Principal Component Analysis

Principal Component Analysis

- Our discussions of **unsupervised learning** have focused on clustering techniques, which seek to “discover” labels on feature data that has no historical labels.
- We will now shift towards unsupervised algorithms that focus on **dimension reduction**.

Principal Component Analysis

- Motivation of Dimension Reduction:
    - Imagine a dataset with 30+ features, how would you understand the key features?
    - Visualization and Data Analysis have limitations when the number of feature dimensions increases.

Principal Component Analysis

- Dimensionality Reduction Outcomes:
  - Understand which features describe the most variance in the data set.
  - Aid human understanding of large feature sets, especially through visualization.

Principal Component Analysis

- Important Note:
    - Dimensionality Reduction algorithms such as PCA **do not** simply choose a subset of the existing features.
    - They create **new** dimensional components that are combinations of proportions of the existing features.

Principal Component Analysis

- Section Overview
  - Theory and Intuition of PCA.
  - Manually create PCA Algorithm.
  - Utilize Scikit-Learn to perform PCA.
  - PCA Exercise Project Overview
  - PCA Exercise Solution

Let’s get started!

Principal Component Analysis

Theory and Intuition - Part One

PIERIAN DATA

Principal Component Analysis

- Let's quickly cover the history and motivation behind the main ideas of PCA (Principal Component Analysis).
- Later on we will focus on the mathematics of PCA.

Principal Component Analysis

- 1901: Karl Pearson publishes “*On Lines and Planes of Closest Fit to Systems of Points in Space*” based on the principal axis theorem in the field of geometry.
- Pearson was the protégé of Francis Galton and the Pearson Correlation Coefficient is named after him.

Principal Component Analysis

- 1933: American mathematician and economist Harold Hotelling independently develops and names Principal Component Analysis in this publication, “*Analysis of a complex of statistical variables into principal components*”.

Principal Component Analysis

- Hotelling’s paper perfectly describes the purpose of PCA:
  - Analyzing a complex set of variables into its principal components.
- Let’s review the motivation and basic idea behind Principal Component Analysis.

Principal Component Analysis

- Principal Component Analysis Outcomes:
  - Reduce number of dimensions in data.
  - Show which features explain the most variance in the data.

Principal Component Analysis

- Dimension Reduction

Principal Component Analysis

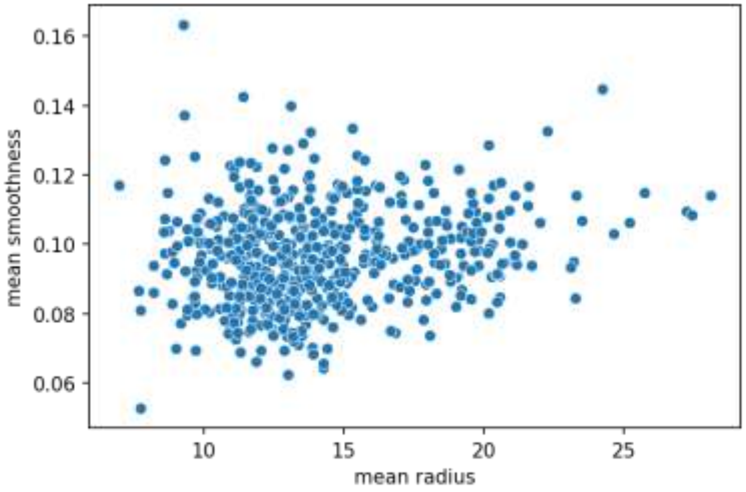
- Dimension Reduction

	mean radius	mean smoothness
0	17.99	0.11840
1	20.57	0.08474
2	19.69	0.10960
3	11.42	0.14250
4	20.29	0.10030
...	...	...
564	21.56	0.11100
565	20.13	0.09780
566	16.60	0.08455
567	20.60	0.11780
568	7.76	0.05263

Principal Component Analysis

- Dimension Reduction

	mean radius	mean smoothness
0	17.99	0.11840
1	20.57	0.08474
2	19.69	0.10960
3	11.42	0.14250
4	20.29	0.10030
...	...	...
564	21.56	0.11100
565	20.13	0.09780
566	16.60	0.08455
567	20.60	0.11780
568	7.76	0.05263



Principal Component Analysis

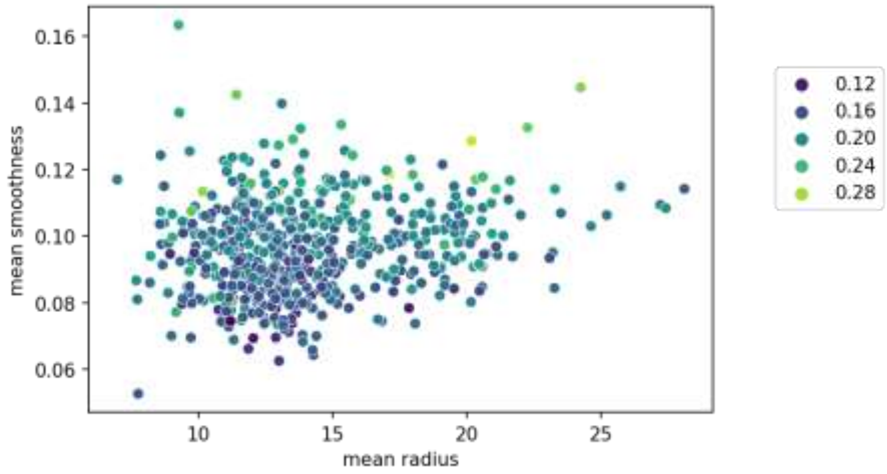
- Dimension Reduction

	mean radius	mean smoothness	mean symmetry
0	17.99	0.11840	0.2419
1	20.57	0.08474	0.1812
2	19.69	0.10960	0.2069
3	11.42	0.14250	0.2597
4	20.29	0.10030	0.1809
...	...	...	...
564	21.56	0.11100	0.1726
565	20.13	0.09780	0.1752
566	16.60	0.08455	0.1590
567	20.60	0.11780	0.2397
568	7.76	0.05263	0.1587

Principal Component Analysis

- Dimension Reduction

	mean radius	mean smoothness	mean symmetry
0	17.99	0.11840	0.2419
1	20.57	0.08474	0.1812
2	19.69	0.10960	0.2069
3	11.42	0.14250	0.2597
4	20.29	0.10030	0.1809
...	...	...	...
564	21.56	0.11100	0.1726
565	20.13	0.09780	0.1752
566	16.60	0.08455	0.1590
567	20.60	0.11780	0.2397
568	7.76	0.05263	0.1587



Principal Component Analysis

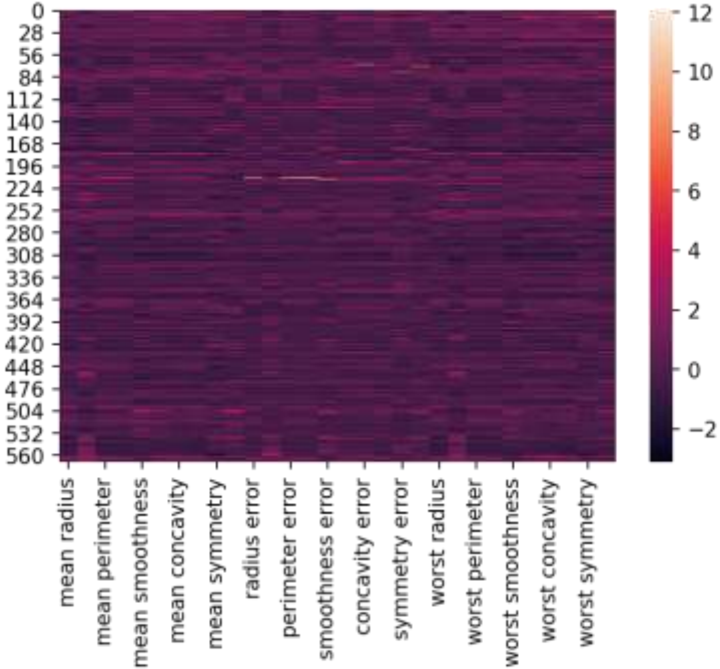
- Dimension Reduction

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...
...	...	...	...	...	...	...	...	...	...	...	...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...

569 rows × 30 columns

Principal Component Analysis

- Dimension Reduction



PIERIAN DATA

Principal Component Analysis

- Dimension Reduction
  - Helps visualize and understand complex data sets.
  - Can also act as a simpler data set for training data for machine learning algorithms.
    - Reduce dimensions then train ML Algorithm on smaller data set.

Principal Component Analysis

- Dimension Reduction
  - Helps reduce N features to a desired smaller set of components through a **transformation**.
  - It does **not** simply select a subset of features.

Principal Component Analysis

- Variance Explained

Principal Component Analysis

- Variance Explained
  - We've often seen that certain features are more important or have more explanatory power than other features.
  - For example, size of a house is probably much more important than the color of a house when explaining the price of a house for sale.

Principal Component Analysis

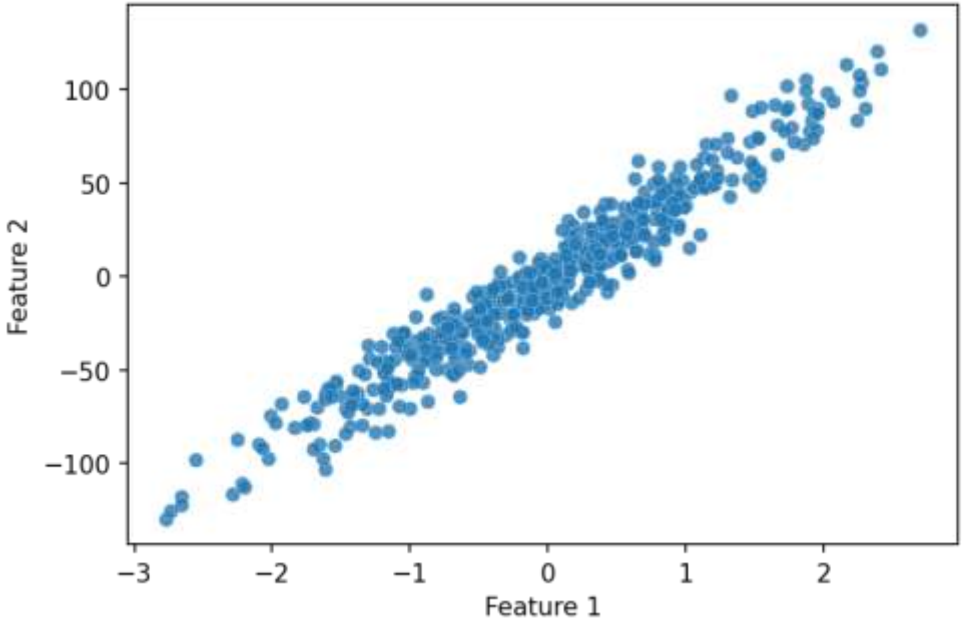
- Variance Explained
  - This idea of more important features is easy to understand when we can directly correlate features to a known label. But what about unlabeled data?
  - What measurement can we use to determine feature “importance”?

Principal Component Analysis

- Variance Explained
  - Measure the proportion to which each feature accounts for dispersion in the data set.

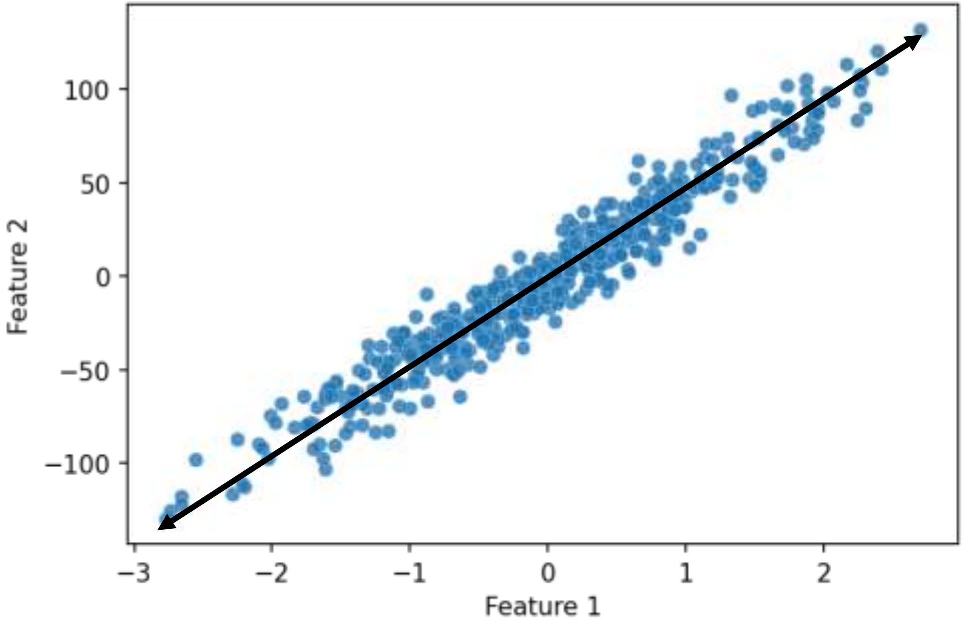
Principal Component Analysis

- Variance Explained



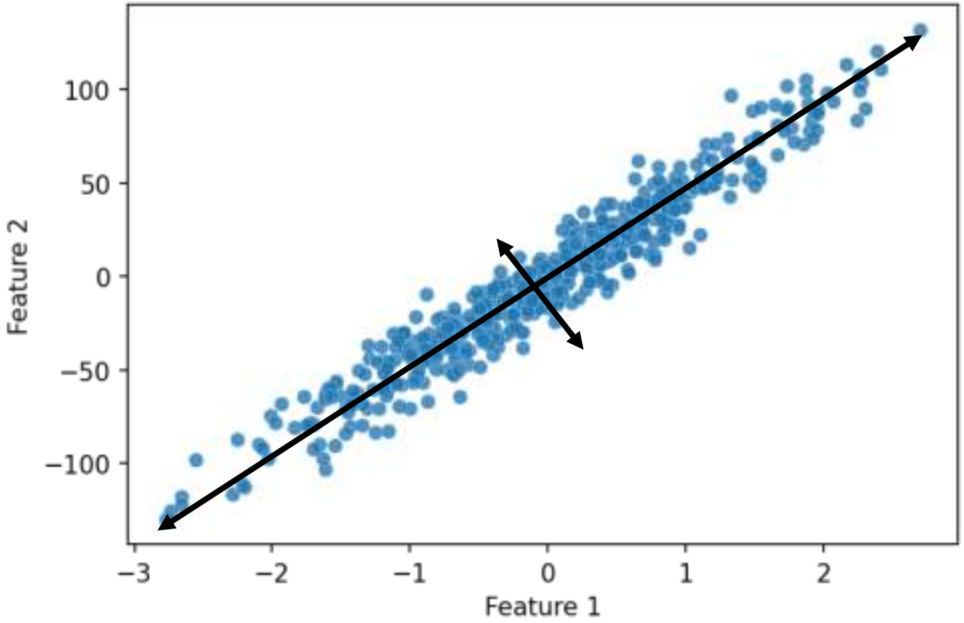
Principal Component Analysis

- Variance Explained



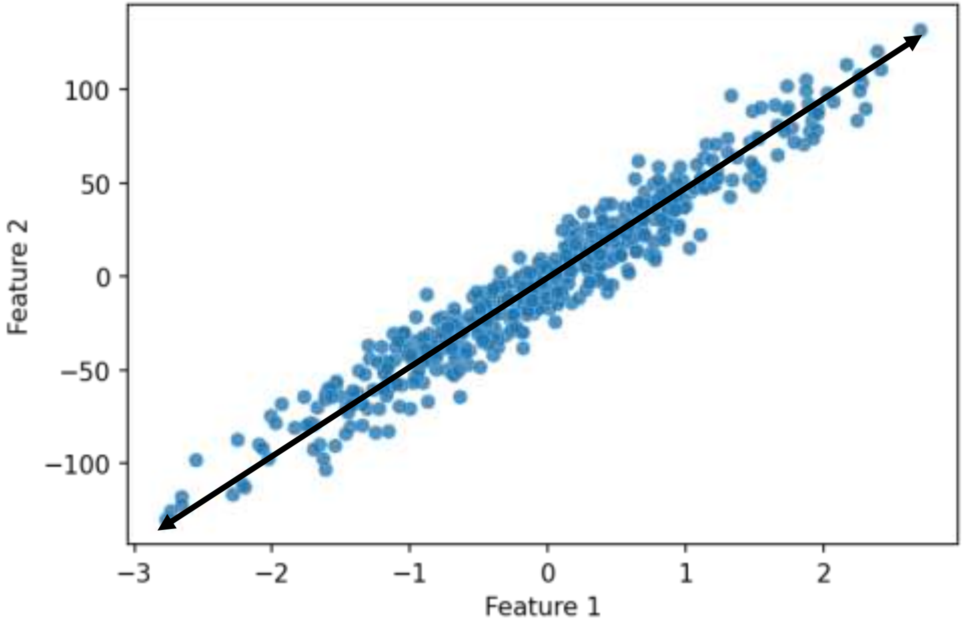
Principal Component Analysis

- Variance Explained



Principal Component Analysis

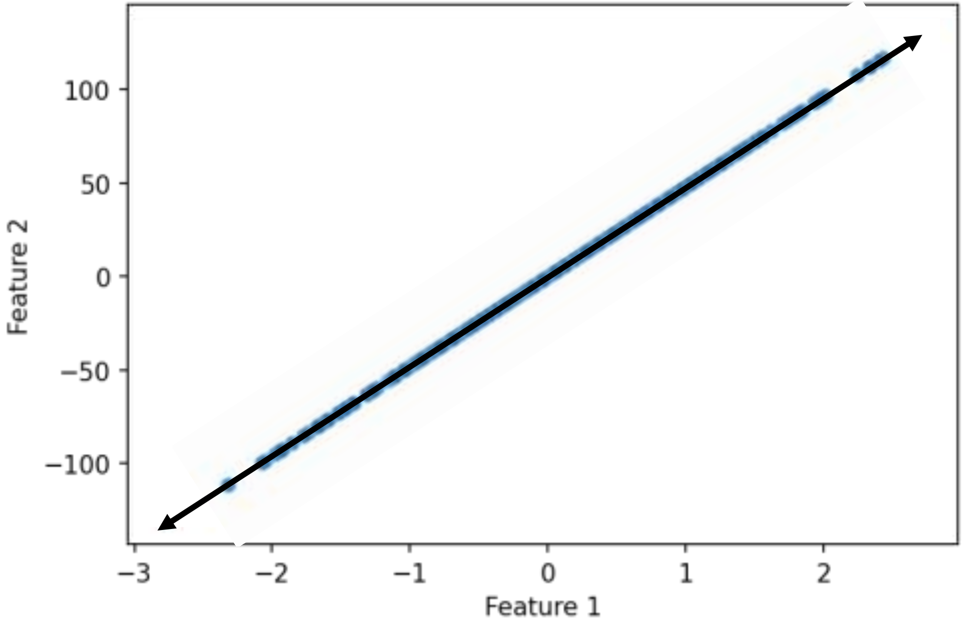
- Variance Explained



PIERIAN DATA

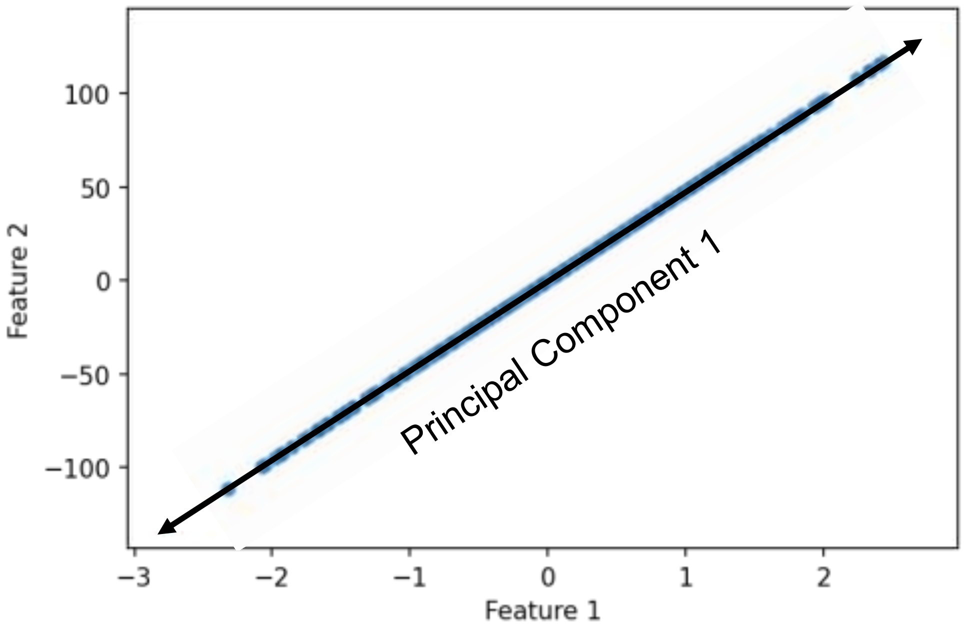
Principal Component Analysis

- Variance Explained



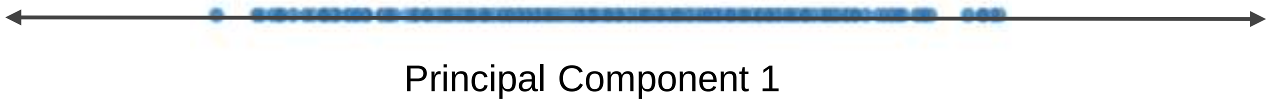
Principal Component Analysis

- Variance Explained



Principal Component Analysis

- Variance Explained



Principal Component Analysis

- Variance Explained
  - Principal Component is a linear combination of original features.
  - The more variance the original feature accounts for, the more influence it has over the principal components.

Principal Component Analysis

- Variance Explained
  - Here we went from 2 features down to 1 principal component.
  - This single principal component can "explain" some percentage of the original data, for example 90% of variance explained by principal component.

Principal Component Analysis

- Variance Explained
  - 100% of the variance in the data is explained by all the original features.
  - We trade off some of the explained variance for less dimensions.
  - This can be significant savings for data sets with many dimensions, but only a few strong features.

Principal Component Analysis

- Let’s continue by exploring how Principal Component Analysis actually works mathematically.

Principal Component Analysis

Theory and Intuition - Part Two

Principal Component Analysis

- Suggested Reading:
    - Section 10.1 of ISLR covers the topic of Principal Component Analysis.

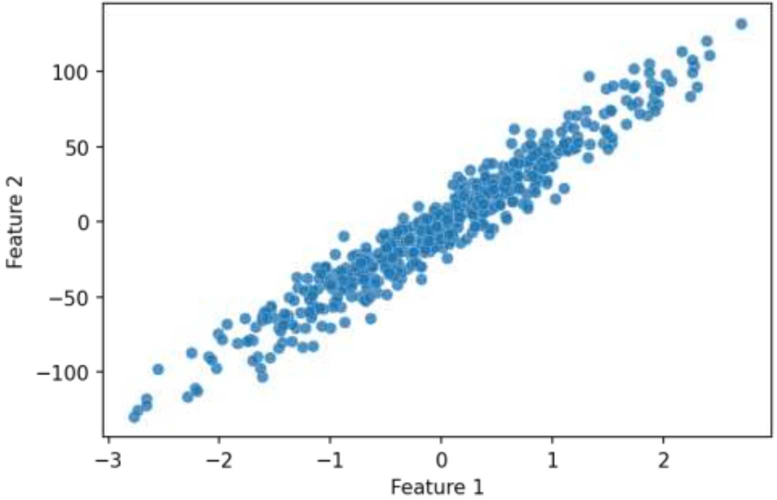
Principal Component Analysis

- Principal Component Analysis operates by creating a new set of dimensions (the principal components) that are normalized linear combinations of the original features.

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \ldots + \phi_{p1}X_p$$

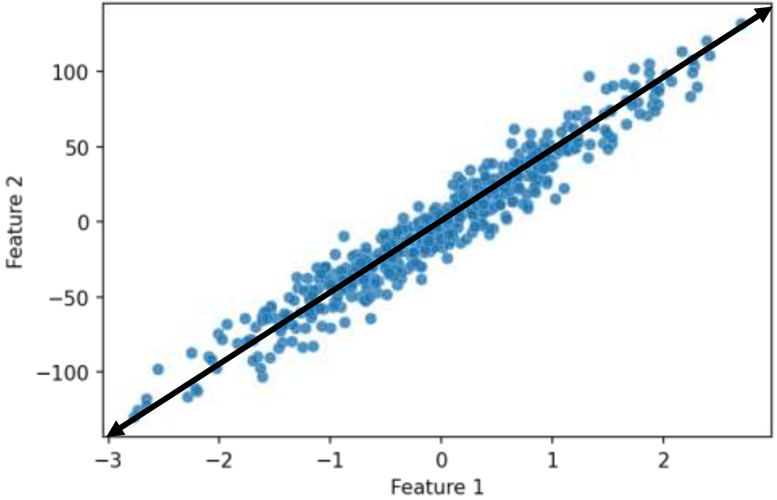
Principal Component Analysis

- Principal Component Analysis:



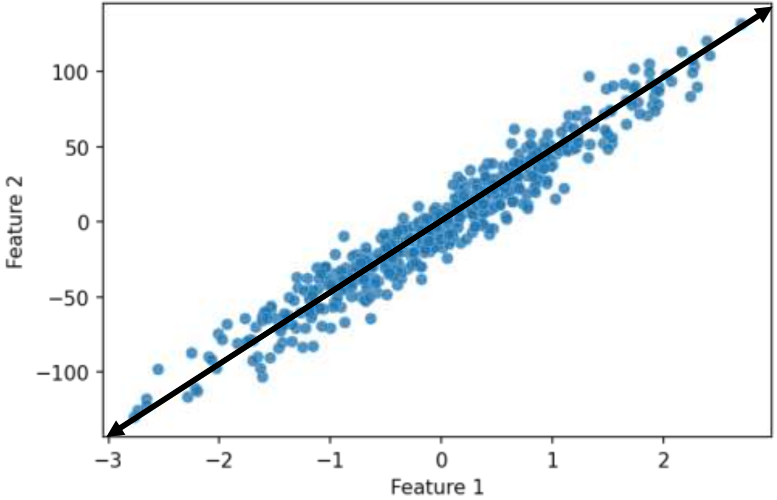
Principal Component Analysis

- Principal Component Analysis:



Principal Component Analysis

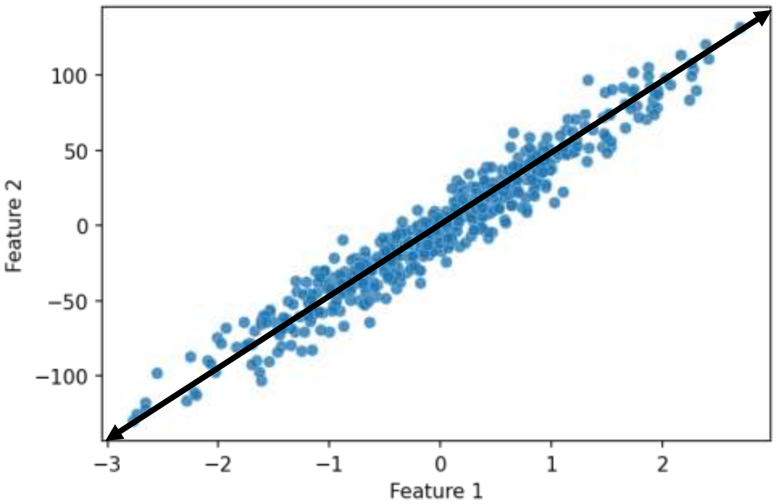
- Principal Component Analysis:



$$Z_1 = \phi_{11} X_1 + \phi_{21} X_2$$

Principal Component Analysis

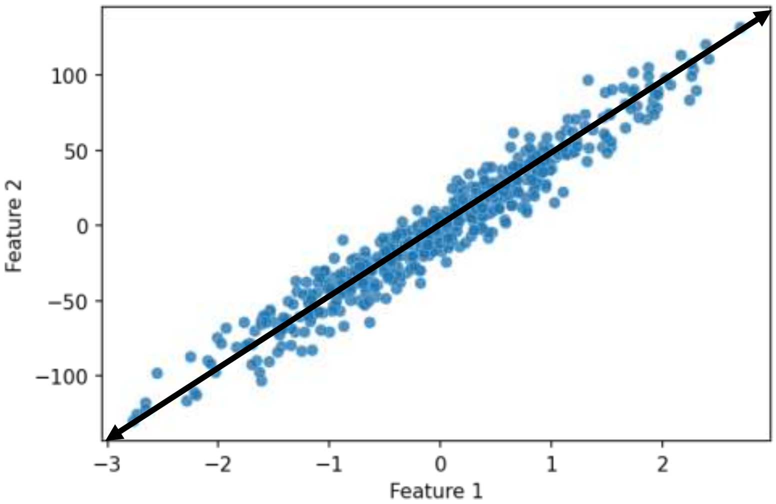
- Principal Component Analysis:



$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2$$

Principal Component Analysis

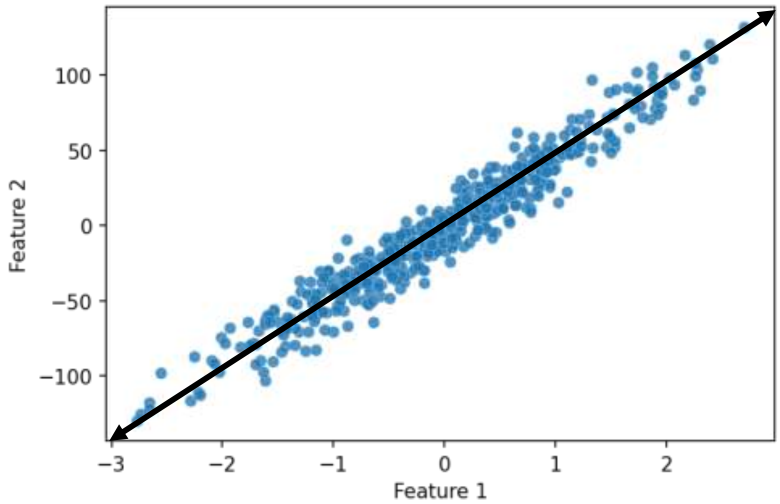
- Principal Component Analysis:



$$Z_1 = \phi_{11} X_1 + \phi_{21} X_2$$

Principal Component Analysis

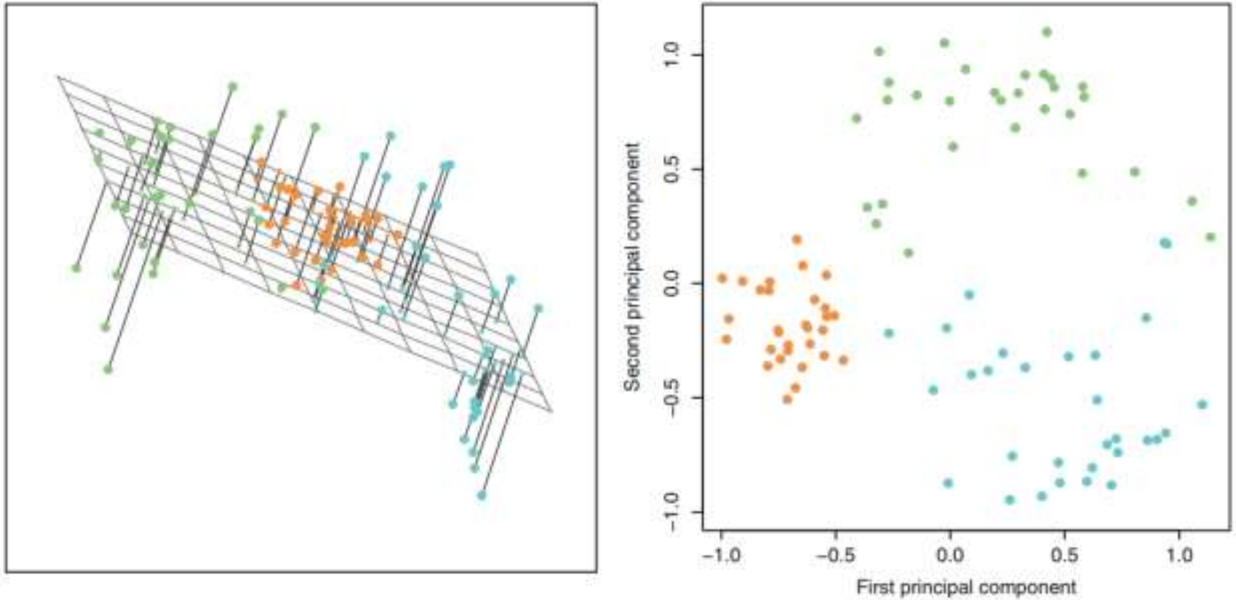
- Principal Component Analysis:



$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2$$

Principal Component Analysis

- Principal Component Analysis:



Principal Component Analysis

- How do we actually calculate these components?
- Let’s walk through the steps visually.

Principal Component Analysis

- Begin with a two dimensional data set:

Principal Component Analysis

- Begin with a two dimensional data set:

X1	X2
1	2
2	1
3	2
4	4
5	3
6	5

Principal Component Analysis

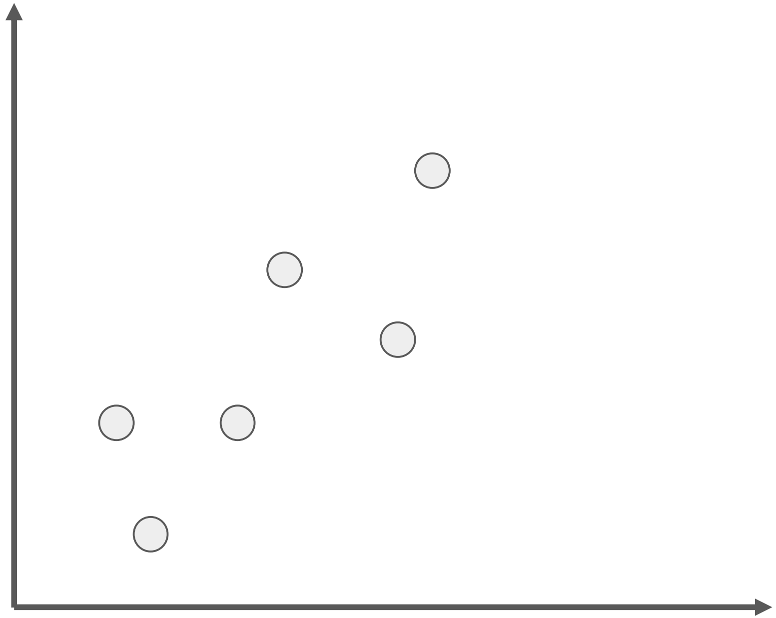
- Begin with a two dimensional data set:

X1	X2
1	2
2	1
3	2
4	4
5	3
6	5

Principal Component Analysis

- Begin with a two dimensional data set:

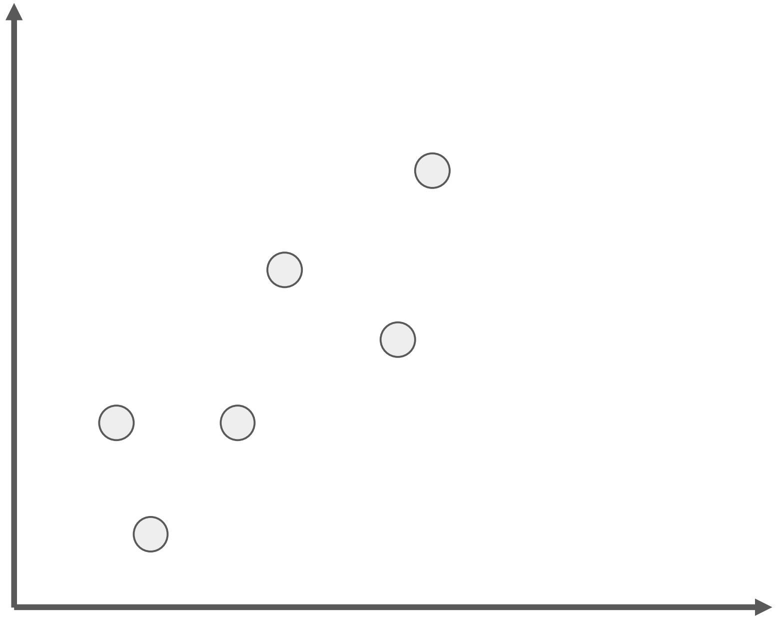
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

- Standardize the data:

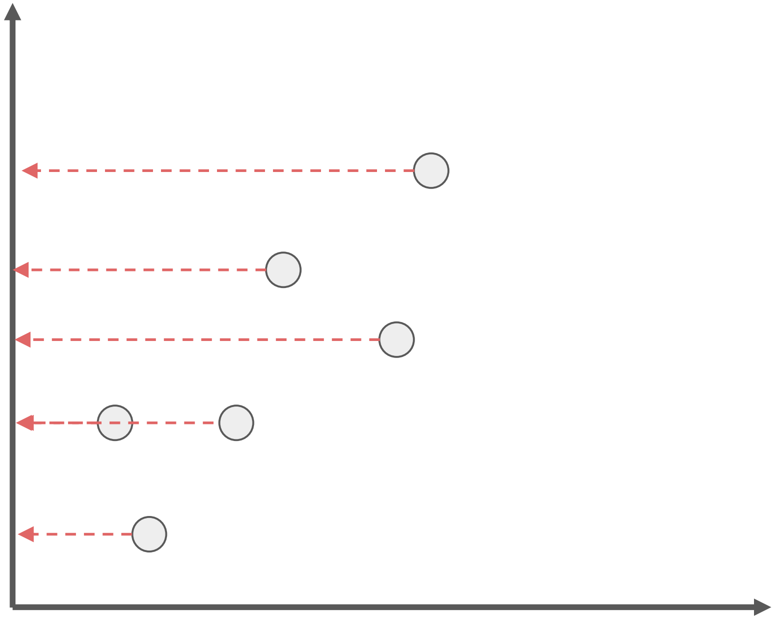
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

- Standardize the data:

X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

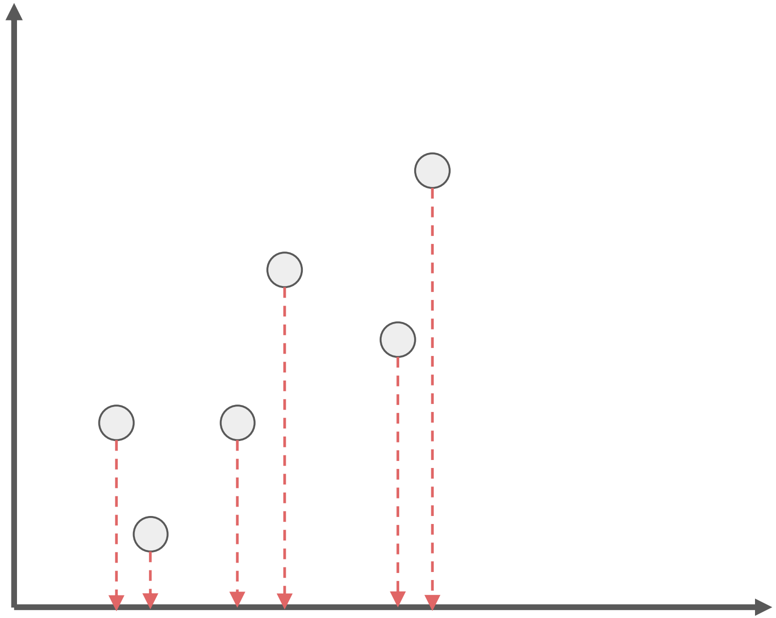
- Standardize the data:

X1	X2
1	2
2	1
3	2
4	4
5	3
6	5

Principal Component Analysis

- Standardize the data:

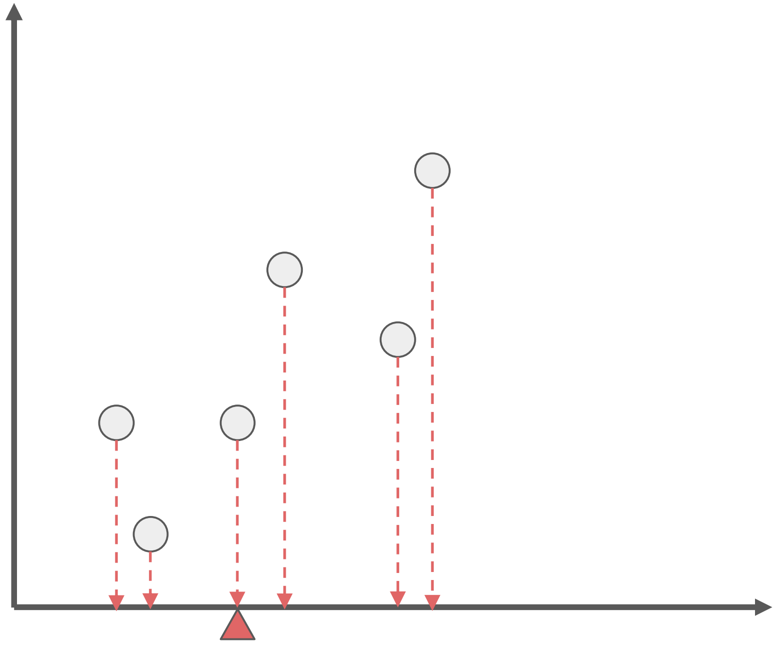
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

- Standardize the data:

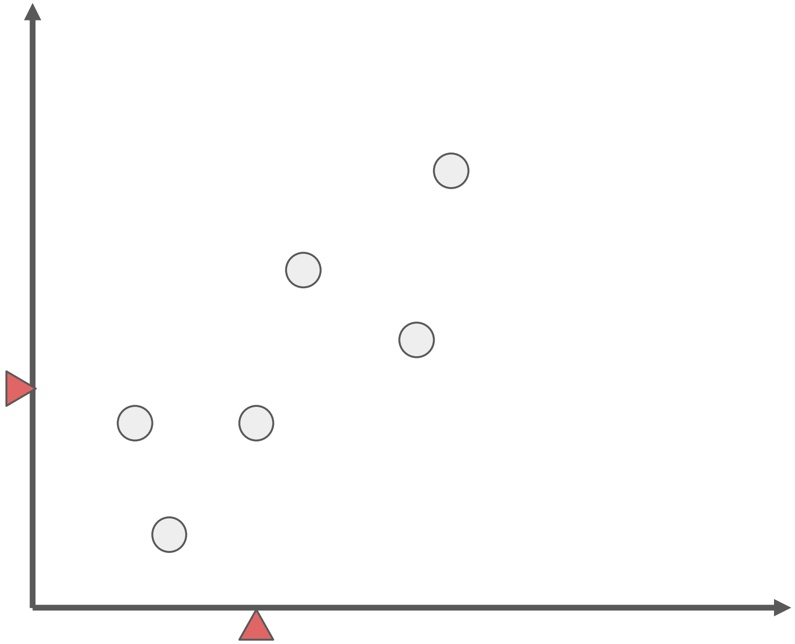
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

- Standardize the data:

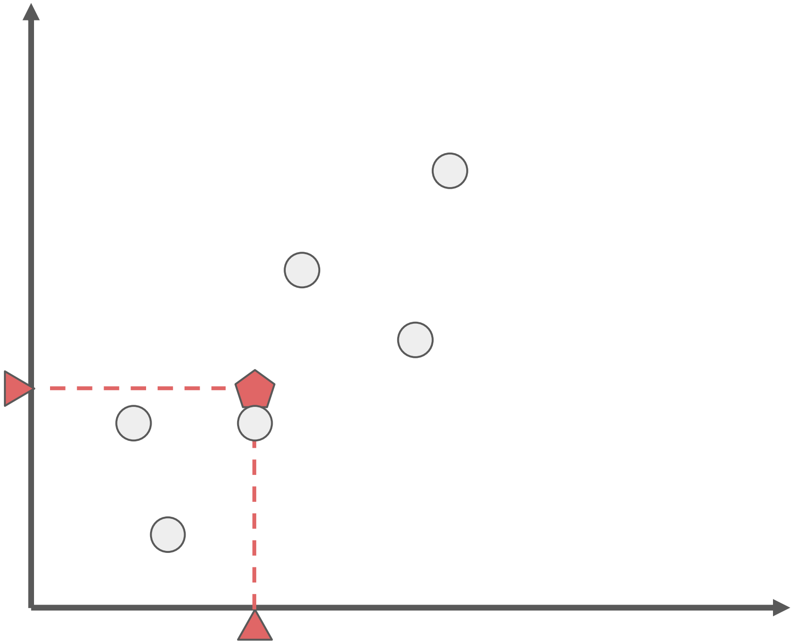
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

- Standardize the data:

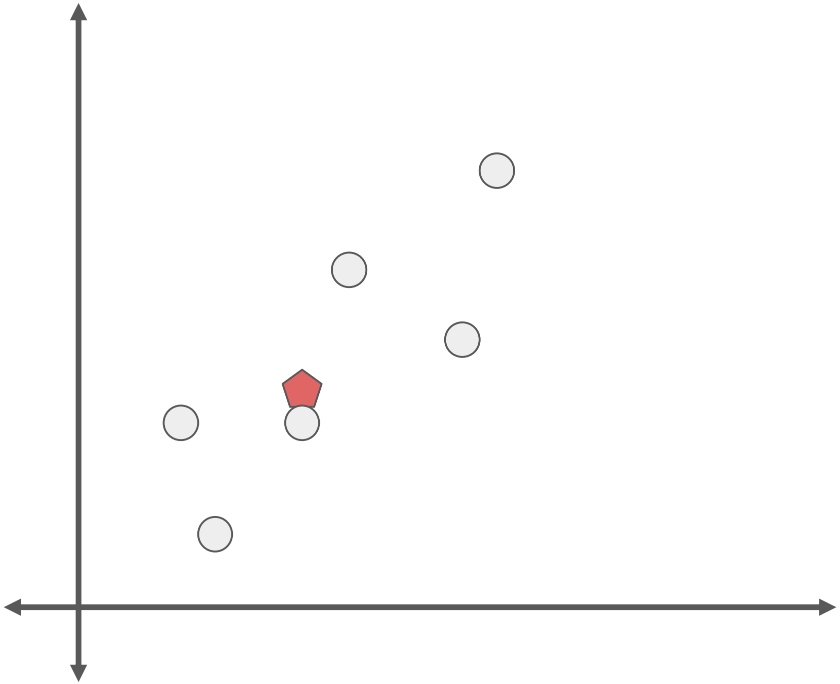
X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



Principal Component Analysis

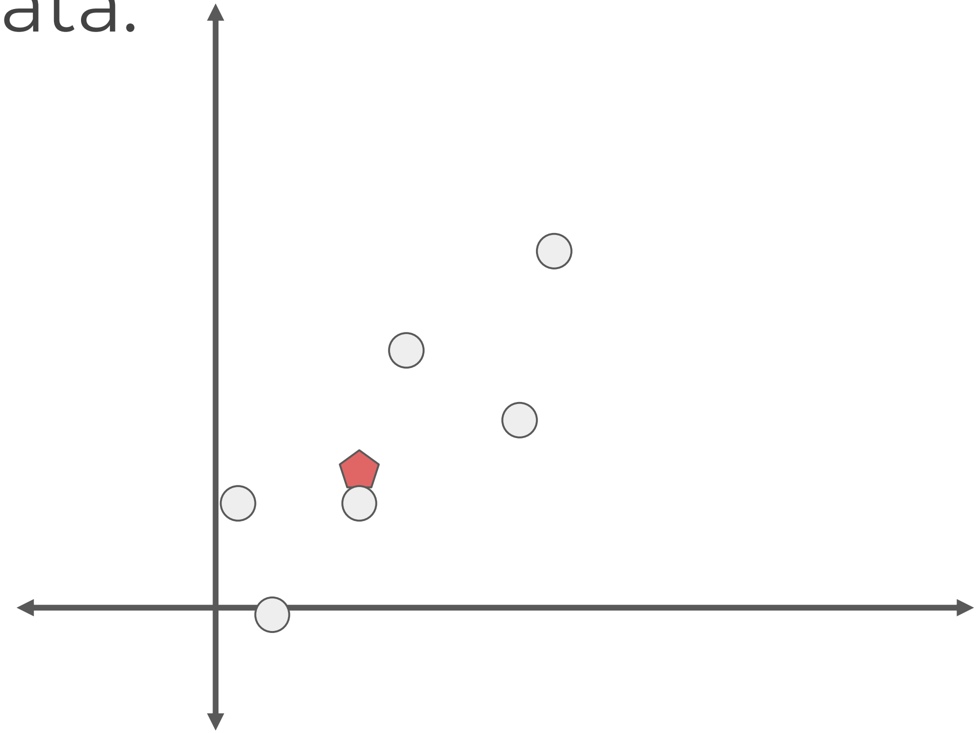
- Standardize the data:

X1	X2
1	2
2	1
3	2
4	4
5	3
6	5



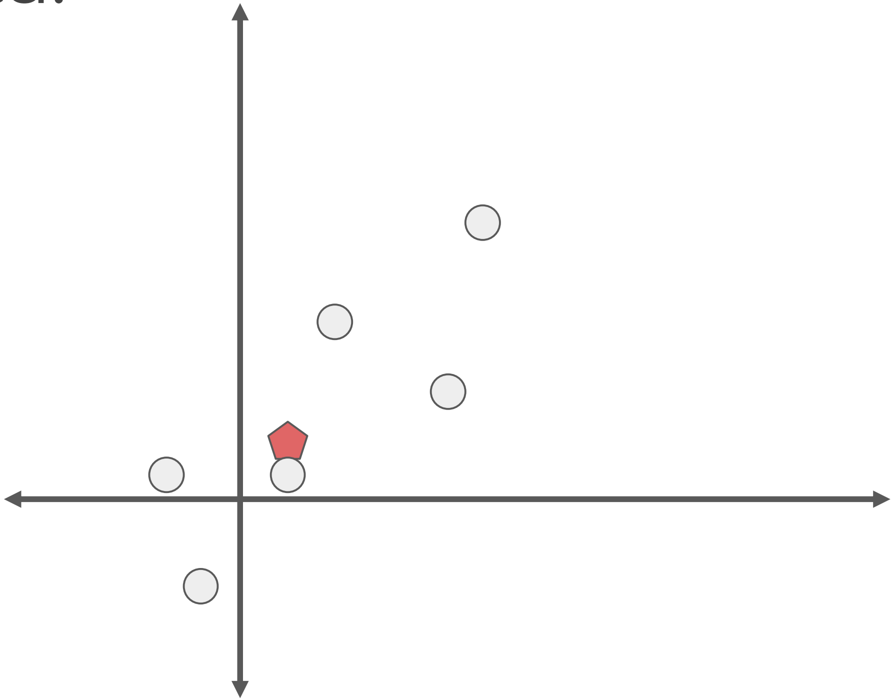
Principal Component Analysis

- Standardize the data:



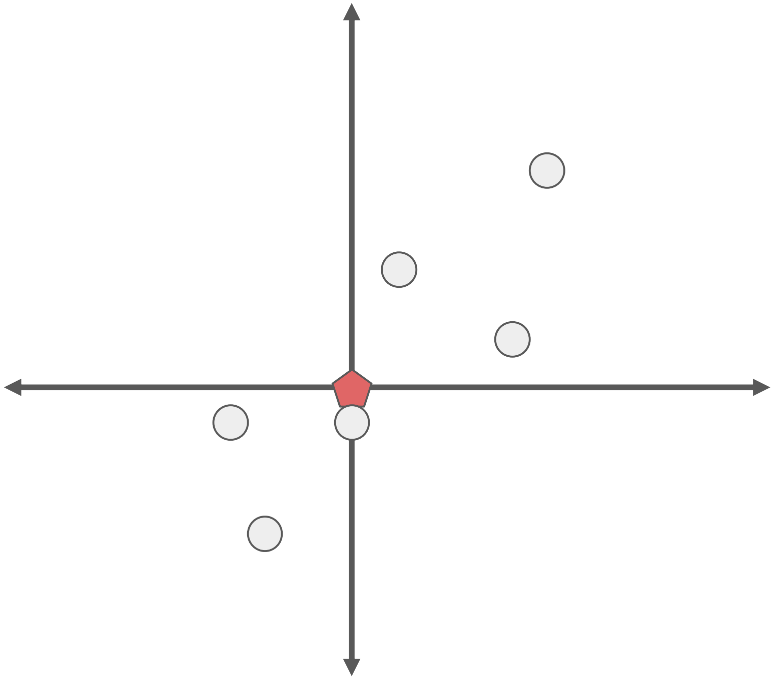
Principal Component Analysis

- Standardize the data:



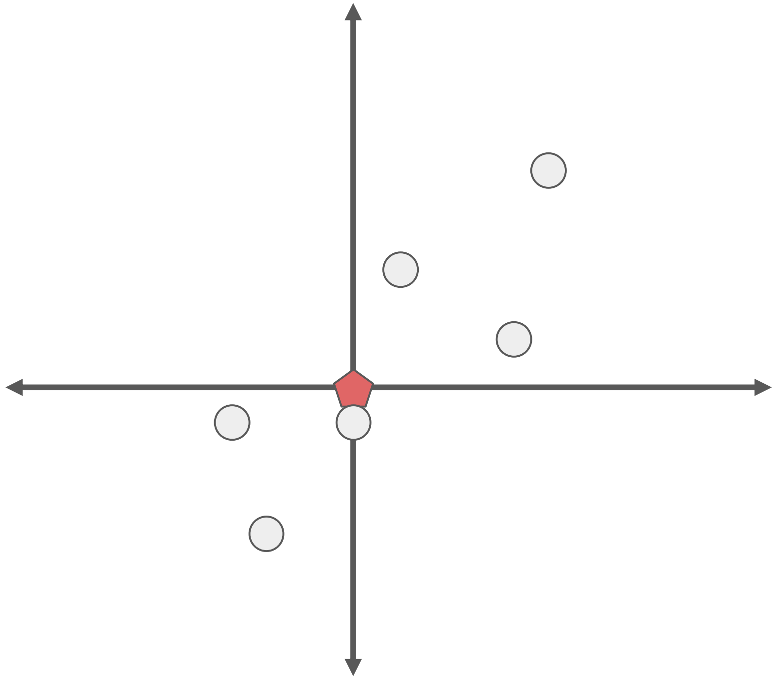
Principal Component Analysis

- Standardize the data:



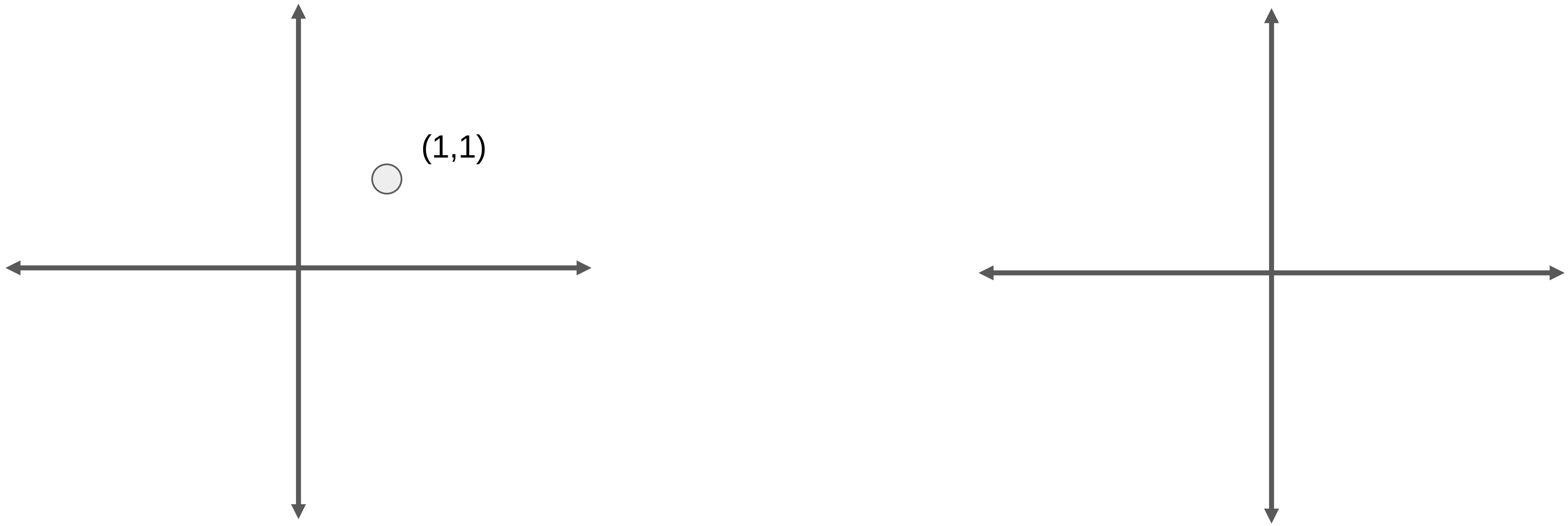
Principal Component Analysis

- Calculate covariance matrix for data:



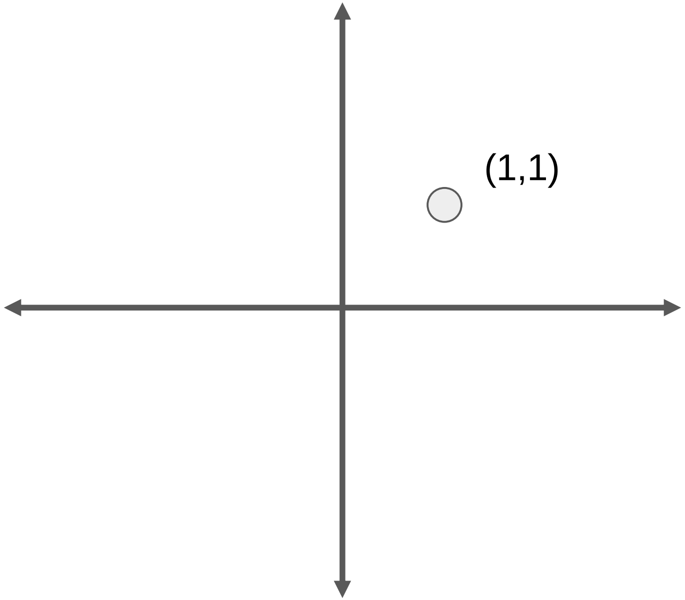
Principal Component Analysis

- Linear transformation of data:

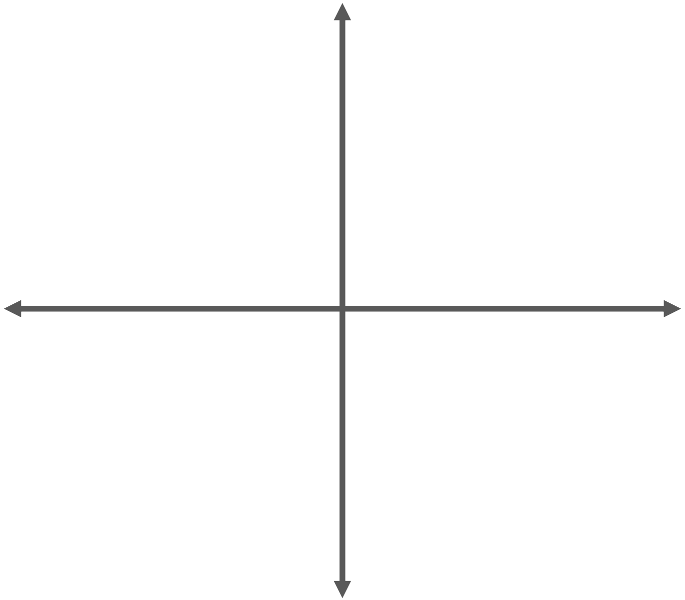


Principal Component Analysis

- Linear transformation of data:



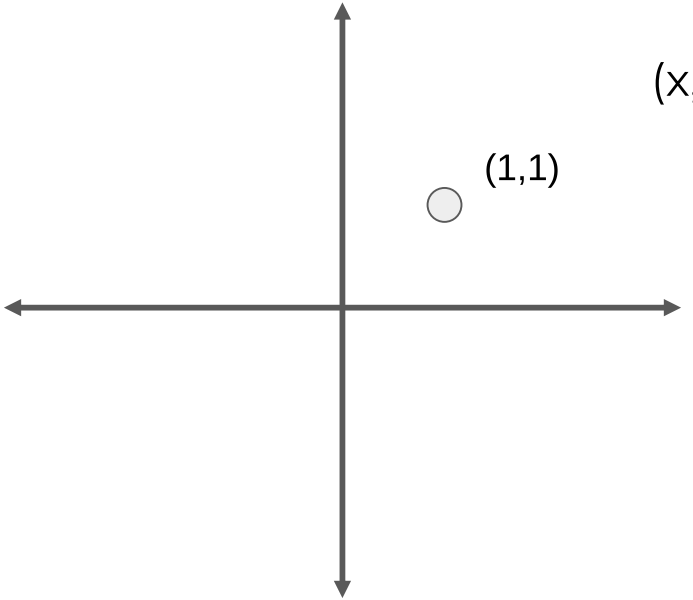
$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



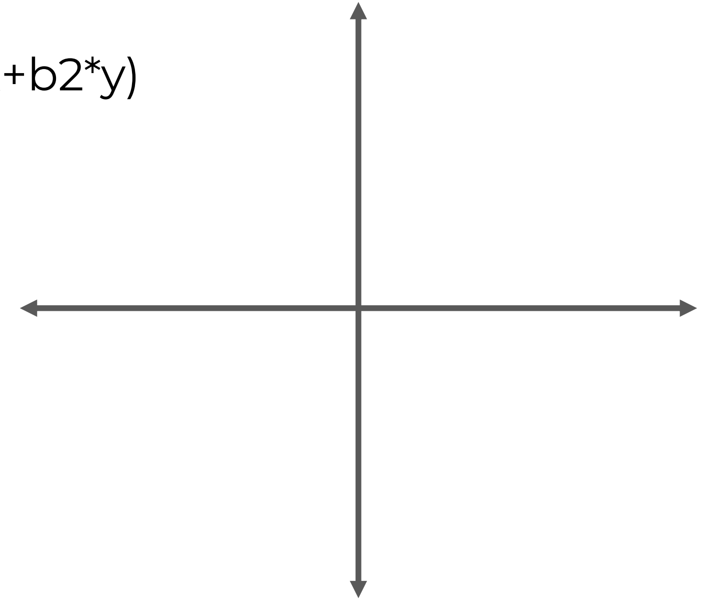
Principal Component Analysis

- Linear transformation of data:

$$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$$

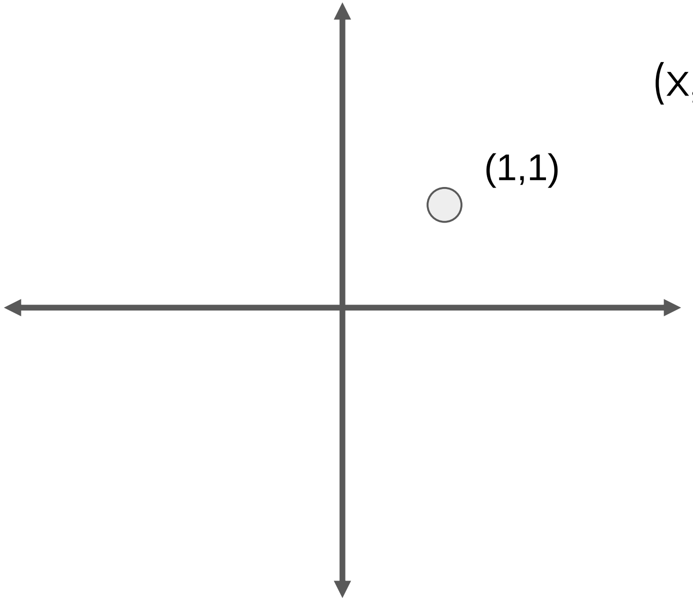


$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



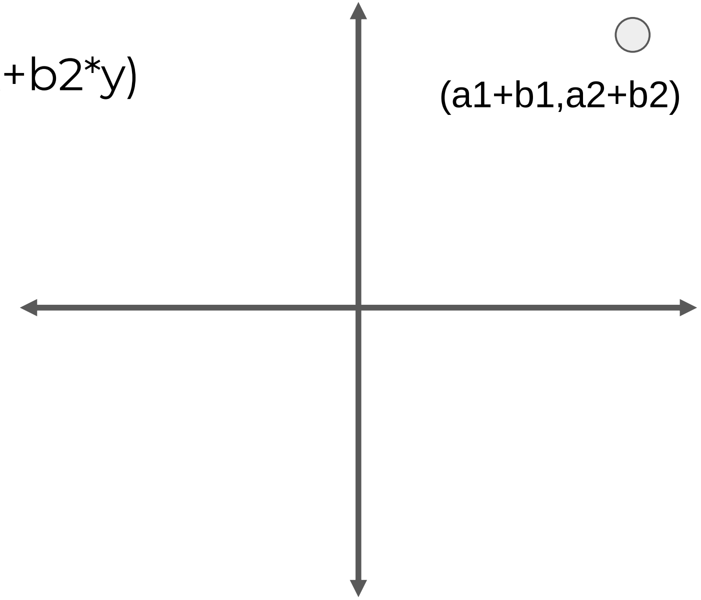
Principal Component Analysis

- Linear transformation of data:



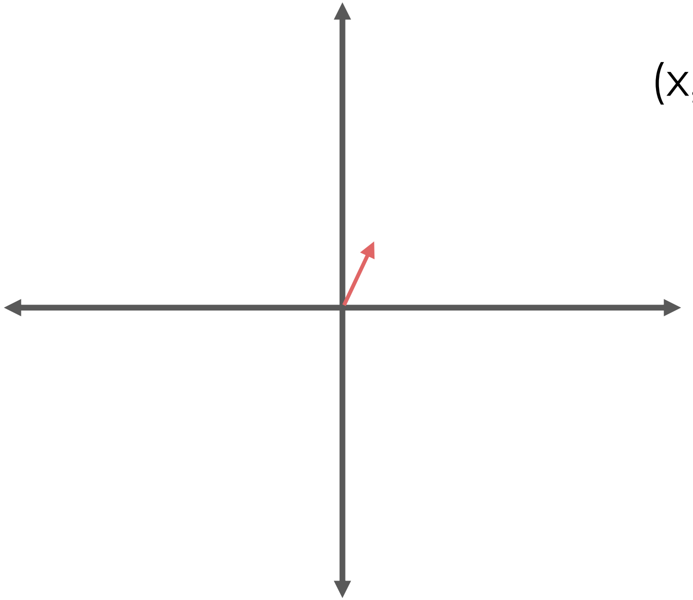
(x,y) → (a1*x+b1*y,a2*x+b2*y)

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



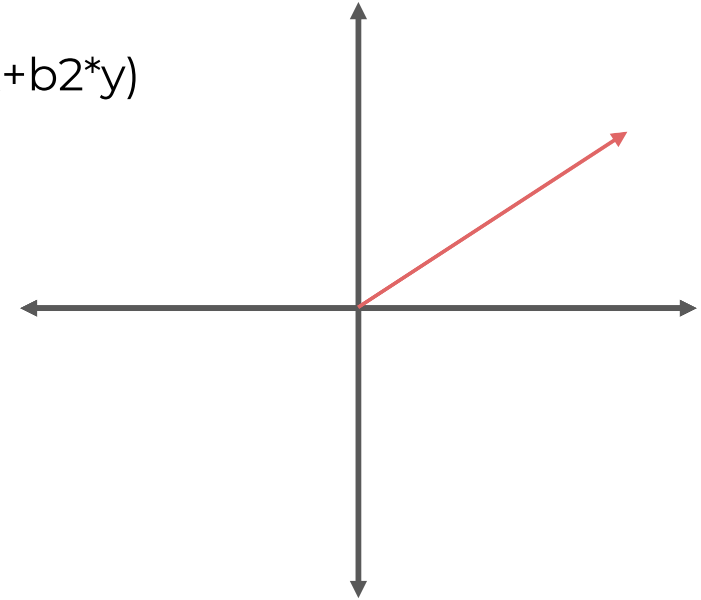
Principal Component Analysis

- Linear transformation of data:



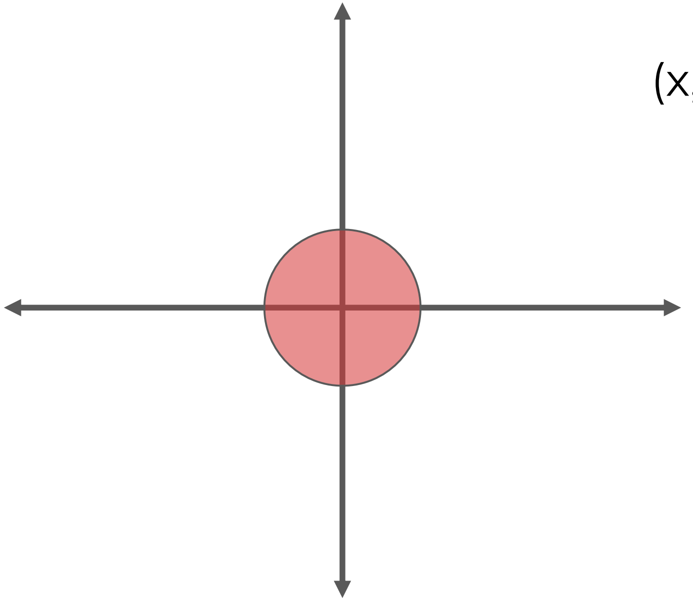
$$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$$

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



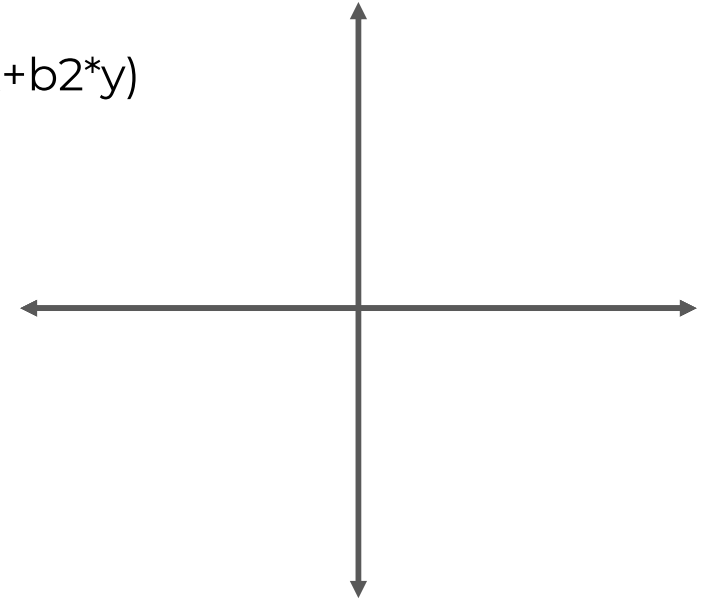
Principal Component Analysis

- Linear transformation of data:



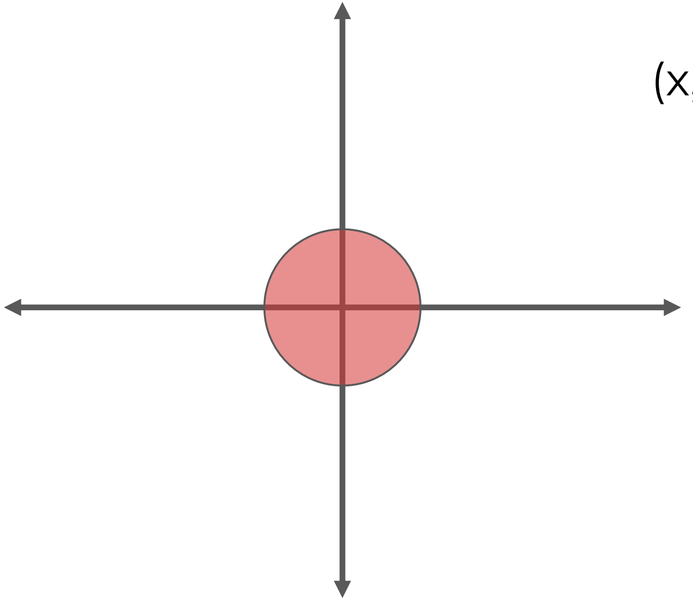
$$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$$

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



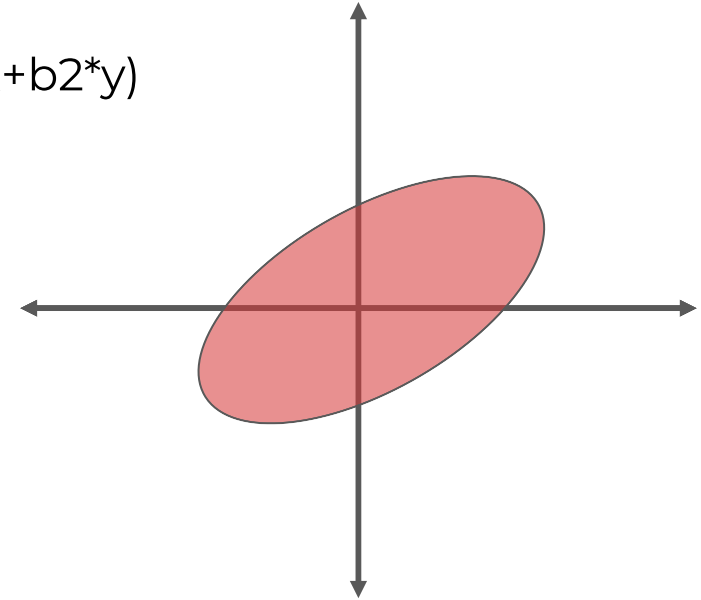
Principal Component Analysis

- Linear transformation of data:



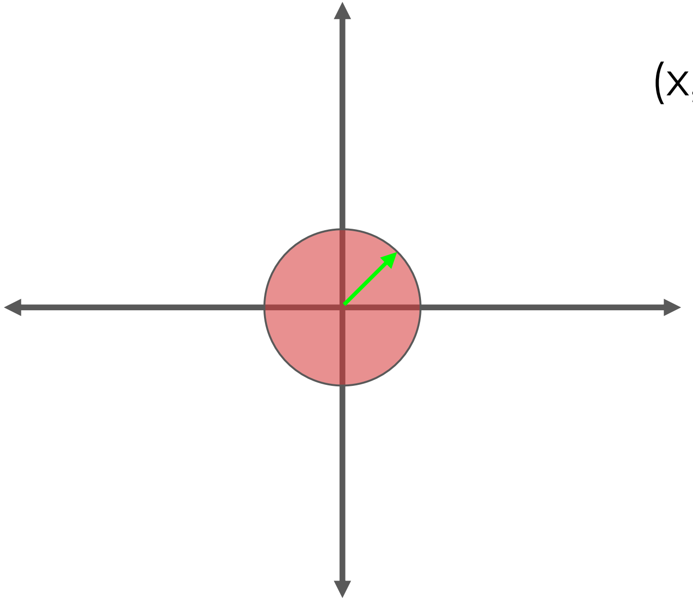
$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



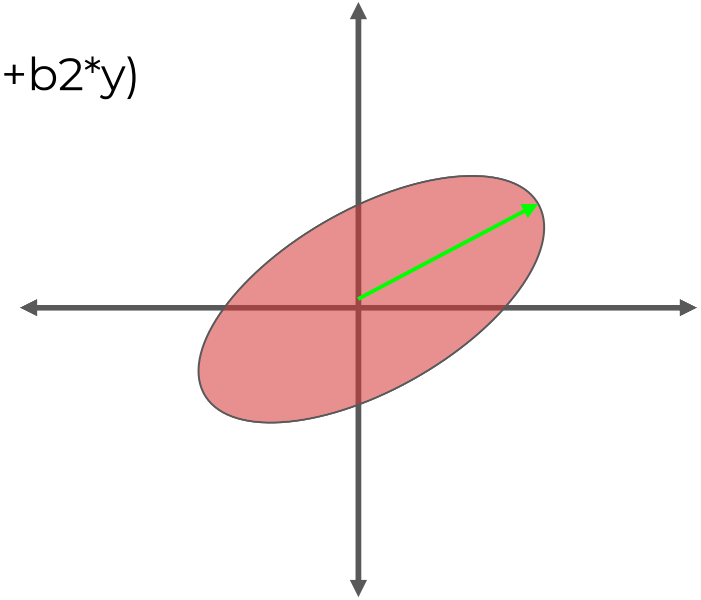
Principal Component Analysis

- Linear transformation of data:



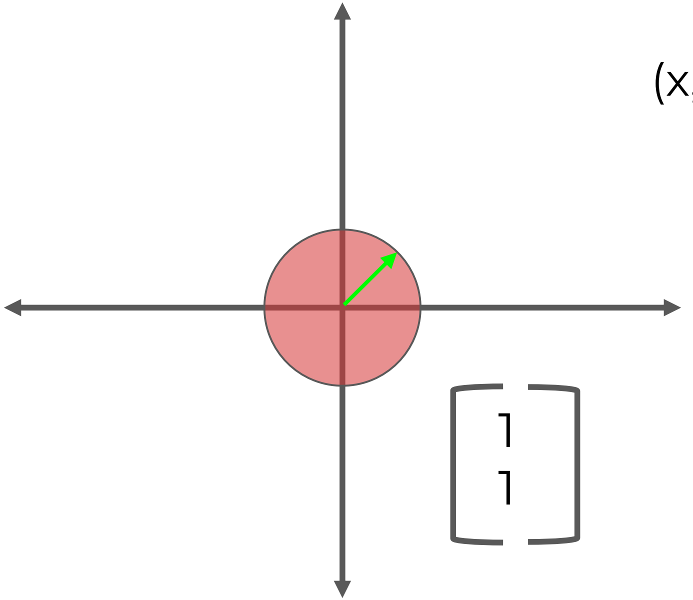
(x,y) → (a1*x+b1*y,a2*x+b2*y)

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$

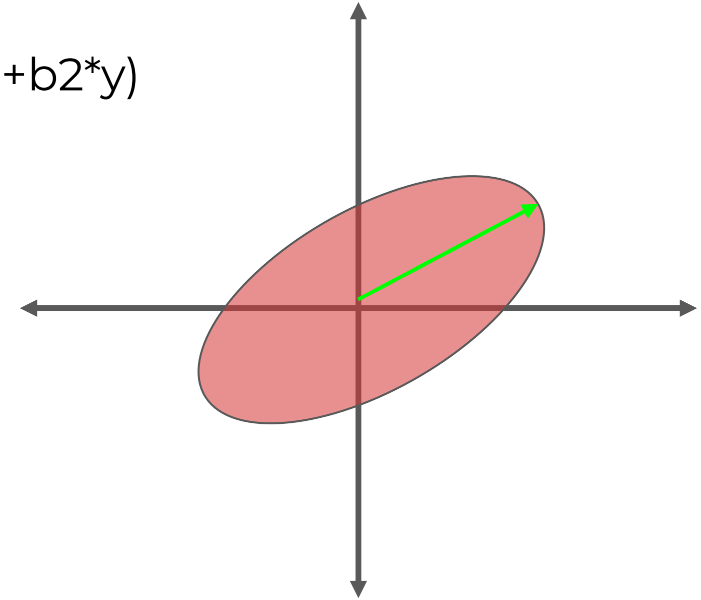


Principal Component Analysis

- EigenVector: Directional Information



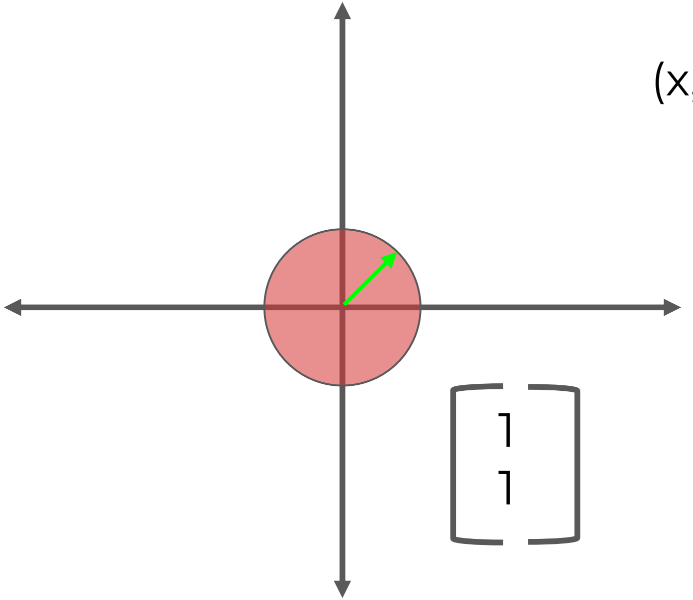
(x,y) → (a1*x+b1*y,a2*x+b2*y)

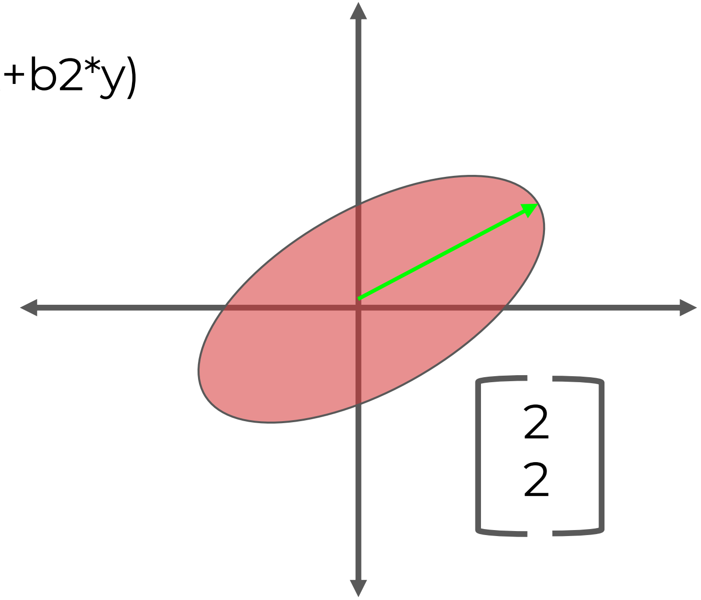


Principal Component Analysis

- EigenVector: Directional Information

$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$

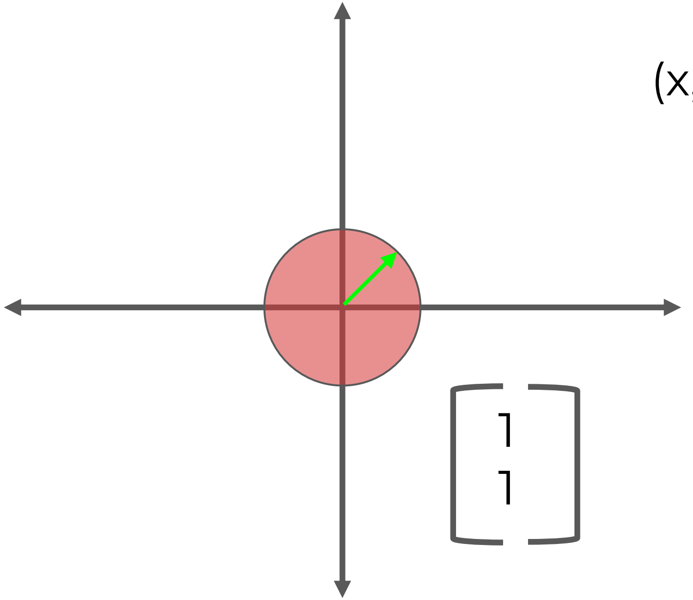




Principal Component Analysis

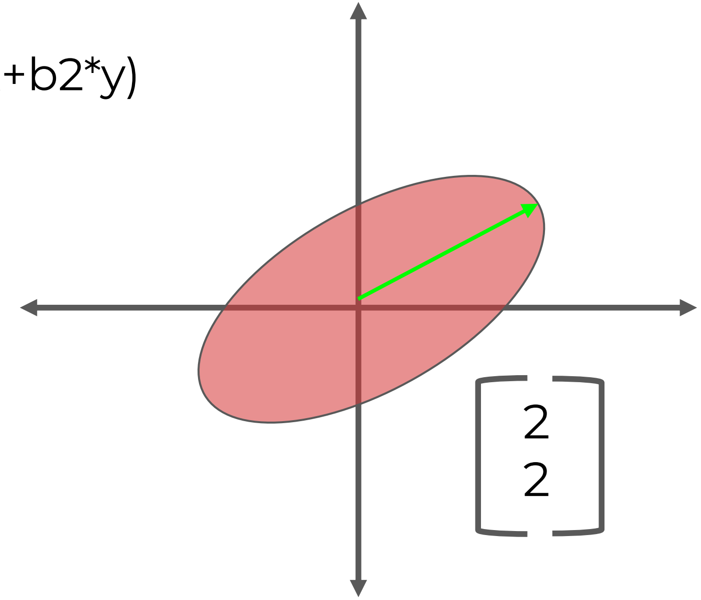
- EigenValue: Magnitude Information

$$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$$



$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$

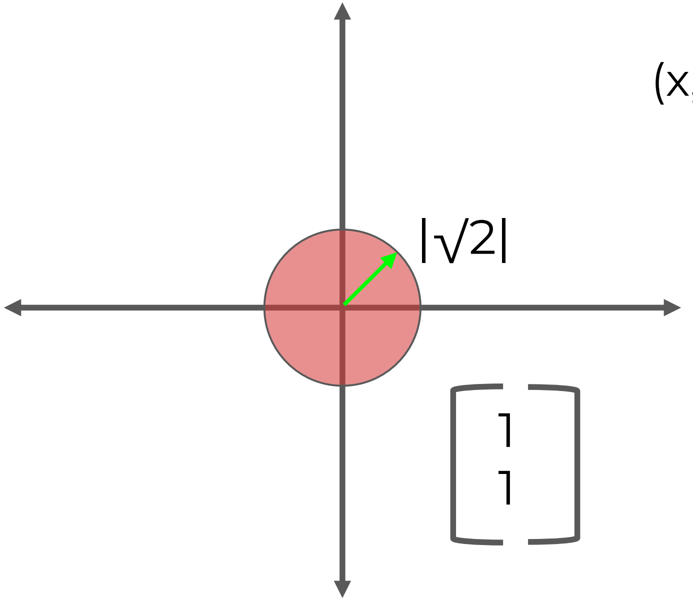


$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Principal Component Analysis

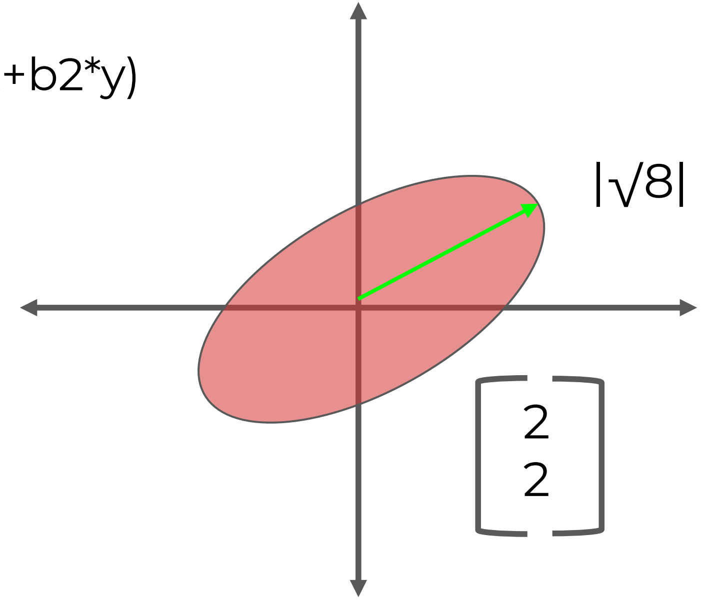
- EigenValue: Magnitude Information

$(x,y) \rightarrow (a1*x+b1*y, a2*x+b2*y)$



$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

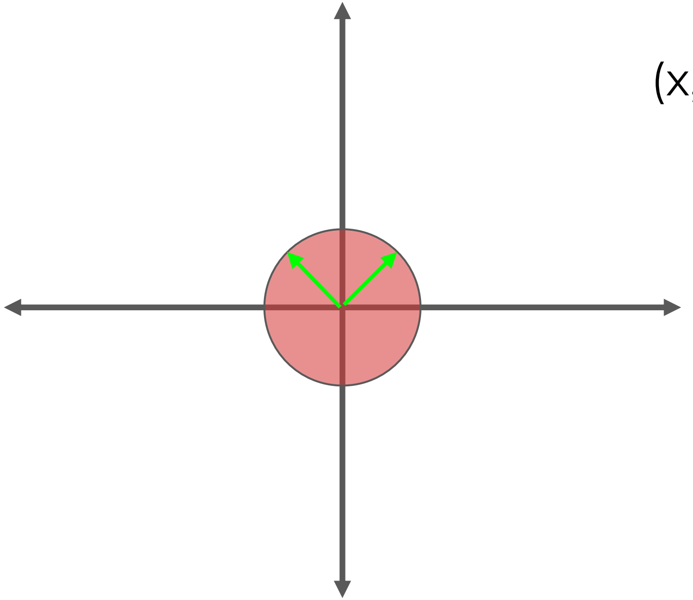
$$\begin{bmatrix} a1 & b1 \\ a2 & b2 \end{bmatrix}$$



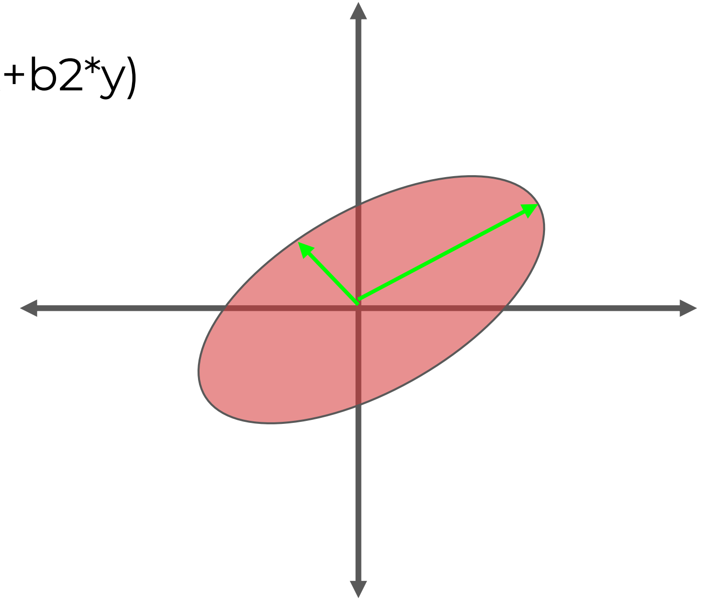
$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Principal Component Analysis

- Orthogonal EigenVector

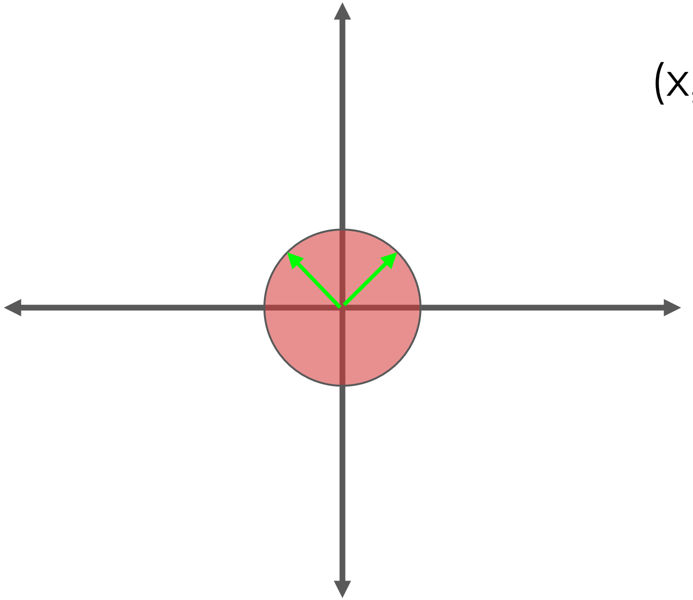


(x,y) → (a1*x+b1*y,a2*x+b2*y)

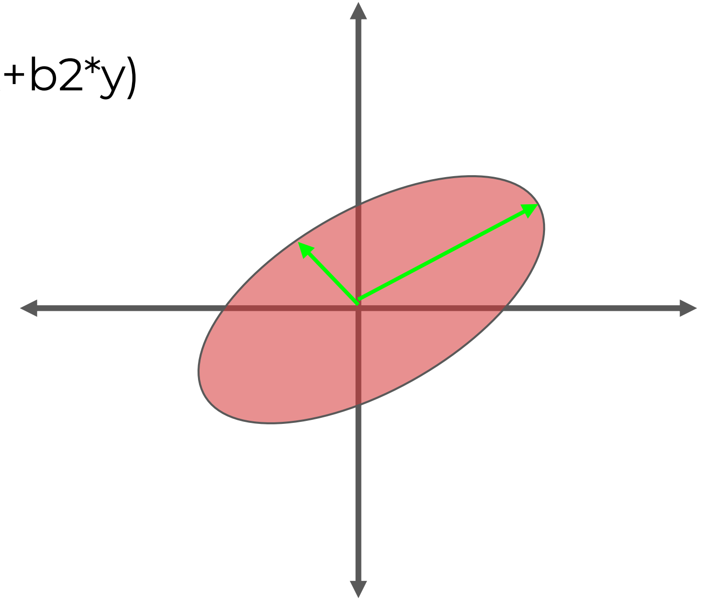


Principal Component Analysis

- EigenVector is just a linear transformation



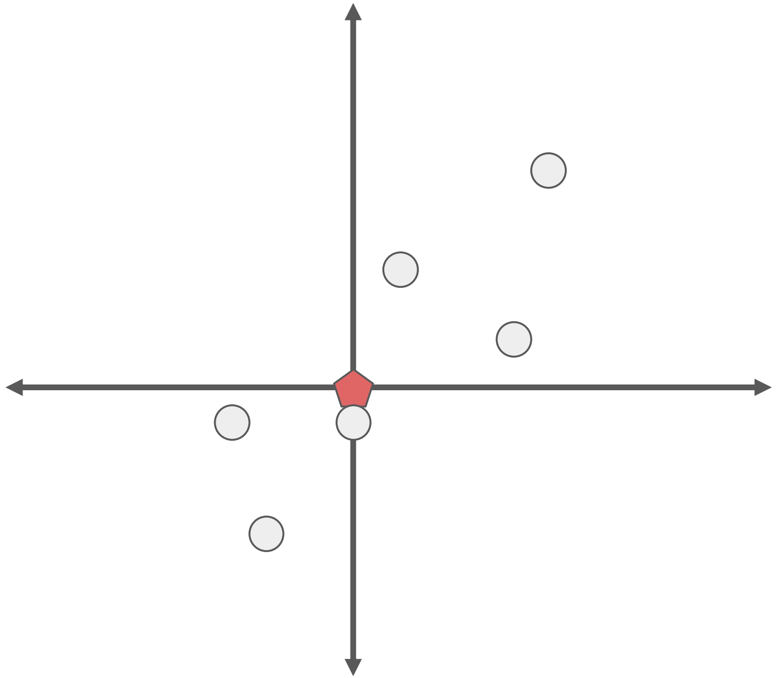
(x,y) → (a1*x+b1*y,a2*x+b2*y)



Principal Component Analysis

- Apply Linear Transformation:

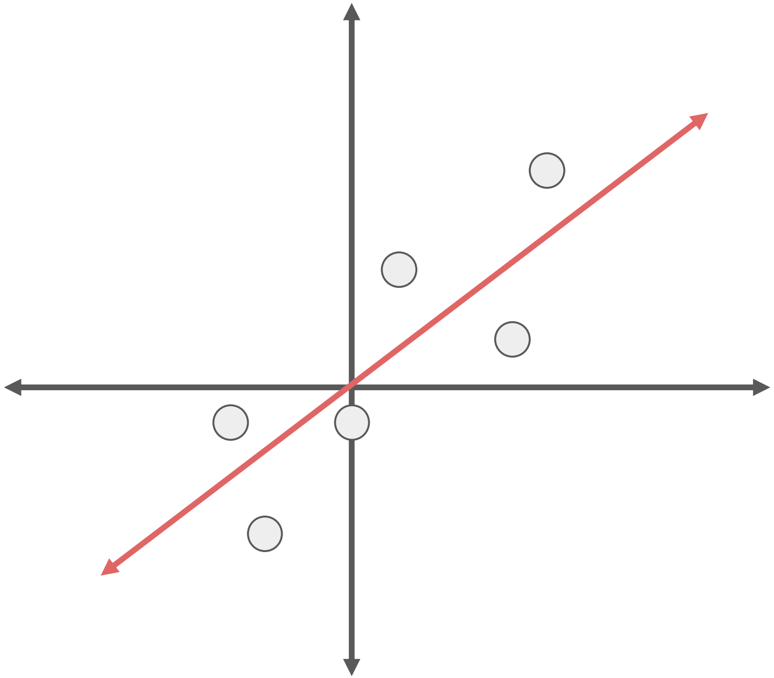
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- Apply Linear Transformation:

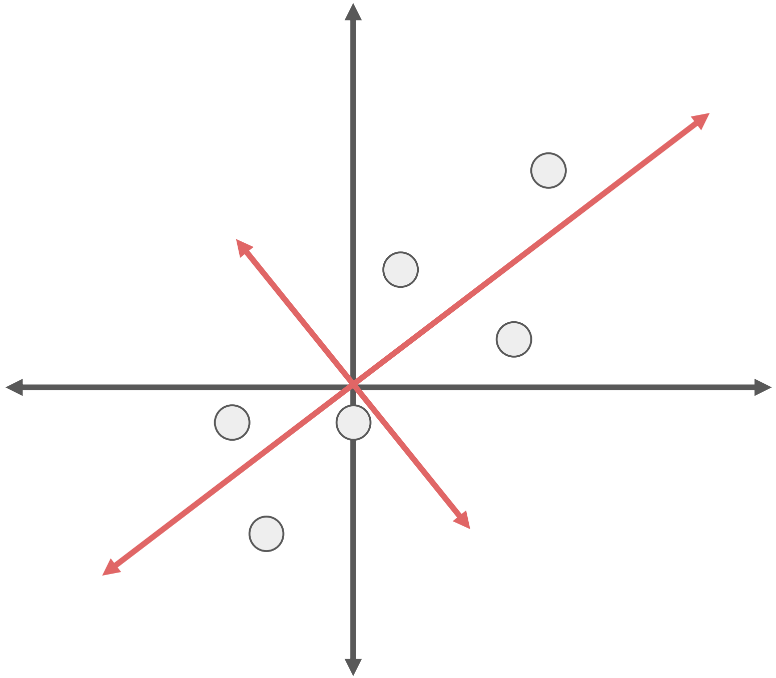
$$\begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}$$



Principal Component Analysis

- Apply Linear Transformation:

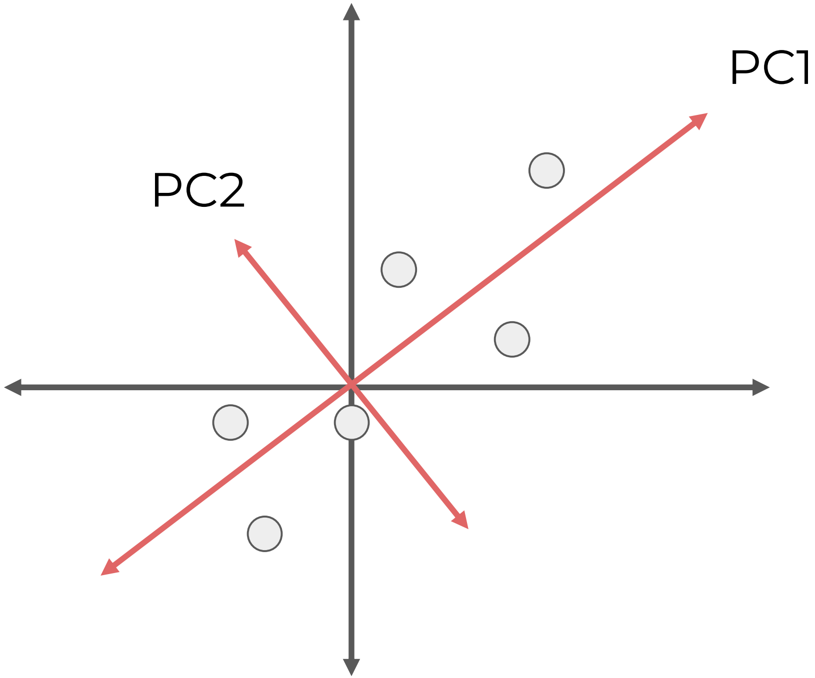
$$\begin{bmatrix} \text{Var}(X) & \text{Cov}(X,Y) \\ \text{Cov}(X,Y) & \text{Var}(Y) \end{bmatrix}$$



Principal Component Analysis

- EigenValue measures variance explained:

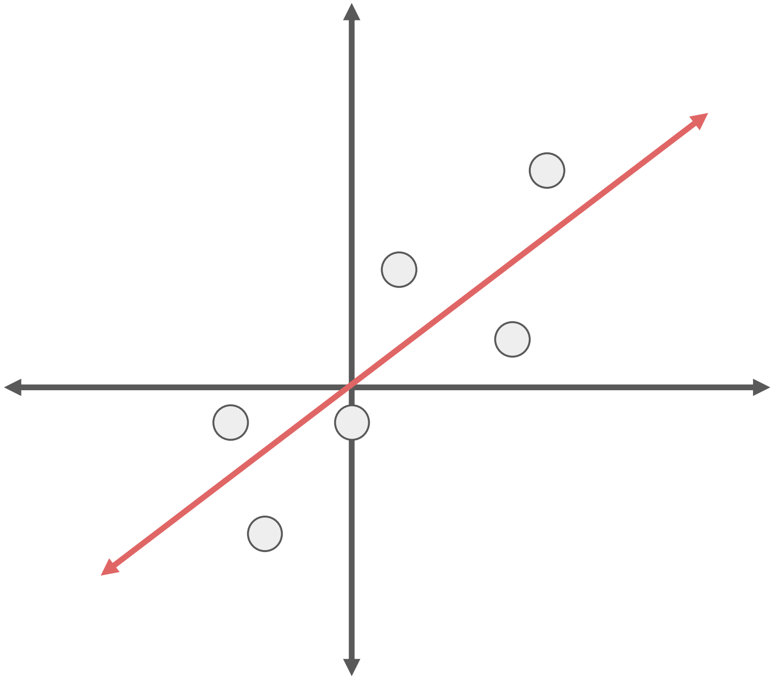
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- EigenValue measures variance explained:

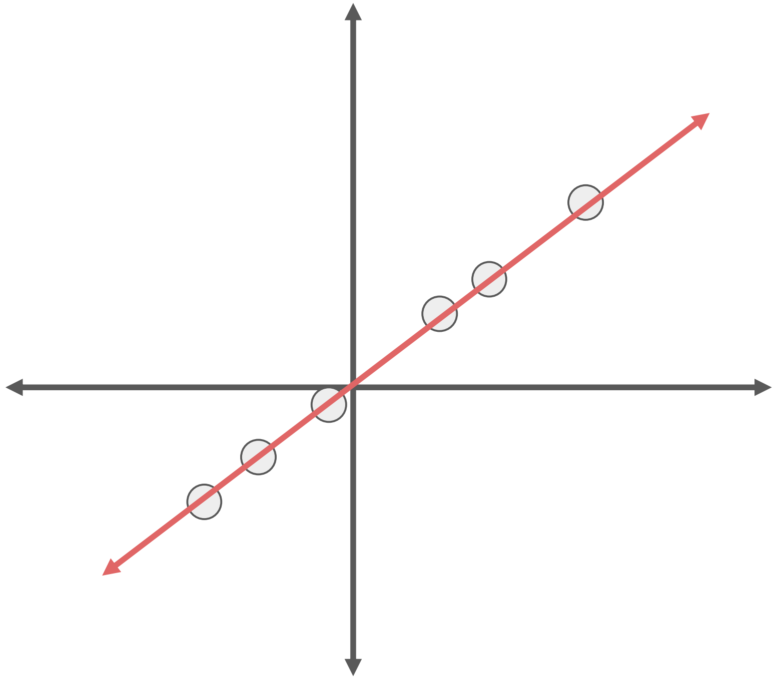
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- EigenValue measures variance explained:

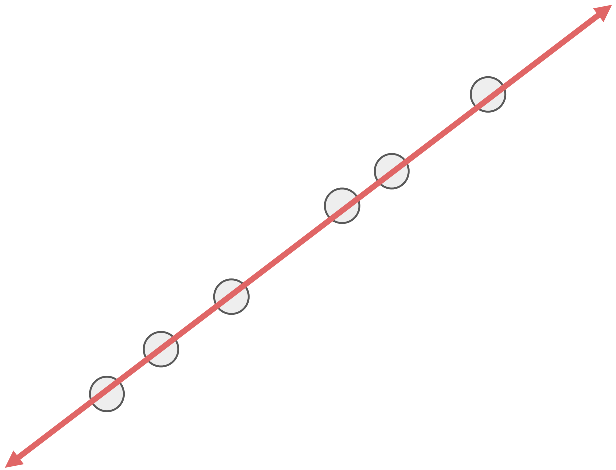
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- EigenValue measures variance explained:

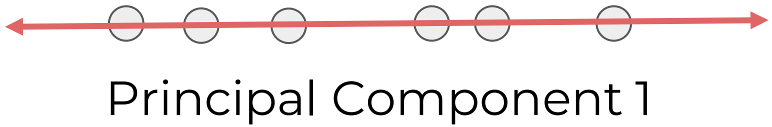
$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- EigenValue measures variance explained:

$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$



Principal Component Analysis

- PCA Steps
  - Get original data
  - Calculate Covariance Matrix
  - Calculate EigenVectors
  - Sort EigenVectors by EigenValues
  - Choose N largest EigenValues
  - Project original data onto EigenVectors

Principal Component Analysis

Manual Implementation

Principal Component Analysis

Scikit-Learn Implementation

Principal Component Analysis

Project Exercise Solutions

PIERIAN DATA