KOMINFO digitalent

PRO Professional Academy

# Digital Talent Scholarship 2022

## Math For Machine Learning - PCA

Lead a sprint through the Machine Learning Track

# Agenda

- Statistics of Datasets
- Inner Products

# Are your students ML-ready?

# Recap

# Pelajaran hari ini

- Kita akan memahami PCA dari sudut pandang geometri.
- Kita akan belajar cara summarize datasets dengan statistik dasar.
- Perubahan mean dan variance saat kita shift atau scale original dataset.
- Akan ada jupyter notebook yang membantu pemahaman matematika dengan coding.

# Introduction

Principal Component Analysis (PCA) adalah salah satu algoritma dimensionality reduction terpenting dalam Machine Learning.

# PART 1

## Statistics

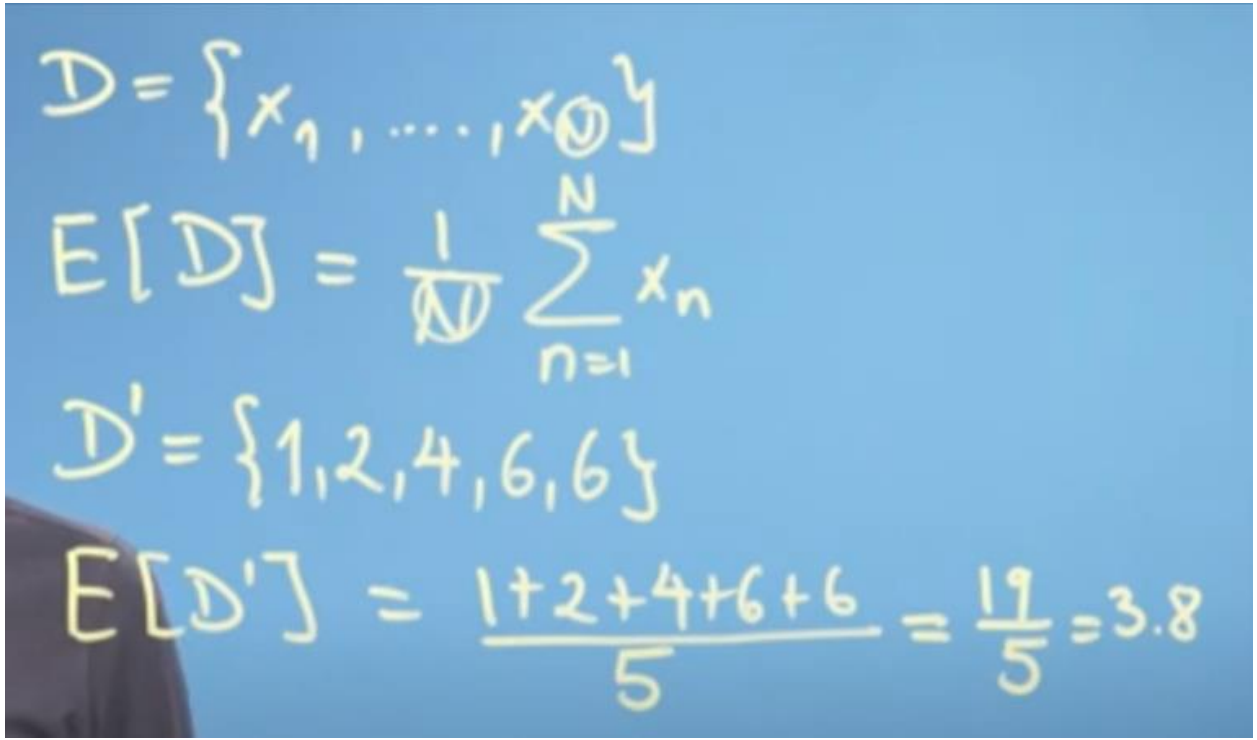
[Calculating the Mean, Variance and Standard Deviation, Clearly Explained!!! - YouTube]

[Covariance, Clearly Explained!!! - YouTube]

# Objektif Pembelajaran

- Compute statistik dasar dari sebuah data sets
- Memahami efek dalam linear transformations terhadap means dan (co)variances
- Menghitung means/variances dari linearly transformed data sets
- Membuat kode yang dapat merepresentasikan gambar sebagai vector
- Membuat kode yang dapat menghitung statistik dasar sebuah datasets

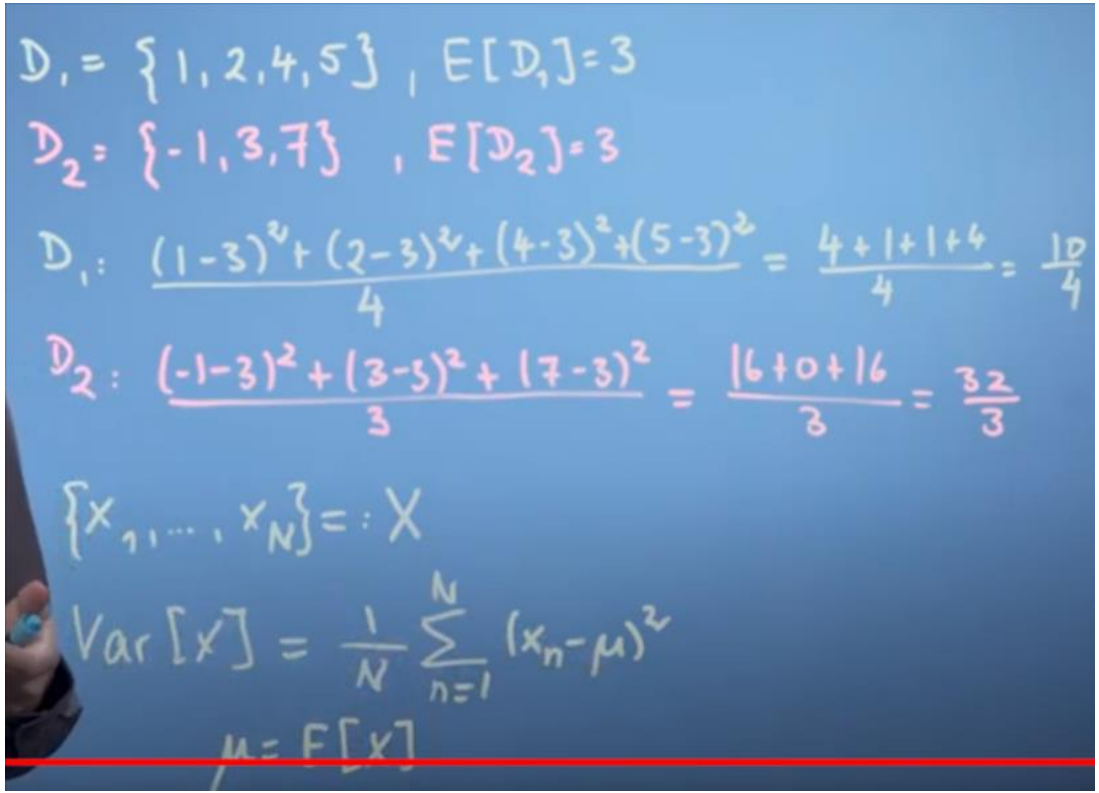
# Mean



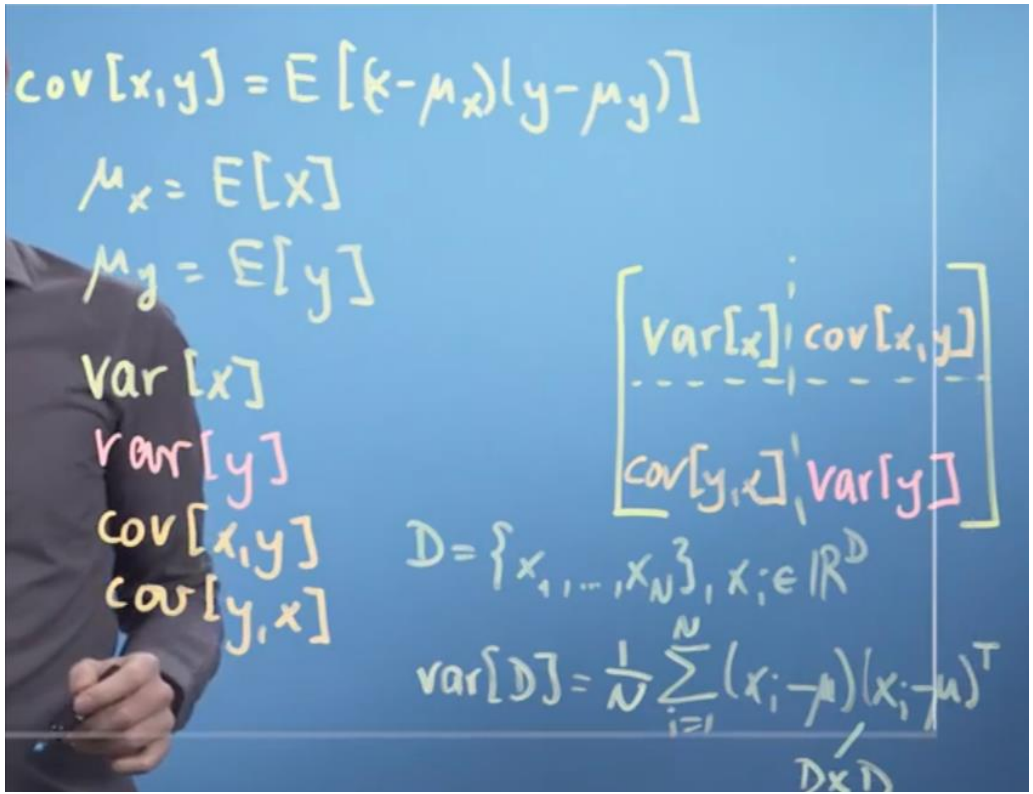

D = Data

E = Expected Value

# Variance


$D_1 = \{1, 2, 4, 5\}$ , $E[D_1] = 3$
$D_2 = \{-1, 3, 7\}$ , $E[D_2] = 3$
$D_1 : \frac{(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2}{4} = \frac{4+1+1+4}{4} = \frac{10}{4}$
$D_2 : \frac{(-1-3)^2 + (3-3)^2 + (7-3)^2}{3} = \frac{16+0+16}{3} = \frac{32}{3}$
$\{x_1, \ldots, x_N\} =: X$
$Var[X] = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2$
$\mu = E[X]$

# Covariance


cov[x,y] = E[(x-μx)(y-μy)]
μx = E[x]
μy = E[y]
Var[x]
var[y]
cov[x,y]
cov[y,x]
var[x] cov[x,y]
cov[y,x] var[y]
D = {x1, ..., xN}, xi ∈ R^D
var[D] = 1/N Σ (xi - μ)(xi - μ)^T
D x D

Range, variance and standard deviation as measures of dispersion | Khan Academy

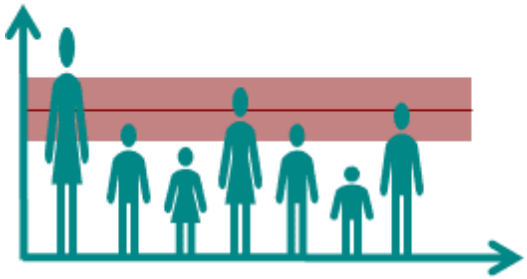
# Spread

The 5 main values of measuring spread:

- Range, bottom to top
- Quartiles, bagi empat
- Interquartile range (IQR), Q1 - Q3
- Variance ($\sigma^2$),
- Standard deviation ($\sigma$), aka. Simpangan baku

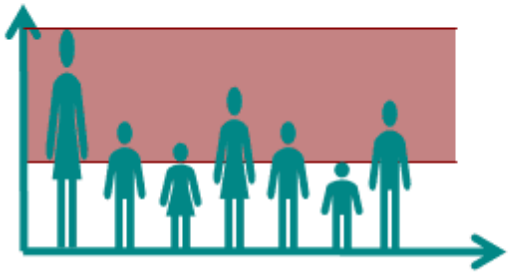
# Spreads

## Standard deviation / variance



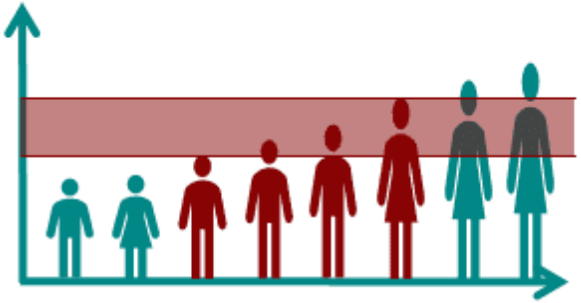
Average distance of all measured values from the mean value

## Range



Distance between lowest and highest value of a distribution

## Quantile distance



Spectrum in which the middle 50% of the values lie.

Difference between the first and the third quartile

# Variance and Standard Deviation

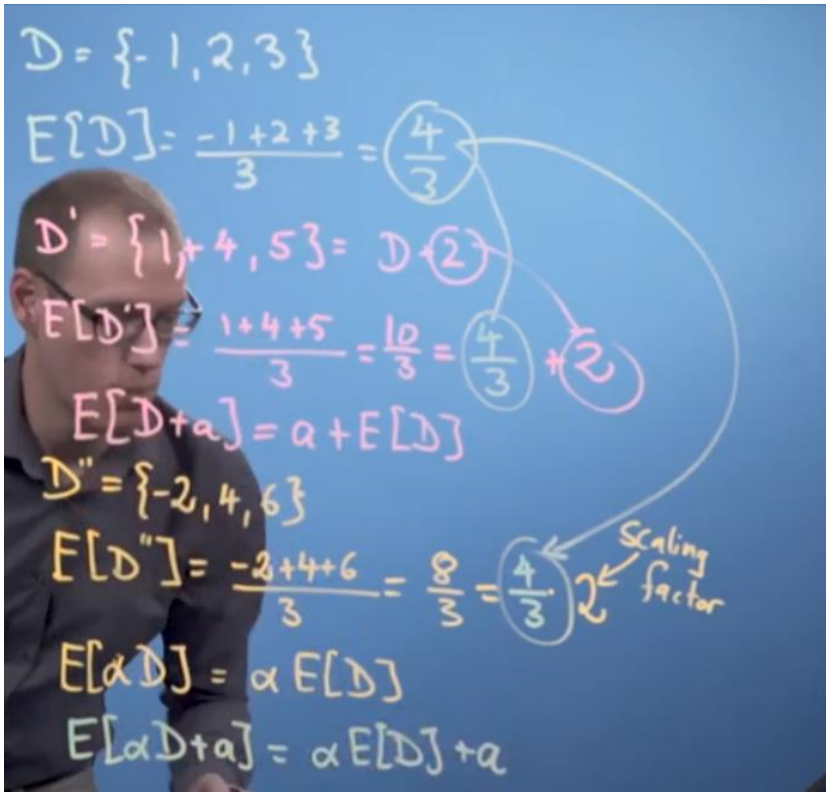
## Variance and Standard Deviation Formula

cuemath
THE MATH EXPERT

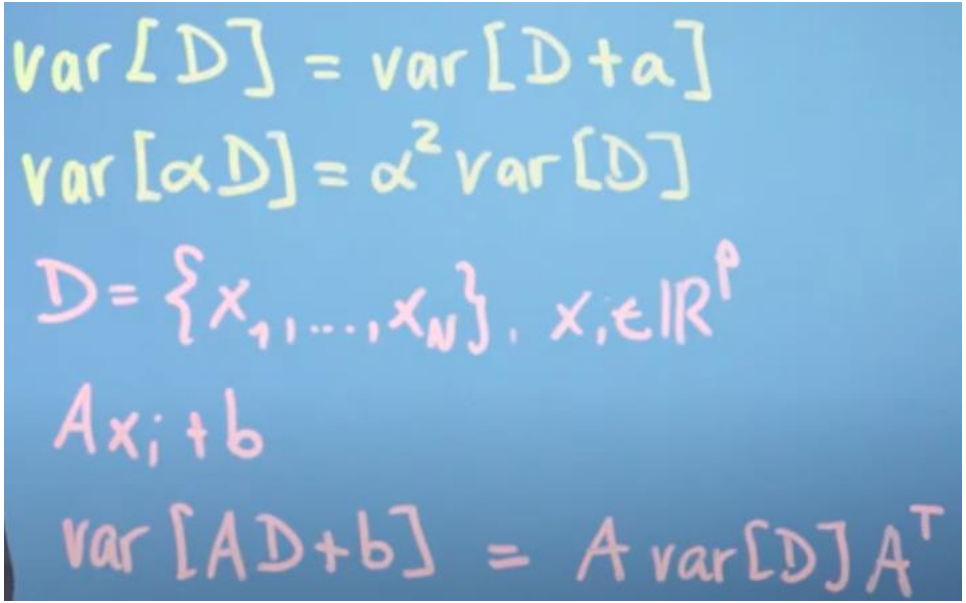
| | Population | Sample |
|---|---|---|
| Variance | $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}$ | $S^2 = \frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}$ |
| Standard Deviation | $\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}}$ | $S = \sqrt{\frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}}$ |

[Linear transformations | Matrix transformations | Linear Algebra | Khan Academy - YouTube]

# Linear Transformation of Datasets



# Linear Transformation of Datasets



# DEMO Mean Covariance Effect on Linear Transformation

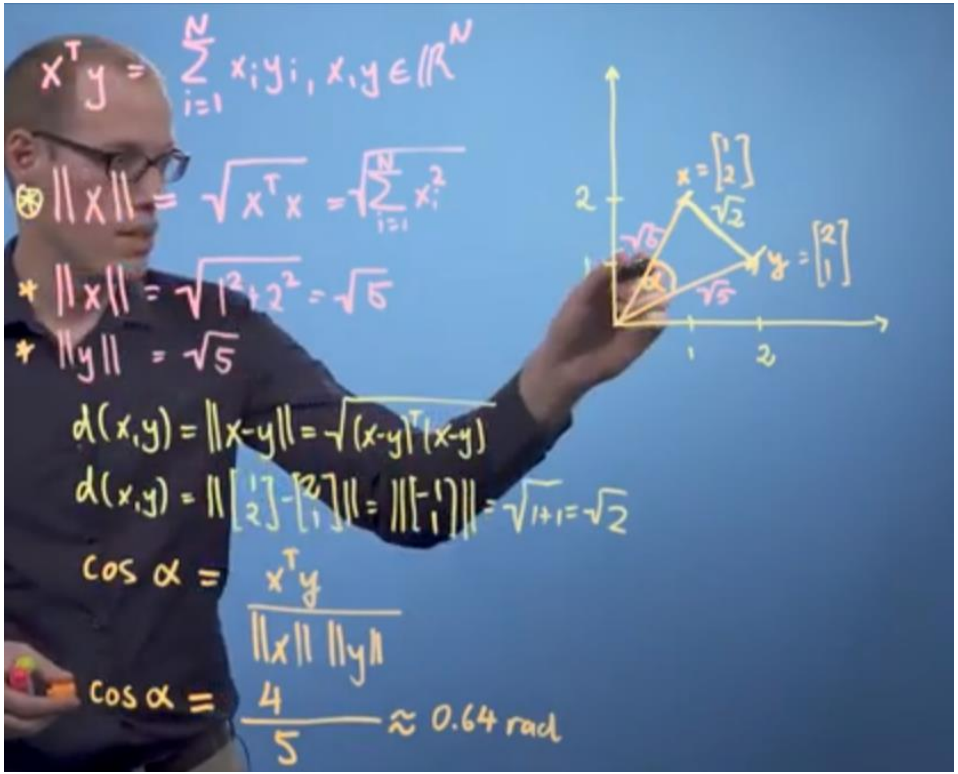
# PART 2

## Inner Product

# Objektif Pembelajaran

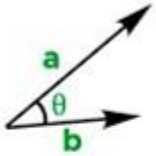
- Memahami apa itu inner products
- Menghitung sudut dan jarak dengan menggunakan inner products
- Menulis kode untuk menghitung jarak dan sudut antar data
- Memahami properti inner products
- Mengetahui bahwa orthogonality tergantung pada inner product

# Dot Product


$x^T y = \sum_{i=1}^{N} x_i y_i,\ x, y \in \mathbb{R}^N$
$\|x\| = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{N} x_i^2}$
$\|x\| = \sqrt{1^2+2^2} = \sqrt{5}$
$\|y\| = \sqrt{5}$
$d(x,y) = \|x-y\| = \sqrt{(x-y)^T(x-y)}$
$d(x,y) = \left\|\begin{bmatrix}1\\2\end{bmatrix}-\begin{bmatrix}2\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}-1\\1\end{bmatrix}\right\| = \sqrt{1+1} = \sqrt{2}$
$\cos\alpha = \frac{x^T y}{\|x\|\,\|y\|}$
$\cos\alpha = \frac{4}{5} \approx 0.64\ \text{rad}$
$x = \begin{bmatrix}1\\2\end{bmatrix}$
$y = \begin{bmatrix}2\\1\end{bmatrix}$
$\sqrt{2}$
$\sqrt{5}$
$\alpha$
2
1
1
2

# Dot Product

## The Vector Dot Product



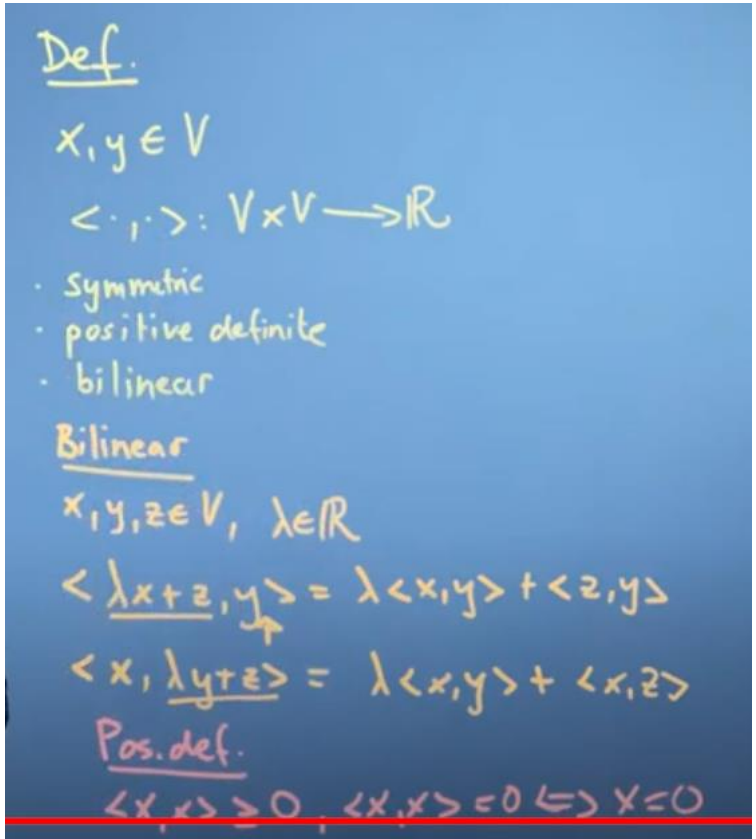

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$$

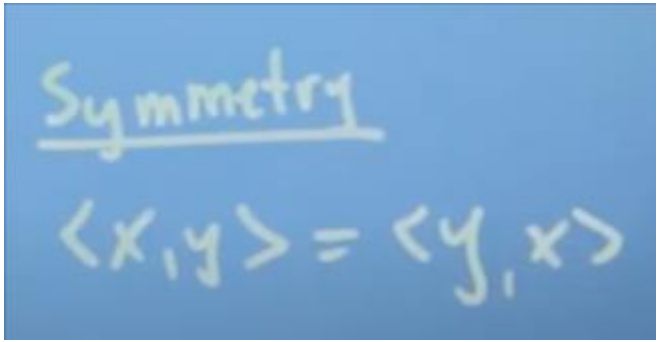
$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

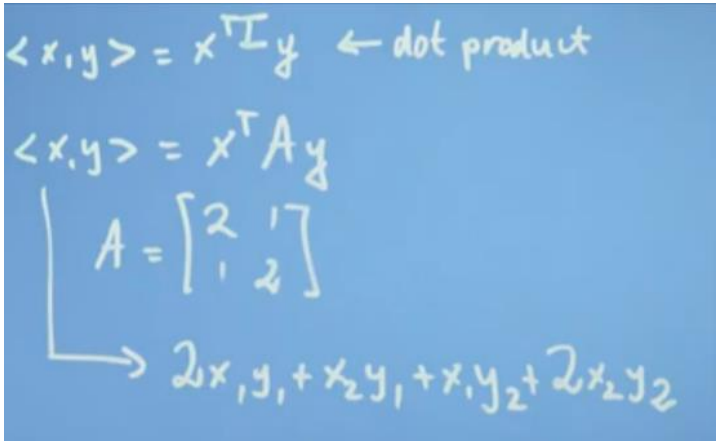
KOMINFO digitalent

PCA - Unit 9 - Section 9.5 - Dot Product, Angle Bewteen Vectors and Work - YouTube

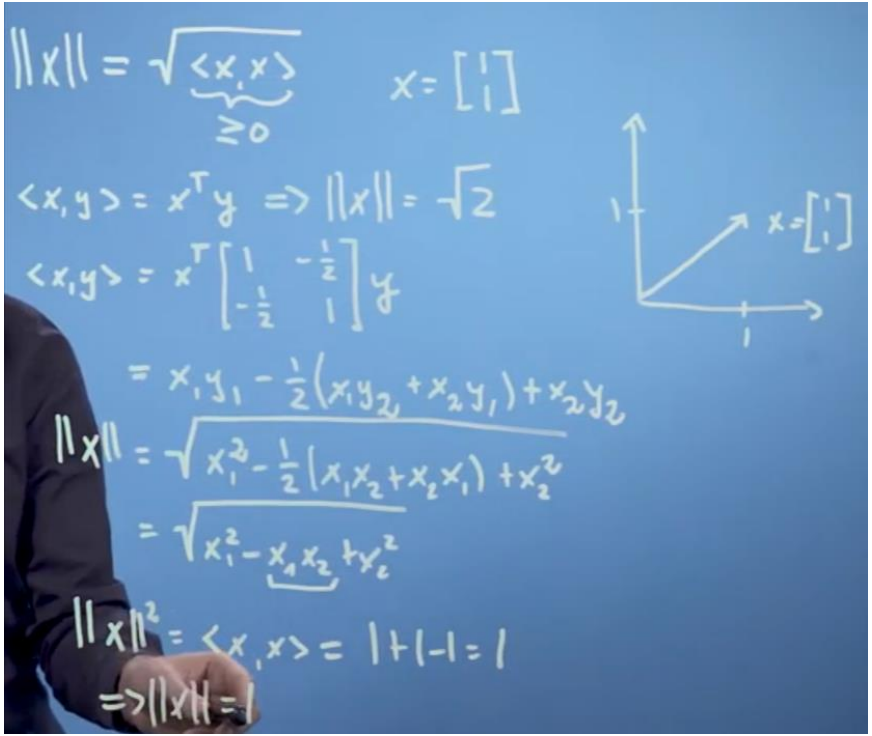
# Inner Product Definition and Properties



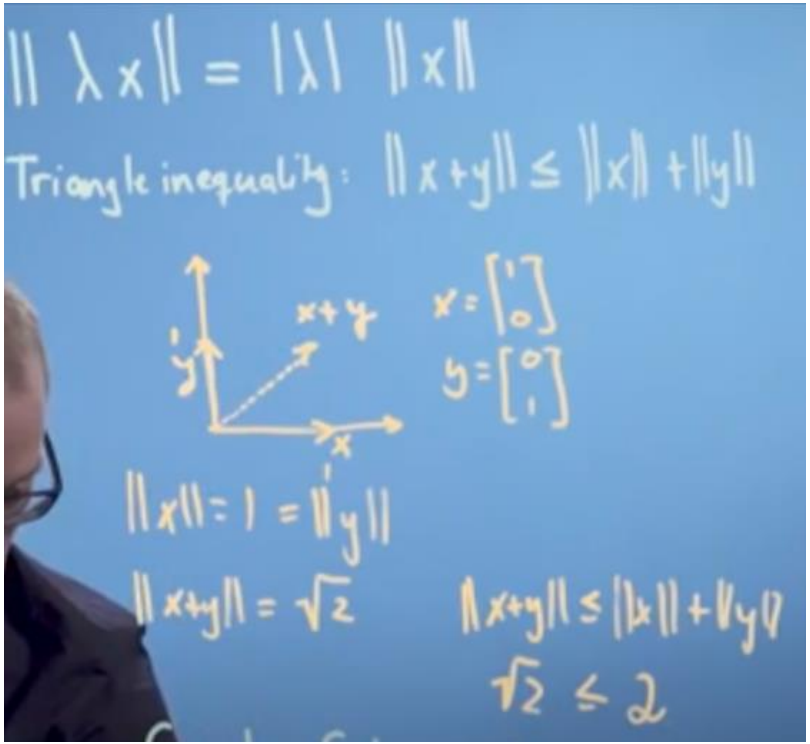



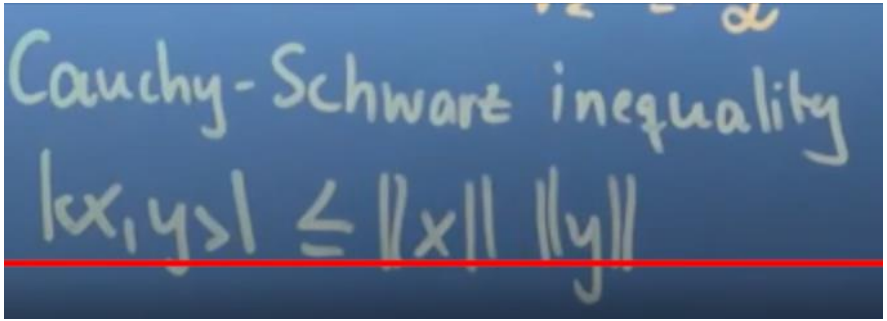



# Inner Product: Length of Vectors

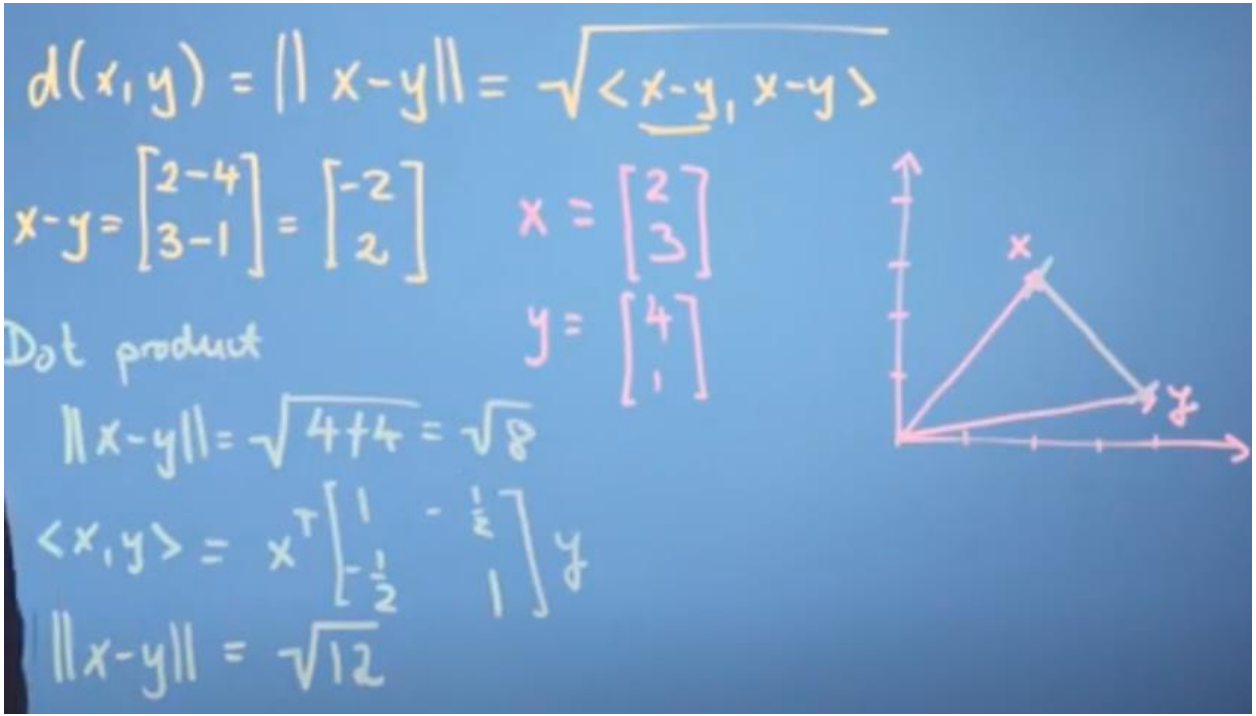


# Inner Product: Length of Vectors

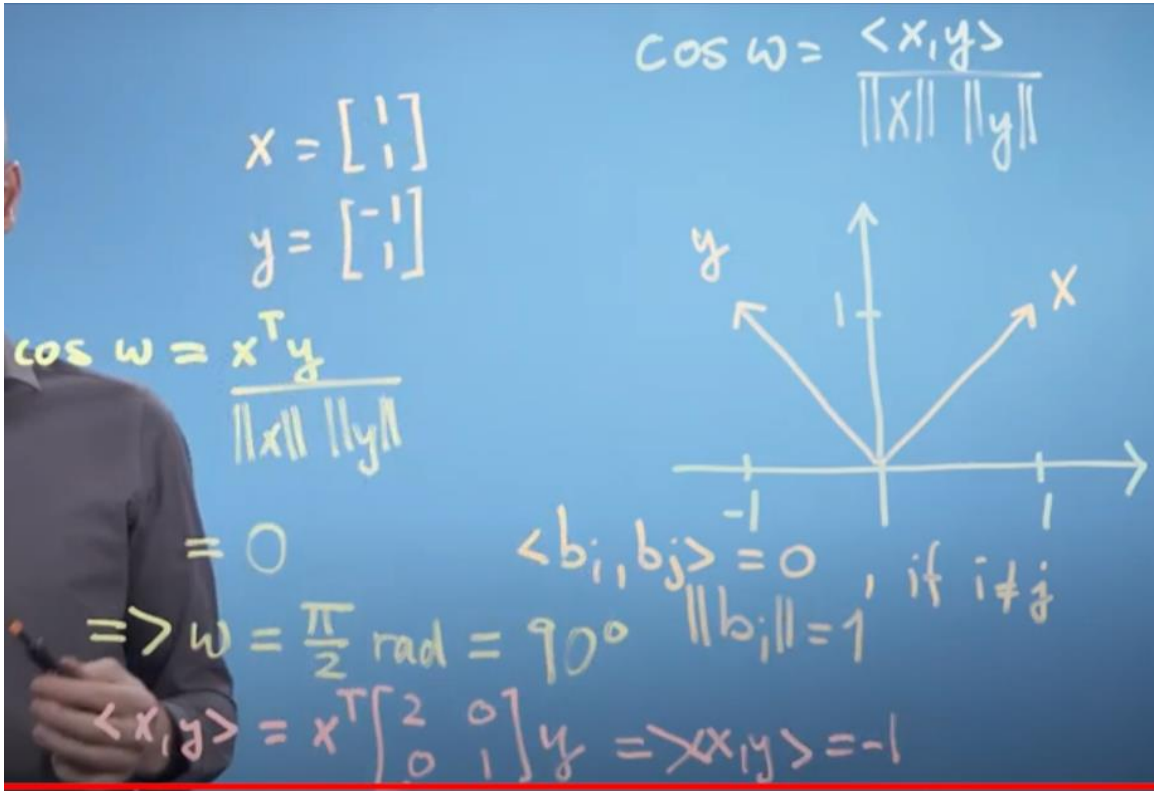


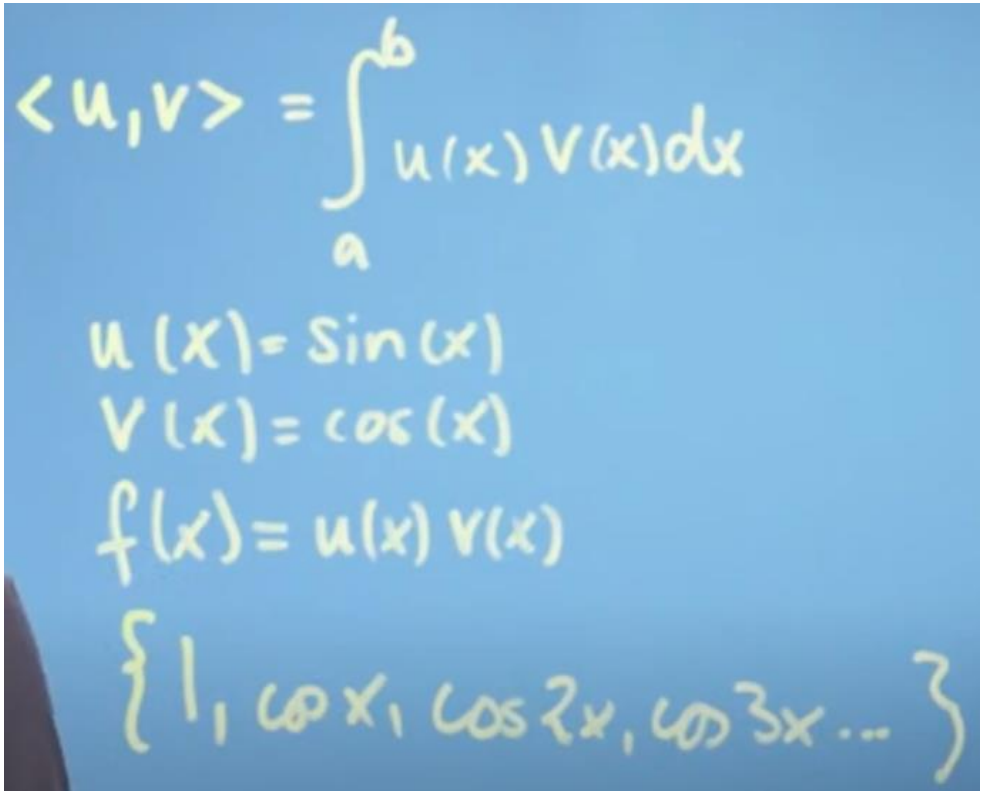
# Inner Product: Distance between Vectors



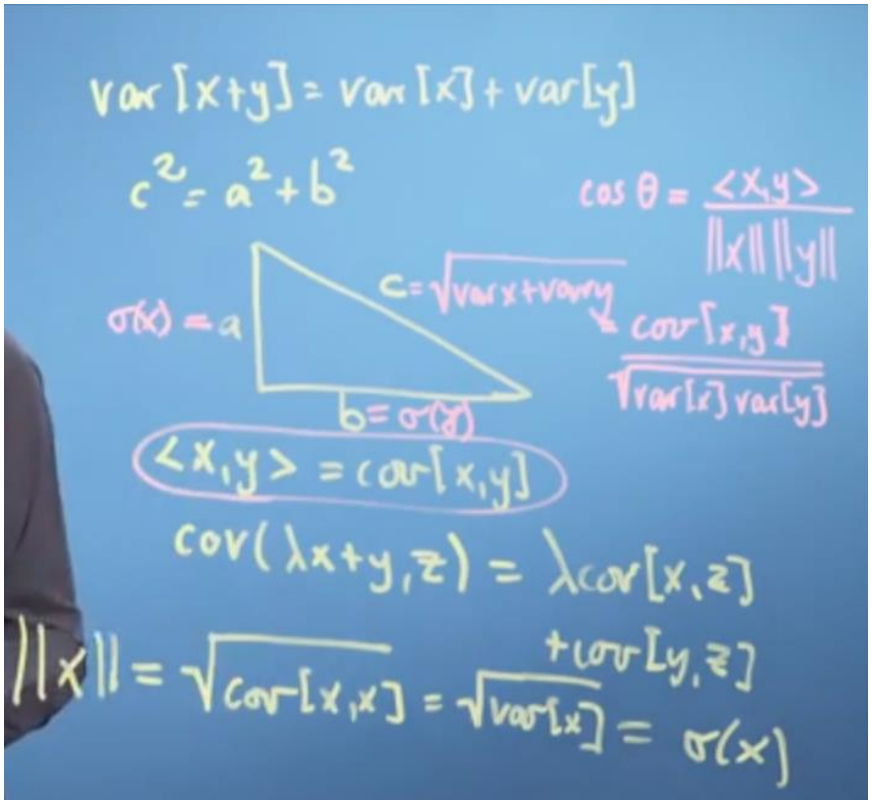
# Inner Product: Angles and Orthogonality



# Unconventional Inner Product for Function (Advanced)



# Unconventional Inner Product (Advanced)



KOMINFO digitalent

# PCA by Imperiall Collage London

# Q & A

# Thank You