THE UNIVERSITY OF WINNIPEG | Professional, Applied and Continuing Education

INTRODUCTION TO MACHINE LEARNING

DIT 45100

Dimensionality Reduction

What is Dimensionality Reduction?

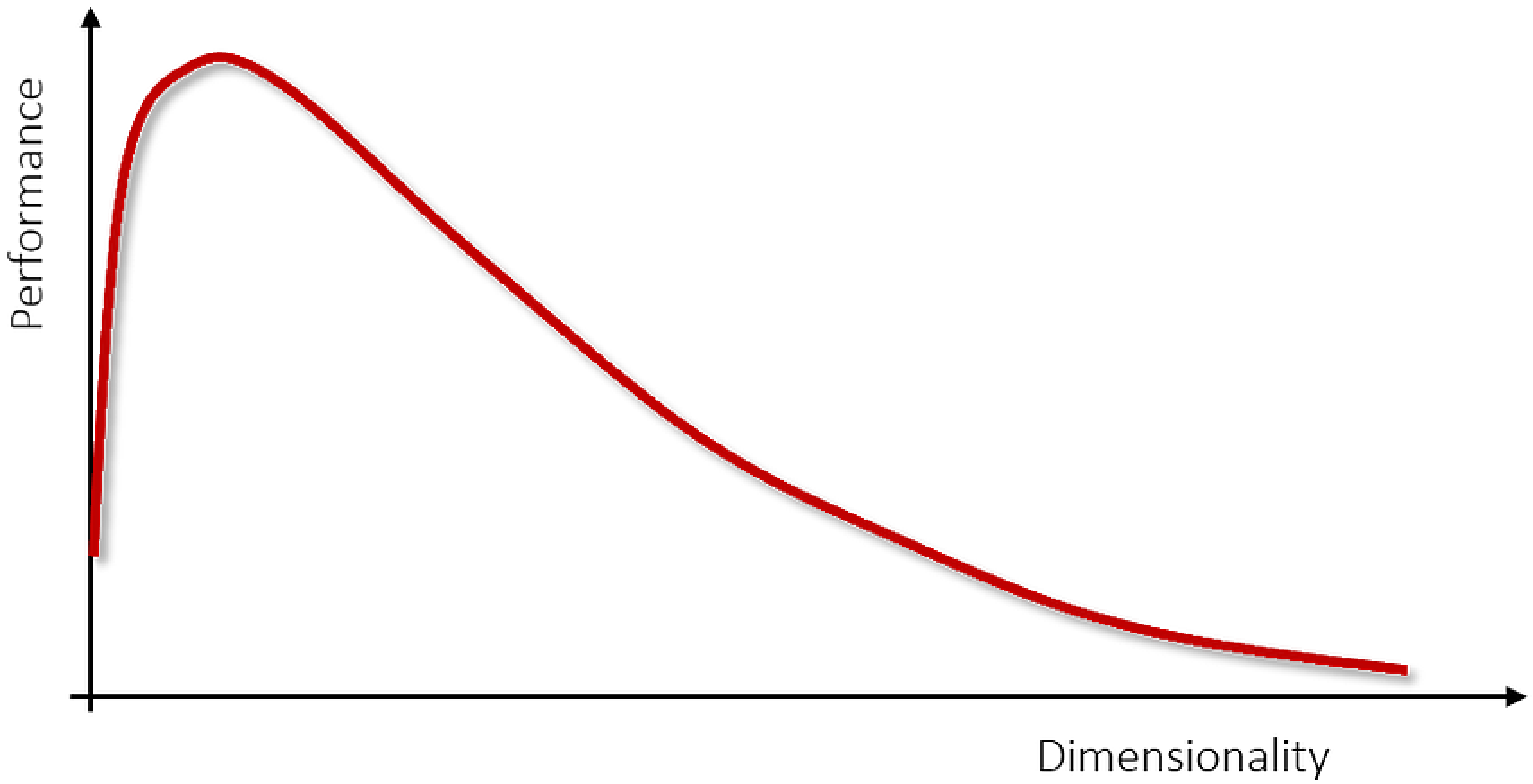
- Transformation of data from a high-dimensional space into a low-dimensional space such that the low-dimensional representation retains meaningful properties of the original data by
  - Maximizing variance (information)
  - Minimizing information loss

Why to Reduce Data Dimensions?

- Curse of dimensionality
- Data visualization and analysis
- Model performance

THE UNIVERSITY OF WINNIPEG | Professional, Applied and Continuing Education

Model Performance

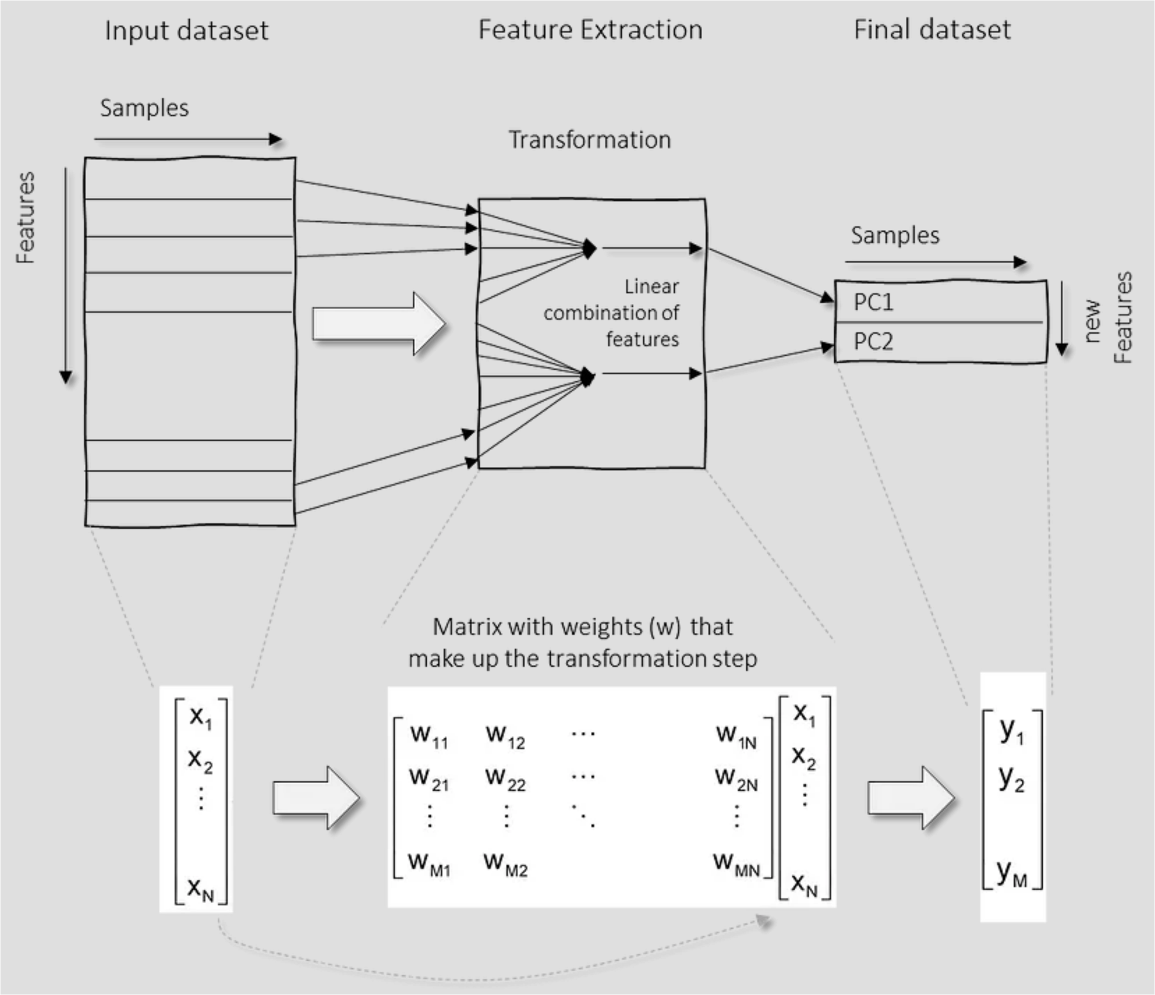


Dimensionality Reduction

How do we reduce data dimensions?

- Feature selection
  - Find a subset of the input variables of features
  - Works well for uncorrelated (or weekly-correlated) features
- Feature extraction
  - Features are (highly) correlated
  - Are on the same scale
  - No outliers

Feature Extraction



Big Idea

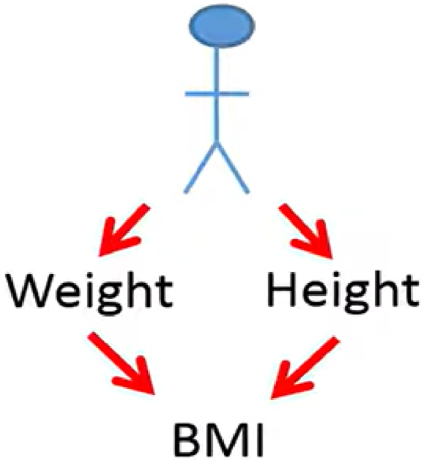
- The big idea is to find new coordinates that:
  - capture the maximum variance (information) in the data, and
  - minimize the information loss

- Principal Component Analysis (PCA) is a statistical technique which does that

Principal Component Analysis (PCA)

- Principal component analysis (PCA) is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data preprocessing.
- The data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified.
- Principal component analysis has applications in many fields such as population genetics, microbiome studies, atmospheric sciences, and more.

Combining variables



Cholesterol = Weight + Height

$\boxed{\text{Cholesterol} = \text{BMI}}$

$$BMI = \frac{Weight_{kg}}{Height_{m}^{2}}$$

Housing Data

Size
Number of rooms
Number of bathrooms

Schools around
Crime rate

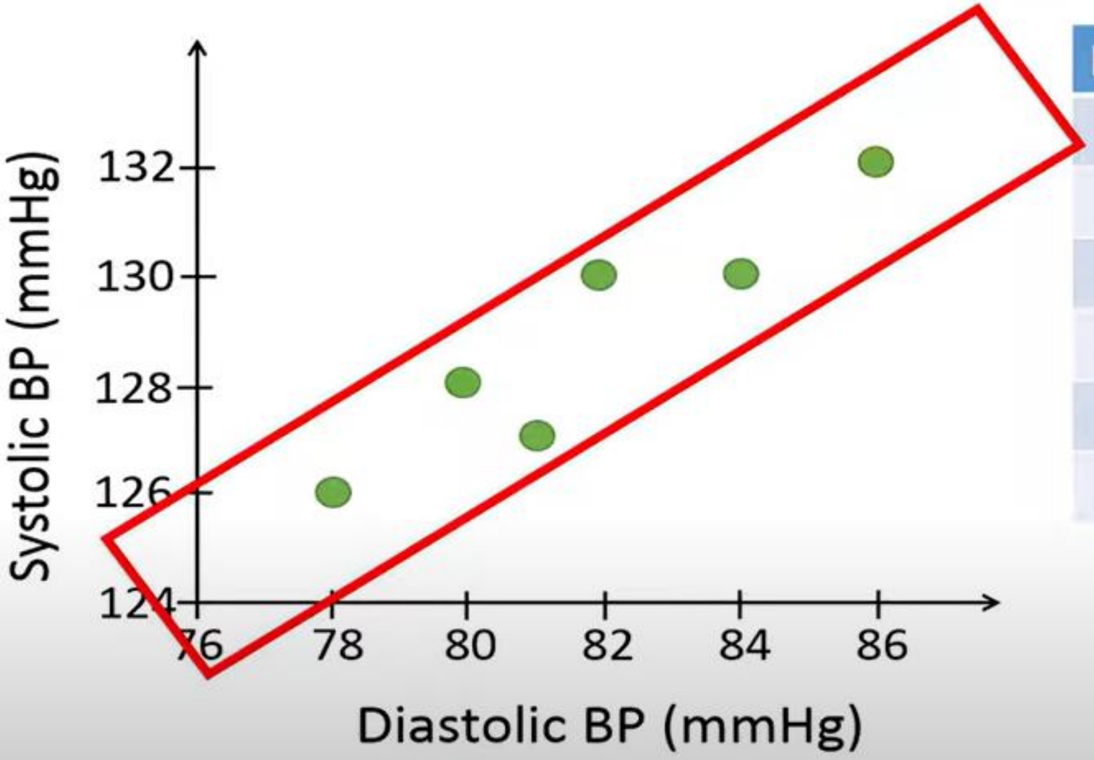
Housing Data

Size
Number of rooms
Number of bathrooms
→ Size feature

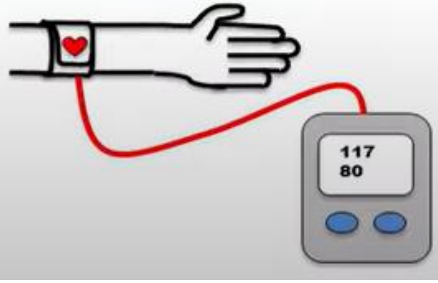
Schools around
Crime rate
→ Location feature

THE UNIVERSITY OF WINNIPEG | Professional, Applied and Continuing Education

Combining variables



Diastolic BP	Systolic BP
78	126
80	128
81	127
82	130
84	130
86	132

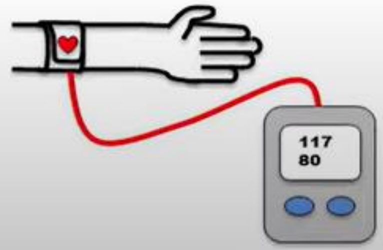


Combining variables

Diastolic BP	Systolic BP
78	126
80	128
81	127
82	130
84	130
86	132

$$Y = \alpha_1 X_1 + \alpha_2 X_2$$

$$BP = \boxed{\alpha_1} DBP + \boxed{\alpha_2} SBP$$



PCA

Principal component analysis (PCA) is a method to find the linear combination that accounts for as much variability as possible.

$$BP = \boxed{\alpha_1} DBP + \boxed{\alpha_2} SBP$$

PCA

Diastolic BP	Systolic BP
78	126
80	128
81	127
82	130
84	130
86	132

	DBP	SBP
DBP	8.17	5.97
SBP	5.97	4.97

$$\boxed{Eig = \begin{bmatrix} -0.8 \\ -0.6 \end{bmatrix}}$$

$$BP = -0.8DBP + (-0.6SBP)$$

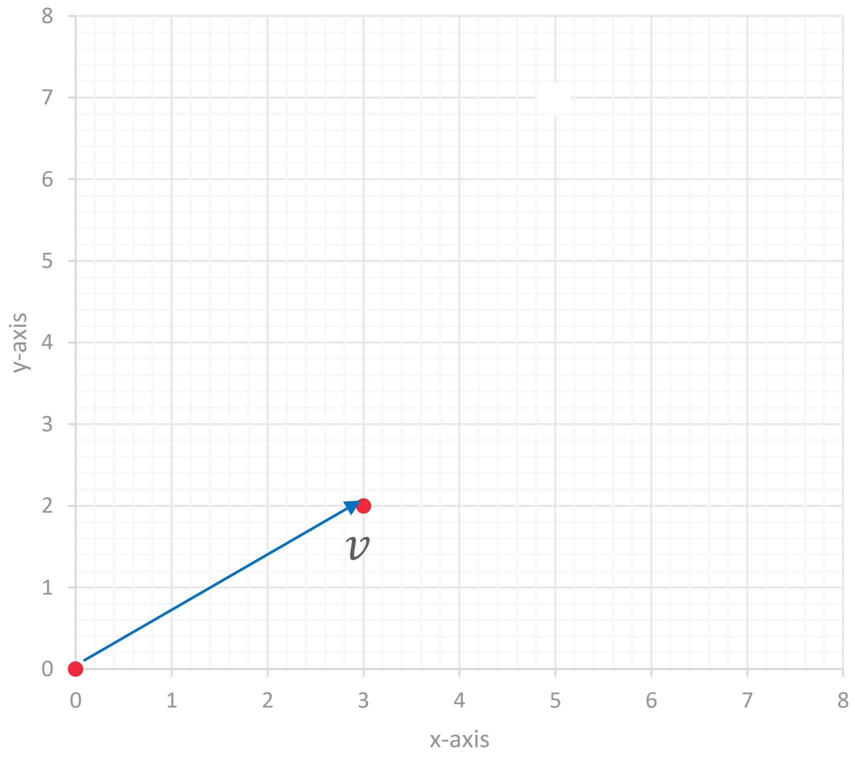
THE UNIVERSITY OF WINNIPEG | Professional, Applied and Continuing Education

Fundamentals

Vectors and Matrices

$$v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

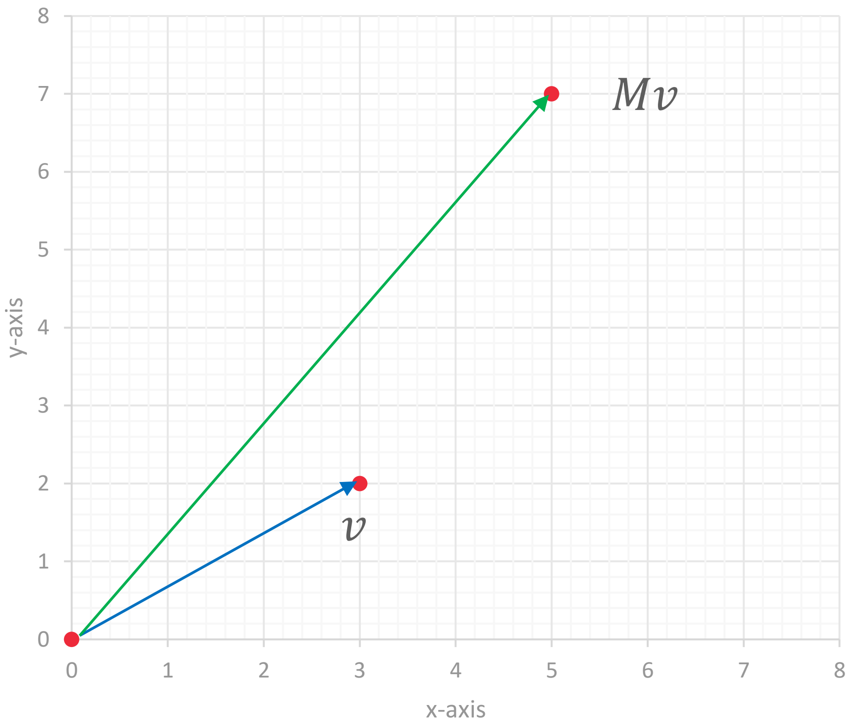


Vectors and Matrices

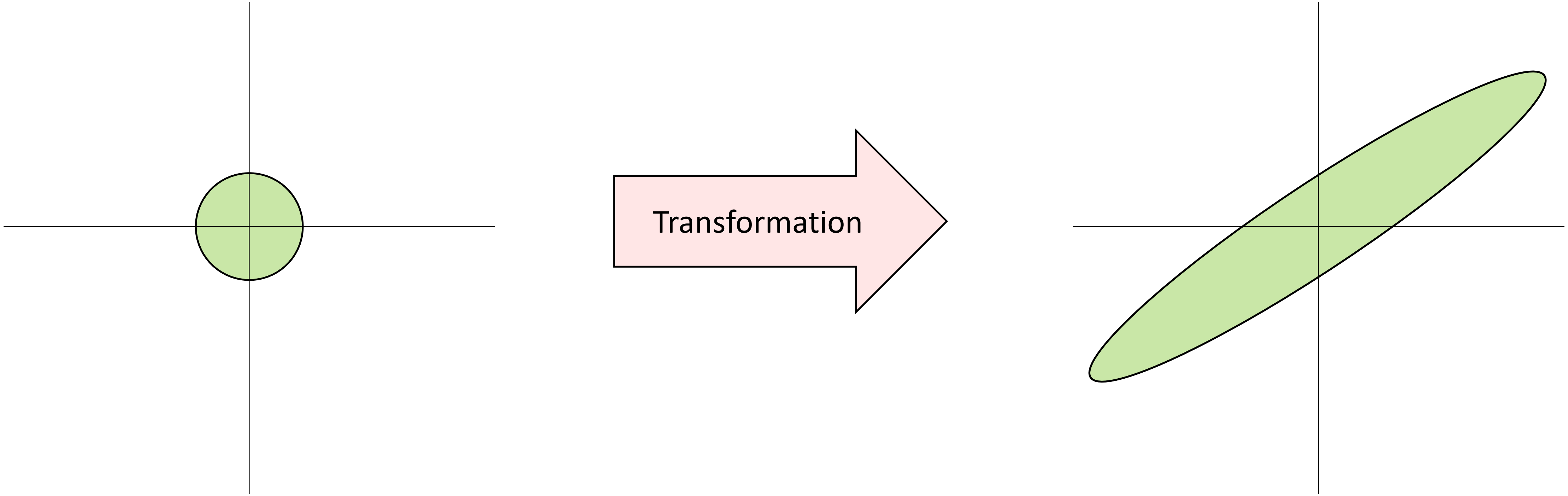
$$v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

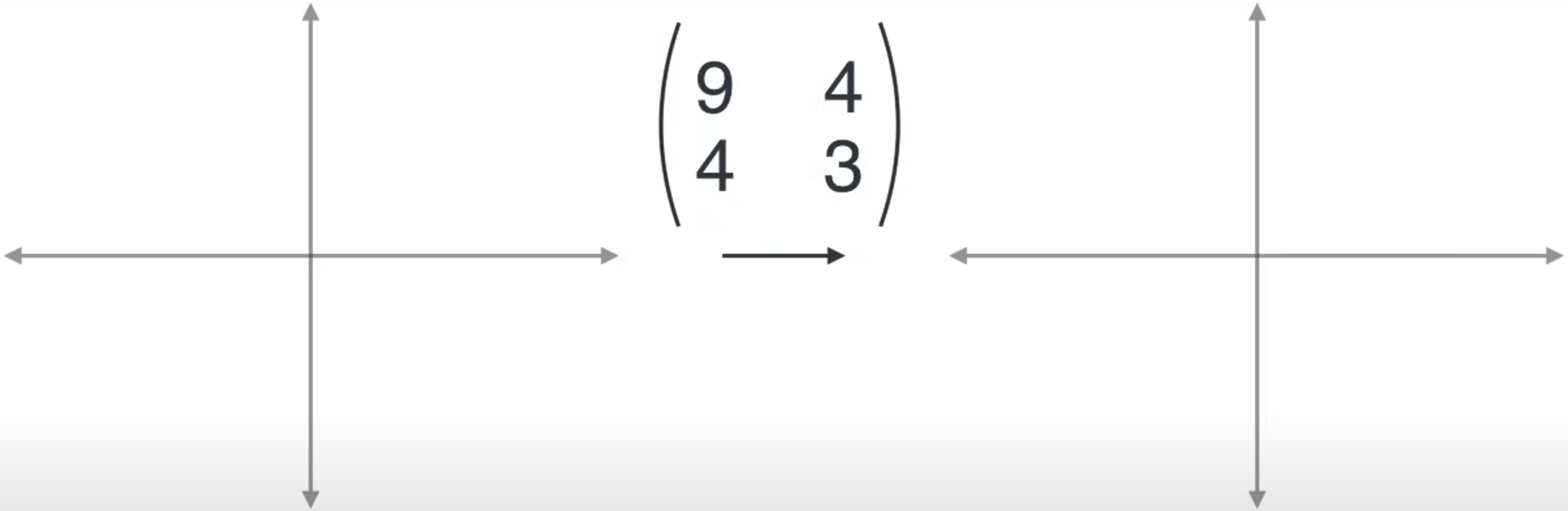
$$Mv = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$



Linear Transformation



Linear Transformations



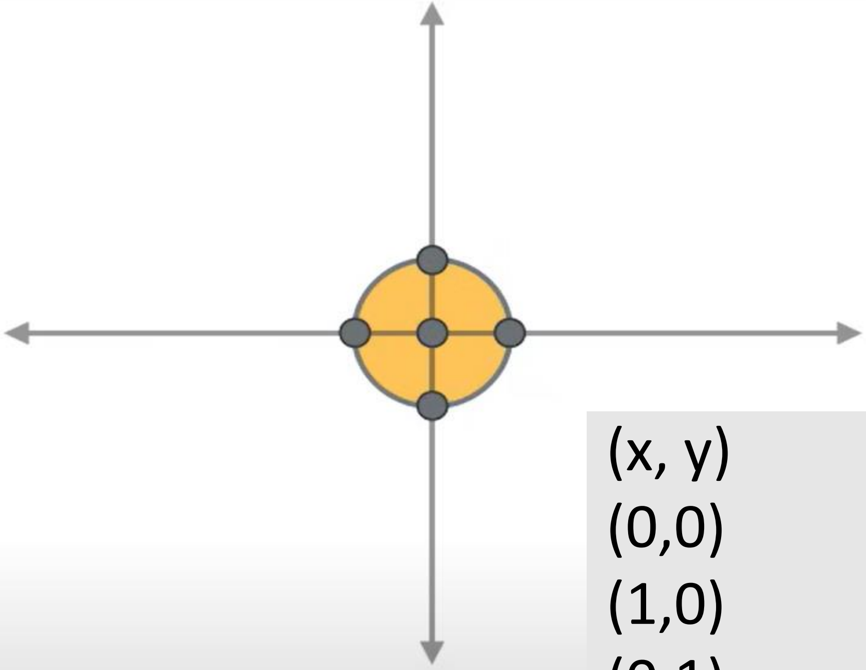
Linear Transformations

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

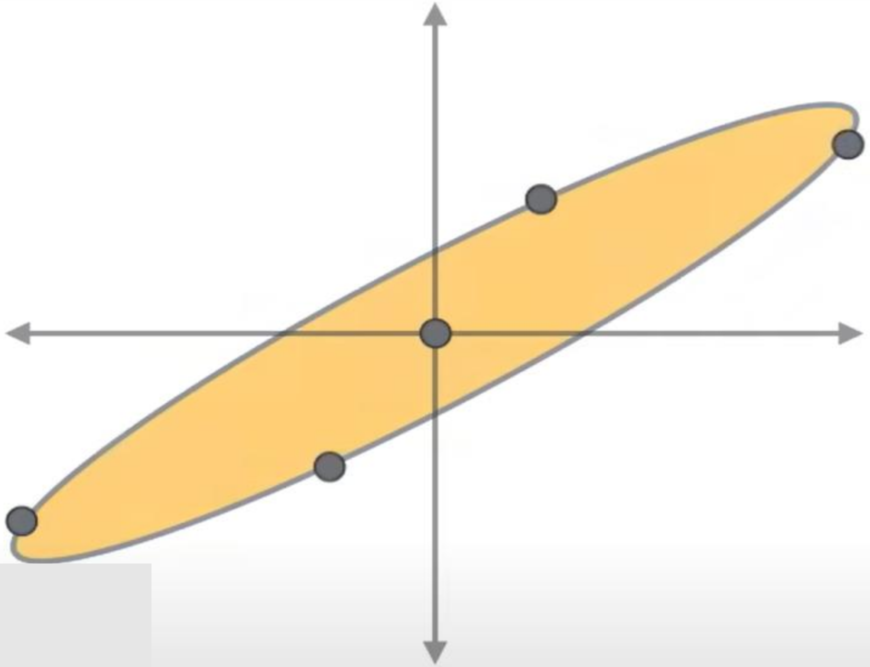
$$(x, y) \longrightarrow (9x+4y,\ 4x+3y)$$

$$(0,0) \qquad\qquad (0,0)$$

Linear Transformations

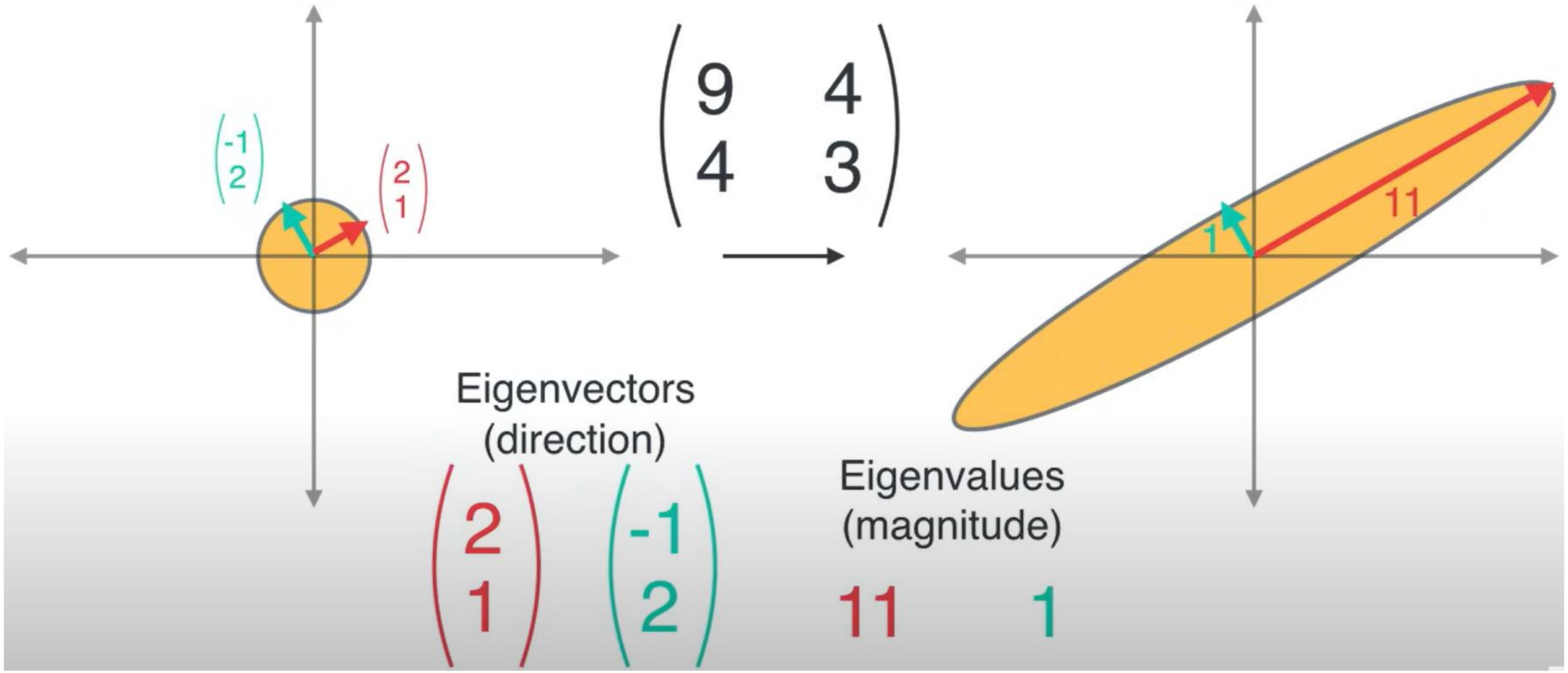


$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} \longrightarrow$$



(x, y)	(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



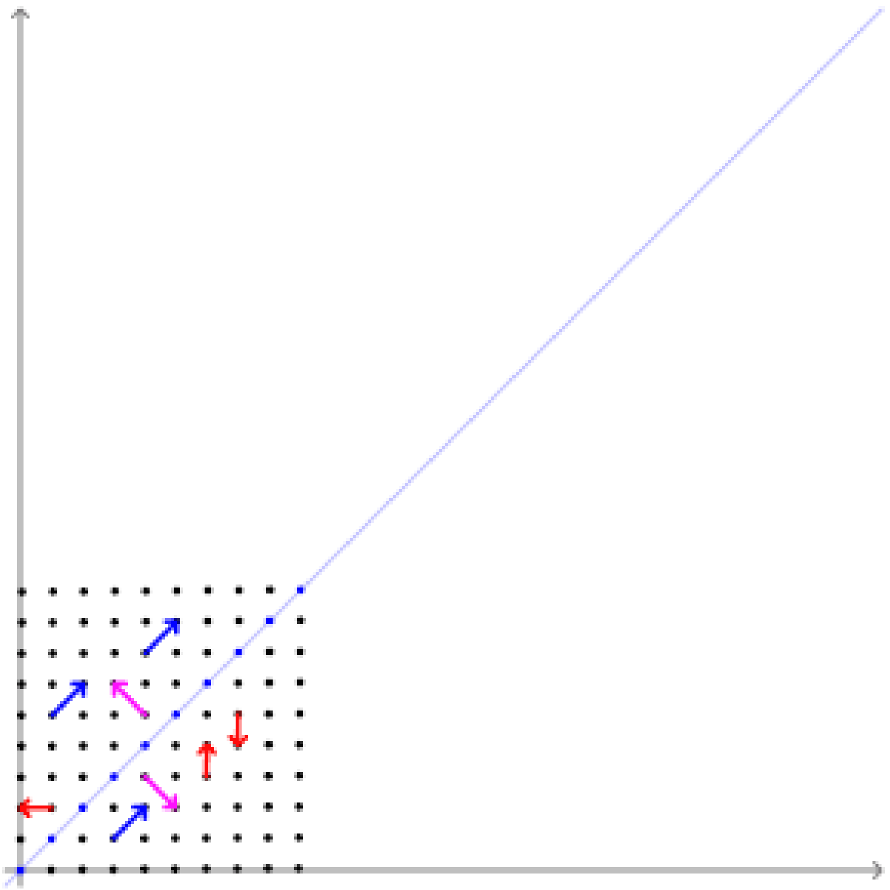
Eigenvalues & Eigenvectors

$$Av = \lambda v$$

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

$$|A - \lambda \mathrm{I}| = 0$$

Eigenvalues & Eigenvectors



https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors

Eigenvalues

$$A = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 9-\lambda & 4 \\ 4 & 3-\lambda \end{bmatrix}$$

$(9 - \lambda)(3 - \lambda) - (4)(4) = \lambda^2 - 12\lambda + 11 = 0$ ← characteristic polynomial

Solving this quadratic, we get $\lambda = 11 \text{ and } \lambda = 1$

Eigenvectors

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

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

Class Activity

$$A = \begin{bmatrix} 17 & -6 \\ 45 & -16 \end{bmatrix}$$

Find the eigenvalues and eigenvectors of the matrix A

Solution

$$A = \begin{bmatrix} 17 & -6 \\ 45 & -16 \end{bmatrix} \rightarrow \quad |A - \lambda I| = \begin{vmatrix} 17 - \lambda & -6 \\ 45 & -16 - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - \lambda - 2 = 0 \rightarrow \boxed{\lambda = 2, -1}$$

$$@\ \lambda = 2 \rightarrow \quad \begin{matrix} 15 & -6 \\ 45 & -18 \end{matrix} \qquad v1 = \begin{matrix} 2 \\ 5 \end{matrix}$$

$$@\ \lambda = -1 \rightarrow \quad \begin{matrix} 18 & -6 \\ 45 & -15 \end{matrix} \qquad v2 = \begin{matrix} 1 \\ 3 \end{matrix}$$

Computing Eigenvalues / Eigenvectors

- Eigenvalue Decomposition (EVD)
  - Numpy.linalg.eig(data)
- Singular Value Decomposition (SVD)
  - Numpy.linalg.svd(data)

Eigenvalue Decomposition (EVD)

- Eigenvalue decomposition (EVD) of a symmetric matrix A is defined as:

$$AQ = Q\Lambda$$

$$A = Q\Lambda Q^{-1}$$

$$A = Q\Lambda Q^{T}$$

EVD

$$A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \rightarrow \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$Q = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Singular Value Decomposition

- Assuming that we are calculating over the field of real numbers, the singular value decomposition (SVD reduced) of an m x n matrix X is defined as:

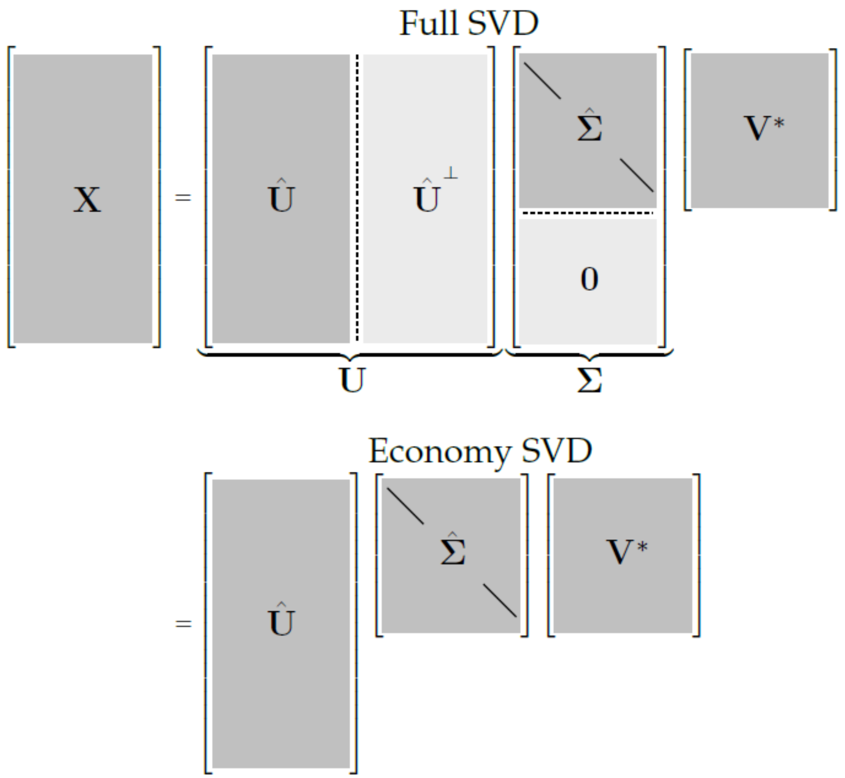
$$X = U\Sigma V^T$$

orthonormal matrix of left singular vectors ($XX^T = U\Sigma^2U^T$)

diagonal matrix of singular values

orthonormal matrix of right singular vectors ($X^TX = V\Sigma^2V^T$)

U and V are unitary matrices
A square matrix U is unitary if $U^TU = UU^T = I$.

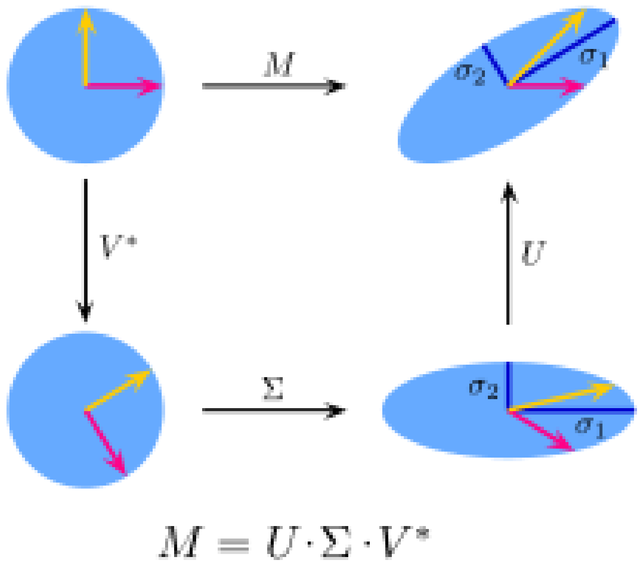


* denotes the complex conjugate transpose, for real-valued matrices, this is the same as the regular transpose $V^* = V^T$

Singular Value Decomposition

Figure: Illustration of the singular value decomposition **UΣV*** of a real 2×2 matrix **M**.

- **Top:** The action of **M**, indicated by its effect on the unit disc D and the two canonical unit vectors e_1 and e_2.
- **Left:** The action of **V***, a rotation, on D, e_1, and e_2.
- **Bottom:** The action of **Σ**, a scaling by the singular values σ_1 horizontally and σ_2 vertically.
- **Right:** The action of **U**, another rotation.

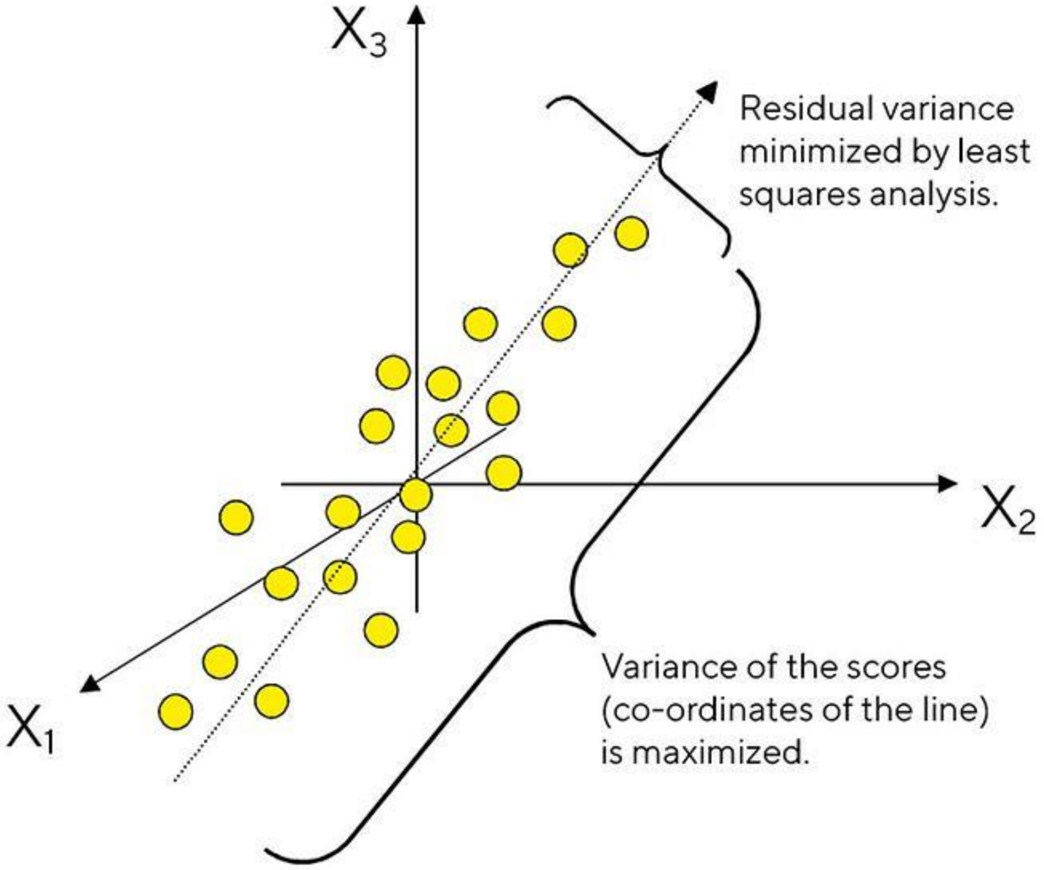


* Stands for complex conjugate transpose

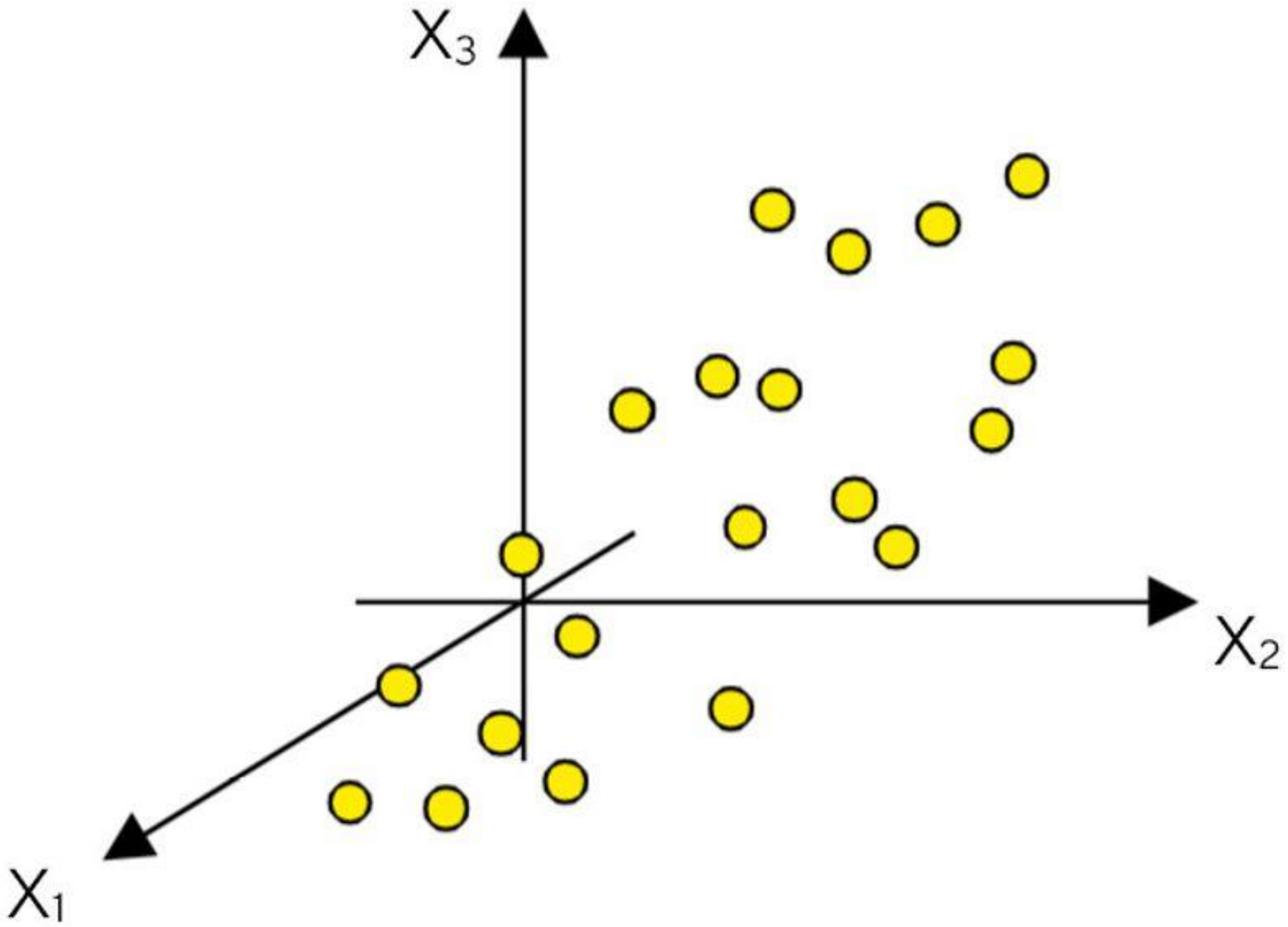
Principal Component Analysis (PCA)

- How PCA works

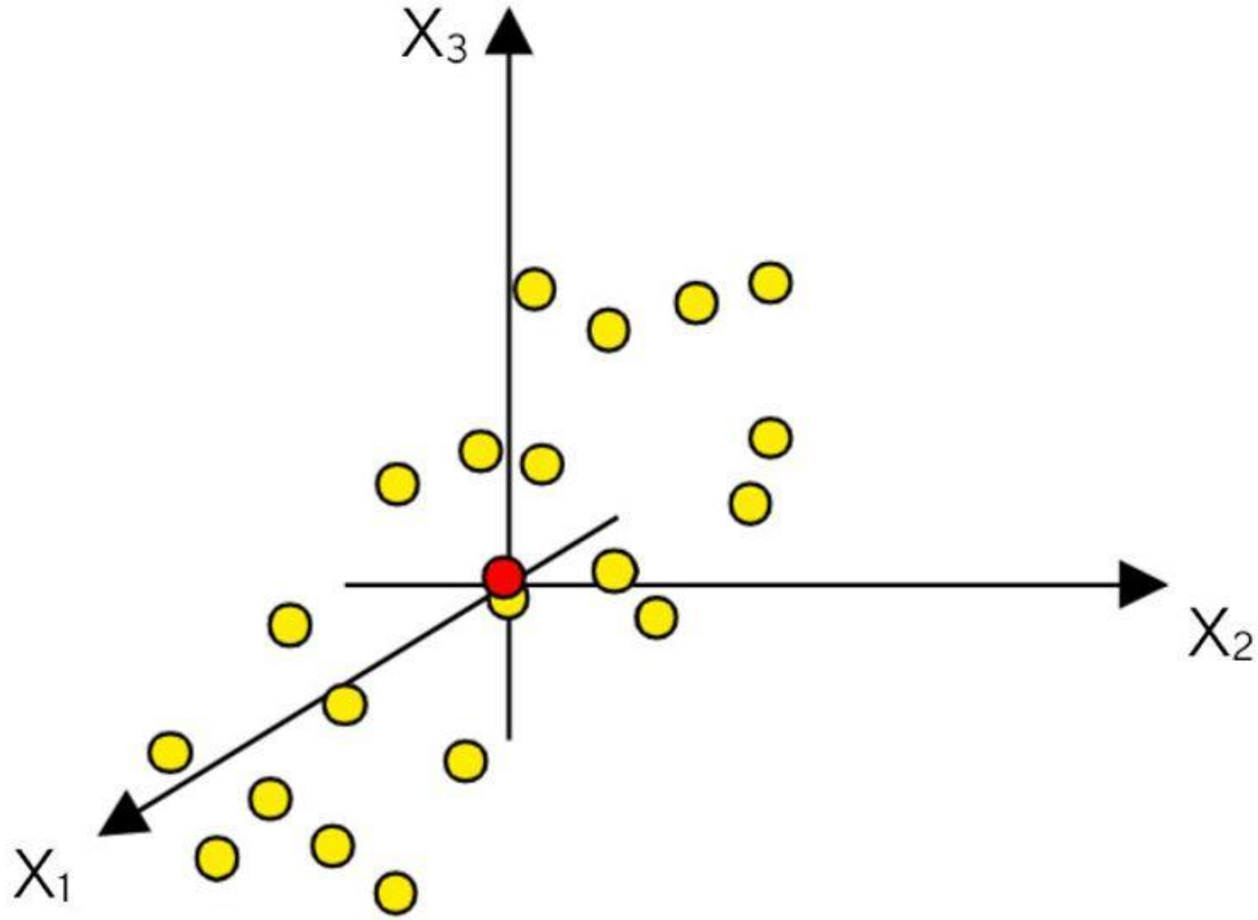
PCA


X3
X2
X1
Residual variance minimized by least squares analysis.
Variance of the scores (co-ordinates of the line) is maximized.

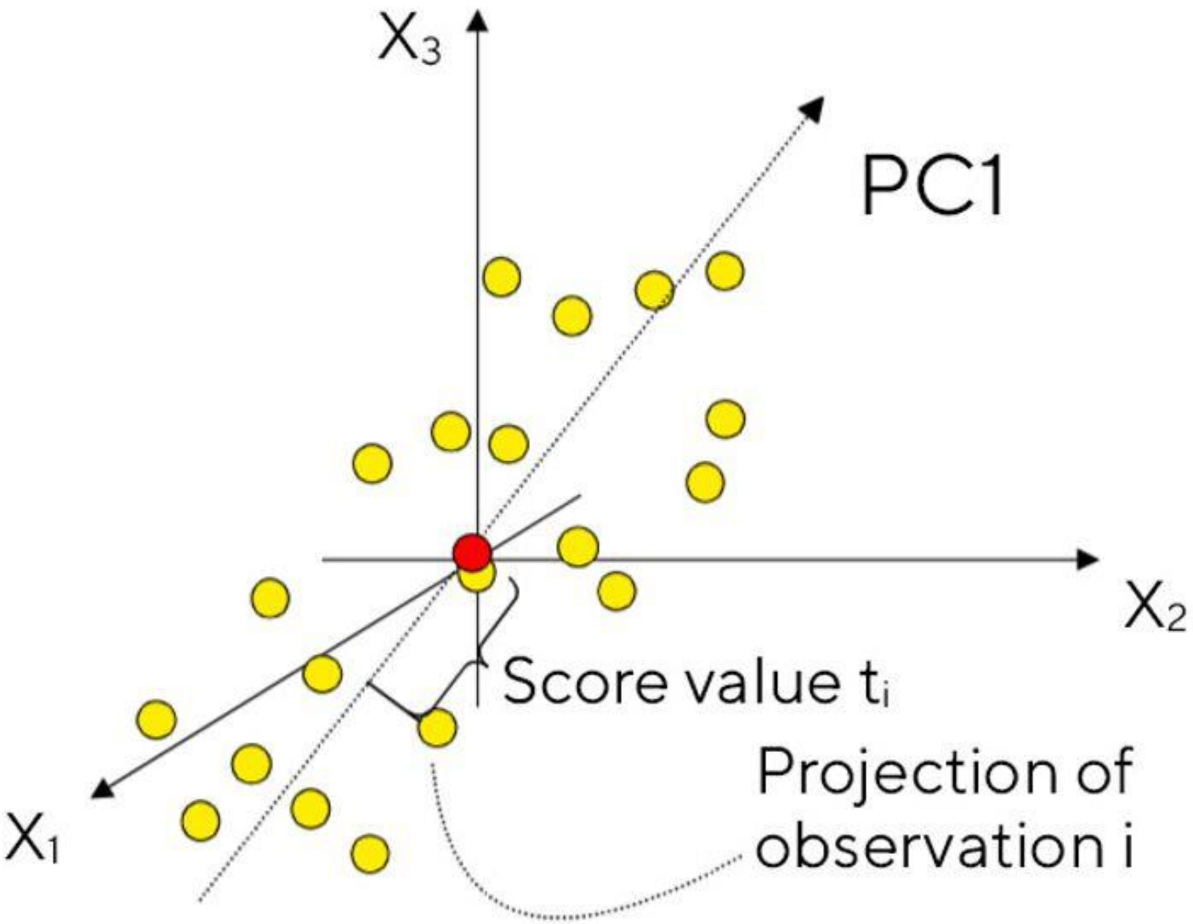
How PCA Works



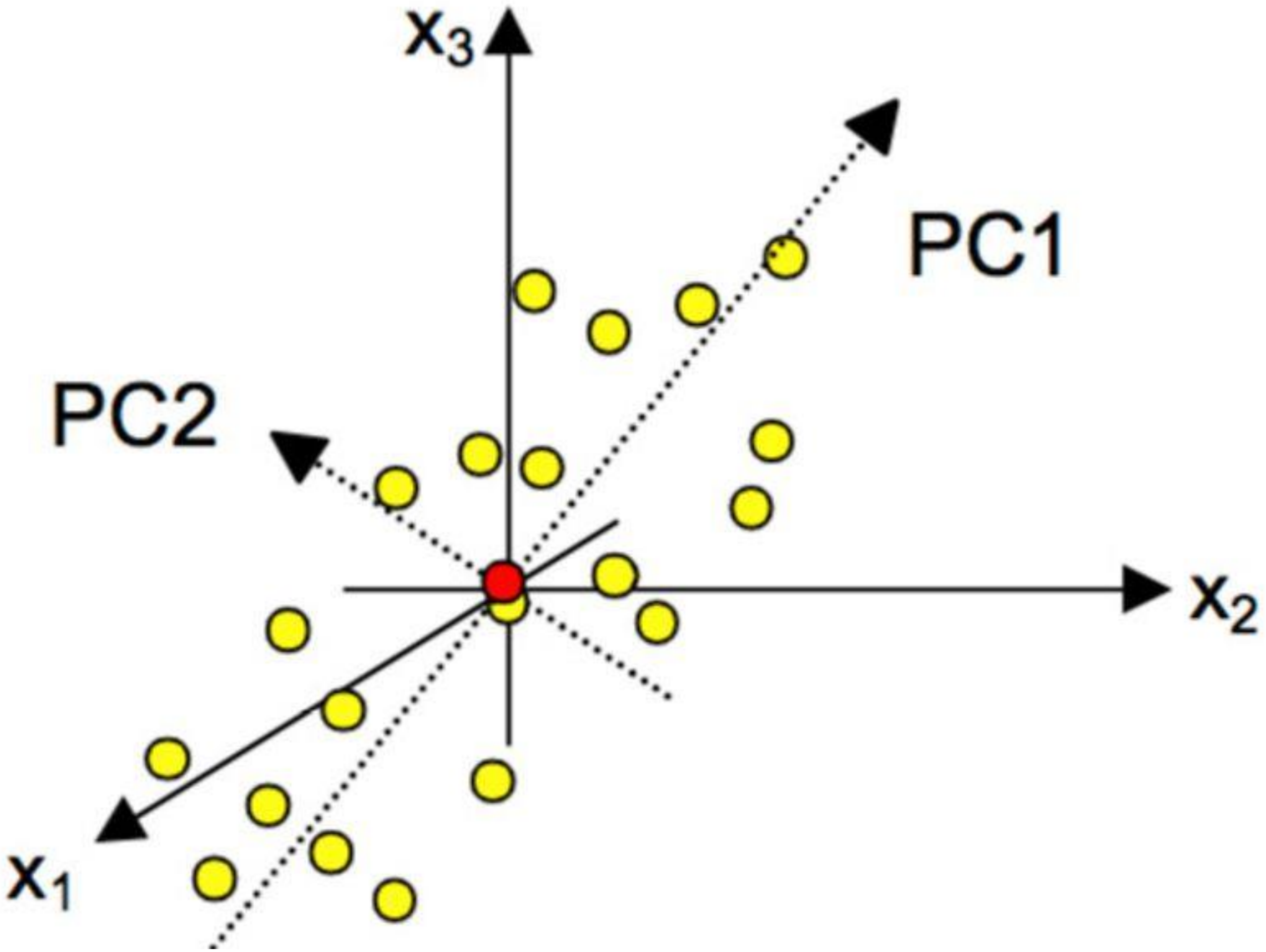
Mean Centering



PCA Projections



PCA Projections



PCA

- Principal components are linear combinations of the original features:
  - PC1 = $a_1 * x_1 + a_2 * x_2 + \ldots. + a_n * x_n$ (first principal component)
  - PC2 = $b_1 * x_1 + b_2 * x_2 + \ldots. + b_n * x_n$ (second principal component)

    …
  - PC1, PC2, … are called PCA sores (or simply scores)
  - Coefficients (a's, b's, …) are called loadings
- Principal components are orthogonal
- PC1 explains the most variance in the data, followed by PC2, PC3, and so on
  - PC1 > PC2 > PC3 > …