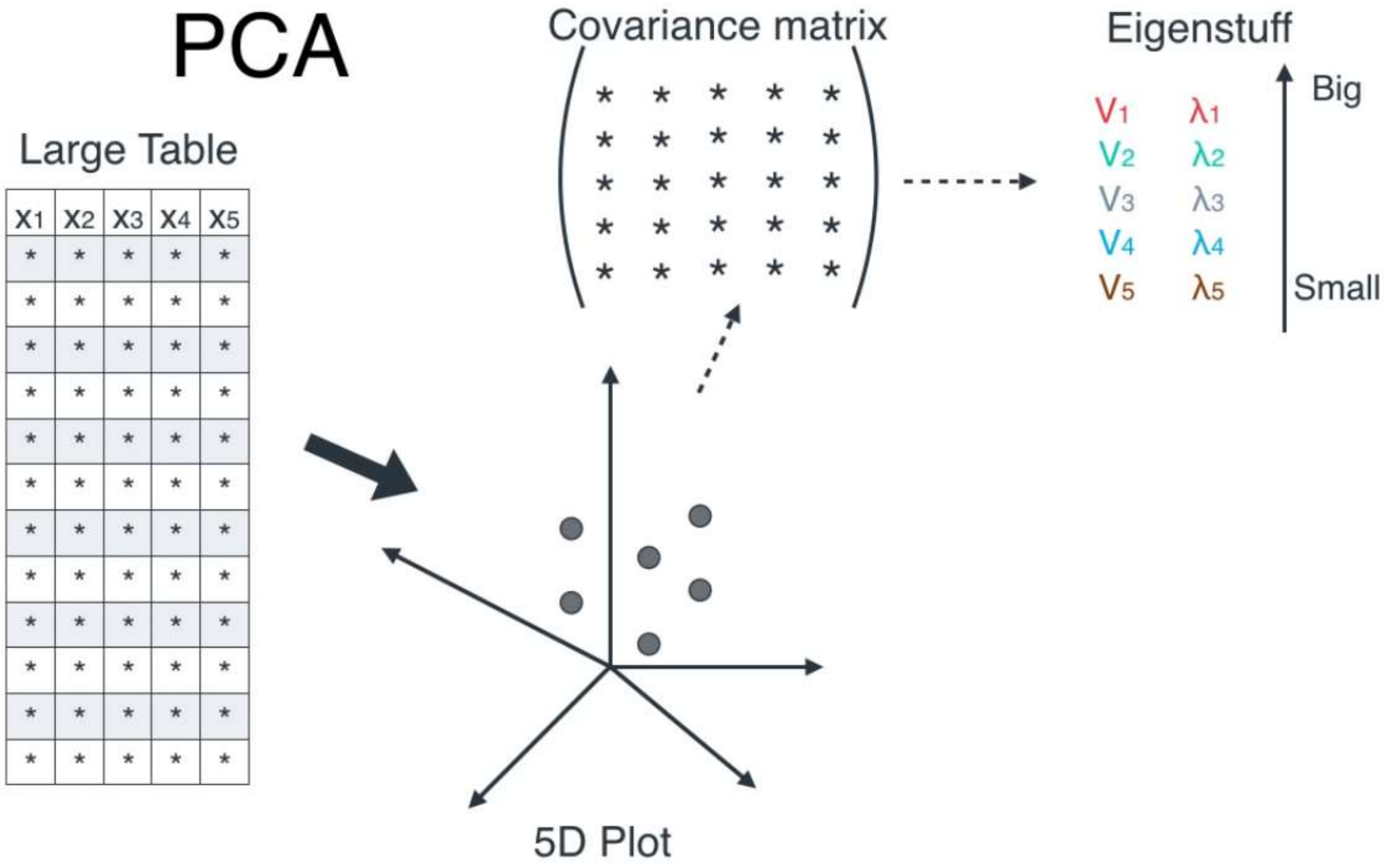


$$\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

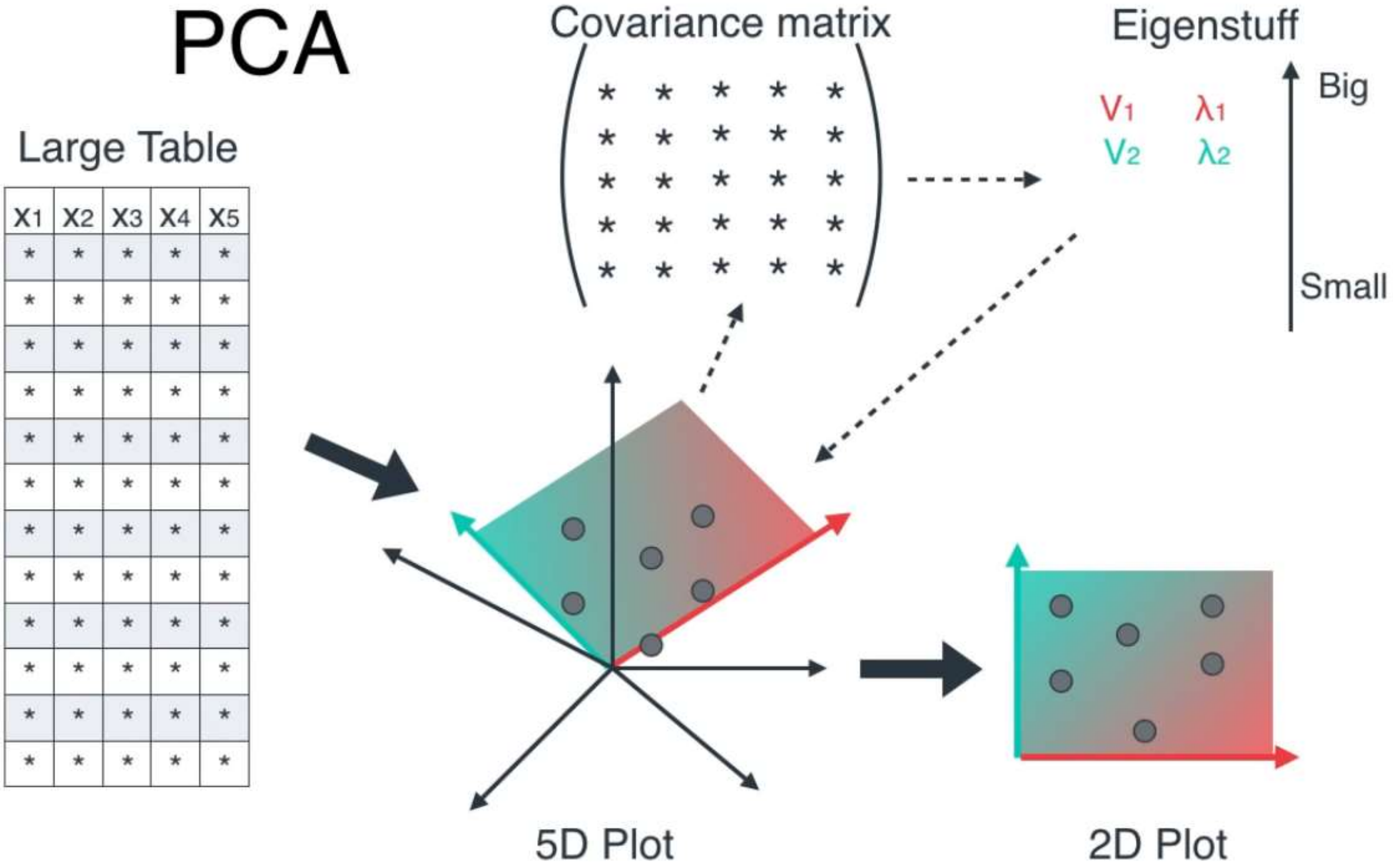
$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ Eigenvectors (direction)

11 Eigenvalues (magnitude)

PCA – Big Picture



PCA


Large Table
X1 X2 X3 X4 X5
Covariance matrix
Eigenstuff
V1 λ1
V2 λ2
Big
Small
5D Plot
2D Plot

X	Y	(XI-XBAR)*2	(YI-YBAR)*2	(XI-XBAR)*(YI-YBAR)	(YI-XBAR)*(XI-YBAR)
4	11	16	6.25	-10	-10
8	4	0	20.25	0	0
13	5	25	12.25	-17.5	-17.5
7	14	1	30.25	-5.5	-5.5
8	8.5	14	23	-11	-11

Covariance Matrix : **A =** $\begin{bmatrix} \mathbf{14} & \mathbf{-11} \\ \mathbf{-11} & \mathbf{23} \end{bmatrix}$

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

$\lambda 1 > \lambda 2$ Eigen Vector of $\lambda 1$,

$$\left(\begin{bmatrix} \mathbf{14} & \mathbf{-11} \\ \mathbf{-11} & \mathbf{23} \end{bmatrix} - \lambda \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}\right) = 0$$

$$\left(\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}\right) = 0$$

(14-λ)(23- λ) – (-11x-11) = 0

λ^2 -37λ+201= 0

$\lambda 1 = 30.3849$, $\lambda 2 = 6.6151$