

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

$$D = \{x_1, \dots, x_N\}$$

$$E[D] = \frac{1}{N} \sum_{n=1}^N x_n$$

$$D' = \{1, 2, 4, 6, 6\}$$

$$E[D'] = \frac{1+2+4+6+6}{5} = \frac{19}{5} = 3.8$$

D = Data

E = Expected Value

Variance

$$D_1 = \{1, 2, 4, 5\}, E[D_1] = 3$$

$$D_2 = \{-1, 3, 7\}, E[D_2] = 3$$

$$D_1: \frac{(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2}{4} = \frac{4+1+1+4}{4} = \frac{10}{4}$$

$$D_2: \frac{(-1-3)^2 + (3-3)^2 + (7-3)^2}{3} = \frac{16+0+16}{3} = \frac{32}{3}$$

$$\{x_1, \dots, x_N\} =: X$$

$$\text{Var}[X] = \frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2$$

$$\mu = E[X]$$

Covariance

$$\text{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)]$$

$$\mu_x = E[x]$$

$$\mu_y = E[y]$$

$$\text{var}[x]$$

$$\text{var}[y]$$

$$\text{cov}[x, y]$$

$$\text{cov}[y, x]$$

$$\begin{bmatrix} \text{var}[x] & \text{cov}[x, y] \\ \text{cov}[y, x] & \text{var}[y] \end{bmatrix}$$

$$D = \{x_1, \dots, x_N\}, x_i \in \mathbb{R}^D$$

$$\text{var}[D] = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T$$

$D \times D$

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

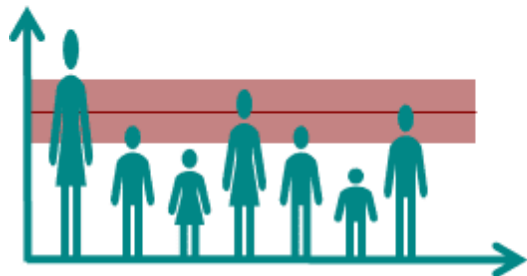
Spread

The 5 main values of measuring spread:

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

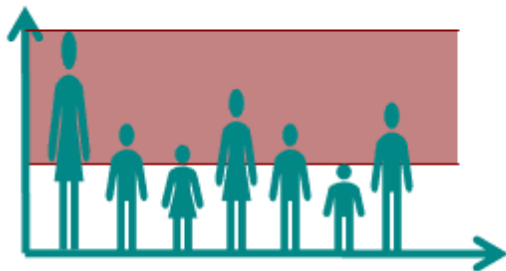
Spreads

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

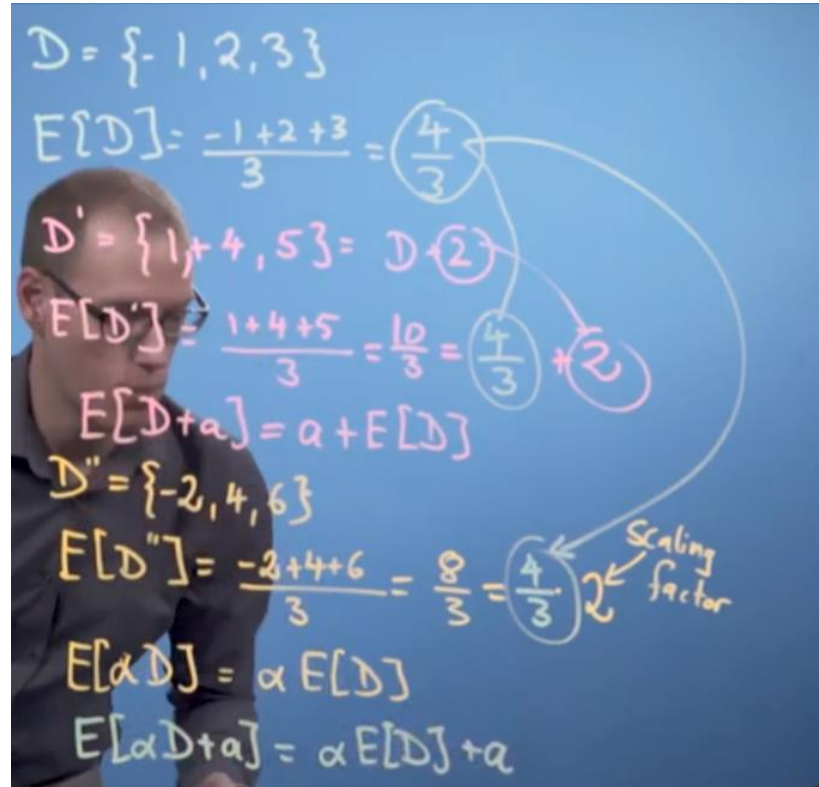
Variance and Standard Deviation Formula



| | 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



$D = \{-1, 2, 3\}$

$E[D] = \frac{-1+2+3}{3} = \frac{4}{3}$

$D' = \{1, 4, 5\} = D + 2$

$E[D'] = \frac{1+4+5}{3} = \frac{10}{3} = \frac{4}{3} + 2$

$E[D+a] = a + E[D]$

$D'' = \{-2, 4, 6\}$

$E[D''] = \frac{-2+4+6}{3} = \frac{8}{3} = \frac{4}{3} \cdot 2$ (Scaling factor)

$E[\alpha D] = \alpha E[D]$

$E[\alpha D + a] = \alpha E[D] + a$

Linear Transformation of Datasets

$$\text{var}[D] = \text{var}[D + a]$$

$$\text{var}[\alpha D] = \alpha^2 \text{var}[D]$$

$$D = \{x_1, \dots, x_N\}, x_i \in \mathbb{R}^p$$

$$Ax_i + b$$

$$\text{var}[AD + b] = A \text{var}[D] A^T$$

DEMO Mean Covariance Effect on Linear Transformation

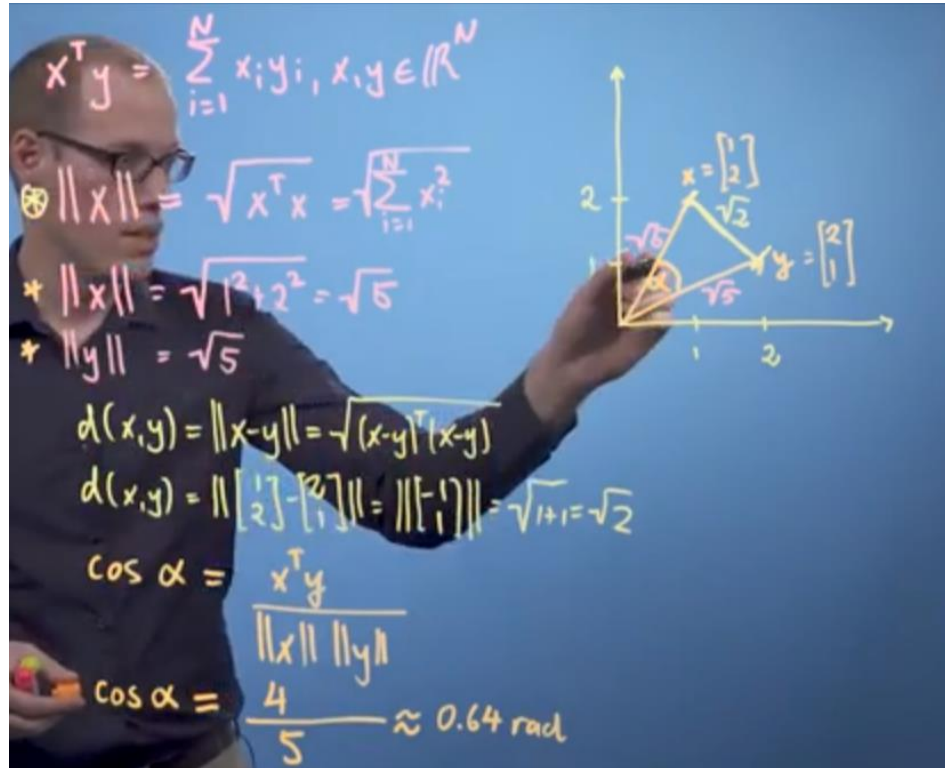
PART 2

Inner Product

Objektif Pembelajaran

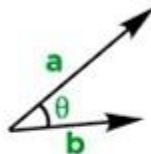
- Memahami apa itu inner products
- Menghitung sudut dan jarak dengan menggunakan inner products
- Menulis kode untuk menghitung jarak dan sudut antar data
- Memahami properti inner products
- Mengetahui bahwa orthogonality tergantung pada inner product

Dot Product



Dot Product

The Vector Dot Product



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

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

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

Inner Product Definition and Properties

Def.

$$x, y \in V$$

$$\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$$

- Symmetric
- positive definite
- bilinear

Bilinear

$$x, y, z \in V, \lambda \in \mathbb{R}$$

$$\langle \lambda x + z, y \rangle = \lambda \langle x, y \rangle + \langle z, y \rangle$$

$$\langle x, \lambda y + z \rangle = \lambda \langle x, y \rangle + \langle x, z \rangle$$

Pos. def.

$$\langle x, x \rangle \geq 0, \langle x, x \rangle = 0 \Leftrightarrow x = 0$$

Symmetry

$$\langle x, y \rangle = \langle y, x \rangle$$

$$\langle x, y \rangle = x^T I y \leftarrow \text{dot product}$$

$$\langle x, y \rangle = x^T A y$$

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

$$\rightarrow 2x_1y_1 + x_2y_1 + x_1y_2 + 2x_2y_2$$

Inner Product: Length of Vectors

$$\|x\| = \sqrt{\underbrace{\langle x, x \rangle}_{\geq 0}} \quad x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\langle x, y \rangle = x^T y \Rightarrow \|x\| = \sqrt{2}$$


$$\langle x, y \rangle = x^T \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} y$$

$$= x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2$$

$$\|x\| = \sqrt{x_1^2 - \frac{1}{2}(x_1 x_2 + x_2 x_1) + x_2^2}$$


$$= \sqrt{x_1^2 - x_1 x_2 + x_2^2}$$

$$\|x\|^2 = \langle x, x \rangle = 1 + 1 - 1 = 1$$

$$\Rightarrow \|x\| = 1$$


Inner Product: Length of Vectors

$\|\lambda x\| = |\lambda| \|x\|$
 Triangle inequality: $\|x+y\| \leq \|x\| + \|y\|$



$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
 $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\|x\| = 1 = \|y\|$
 $\|x+y\| = \sqrt{2}$

$\|x+y\| \leq \|x\| + \|y\|$
 $\sqrt{2} \leq 2$


Cauchy-Schwarz inequality

$$|x, y| \leq \|x\| \|y\|$$

Inner Product: Distance between Vectors

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$$
$$x - y = \begin{bmatrix} 2 - 4 \\ 3 - 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \quad x = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad y = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Dot product

$$\|x - y\| = \sqrt{4 + 4} = \sqrt{8}$$
$$\langle x, y \rangle = x^T \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} y$$
$$\|x - y\| = \sqrt{12}$$


Inner Product: Angles and Orthogonality

Handwritten mathematical derivations and a vector diagram on a blue background.

Top right formula:

$$\cos \omega = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

Left side definitions:

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

$$y = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Left side formula:

$$\cos \omega = \frac{x^T y}{\|x\| \|y\|}$$

Left side result:

$$= 0$$

Left side conclusion:

$$\Rightarrow \omega = \frac{\pi}{2} \text{ rad} = 90^\circ$$

Bottom left calculation:

$$\langle x, y \rangle = x^T \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} y = \langle x, y \rangle = -1$$

Right side diagram:

A 2D coordinate system with x and y axes. The x-axis has tick marks at -1 and 1. The y-axis has a tick mark at 1. Two vectors, labeled x and y, originate from the origin. Vector x points into the first quadrant, and vector y points into the second quadrant. The angle between them is 90 degrees.

Bottom right formula:

$$\langle b_i, b_j \rangle = 0, \text{ if } i \neq j$$

$$\|b_i\| = 1$$

Unconventional Inner Product for Function (Advanced)

$$\langle u, v \rangle = \int_a^b u(x) v(x) dx$$

$$u(x) = \sin(x)$$

$$v(x) = \cos(x)$$

$$f(x) = u(x) v(x)$$

$$\{1, \cos x, \cos 2x, \cos 3x, \dots\}$$

Unconventional Inner Product (Advanced)

$$\begin{aligned} \text{var}[x+y] &= \text{var}[x] + \text{var}[y] \\ c^2 &= a^2 + b^2 \\ \cos \theta &= \frac{\langle x, y \rangle}{\|x\| \|y\|} \\ \sigma(x) &= a \\ c &= \sqrt{\text{var}[x] + \text{var}[y]} \\ b &= \sigma(y) \\ \langle x, y \rangle &= \text{cov}[x, y] \\ \text{cov}(\lambda x + y, z) &= \lambda \text{cov}[x, z] + \text{cov}[y, z] \\ \|x\| &= \sqrt{\text{cov}[x, x]} = \sqrt{\text{var}[x]} = \sigma(x) \end{aligned}$$

PCA by Imperiall Collage London

Q & A

Thank You