

# Assignment 1: Example #2

PCA on Height Weight & Age

3D linear moving features into one principal component

Person	Height	Weight	Age
A	150	50	20
B	160	60	25
C	170	65	30
D	180	80	35
E	190	88	40

Matrix X:

$$X = \begin{bmatrix} 150 & 50 & 20 \\ 160 & 60 & 25 \\ 170 & 65 & 30 \\ 180 & 80 & 35 \\ 190 & 88 & 40 \end{bmatrix}$$

Mean:

$$850/5 = 170$$

$$340/5 = 68$$

$$150/5 = 30$$

Let's standardize the data.

$$Z = \frac{X - \bar{X}(\text{mean})}{\text{Std.}}$$

$$S.D = \sqrt{\frac{(x - \bar{x})^2}{n-1}}$$

Mean & Std.

Feature	Mean	S.D
Height	170	15.81
Weight	68	13.04
Age	30	7.91

Standardized Matrix  $Z$

$$Z = \begin{bmatrix} -1.26 & -1.88 & 1.26 \\ -0.63 & -0.61 & -0.63 \\ 0 & -0.22 & 0 \\ 0.63 & 0.92 & 0.63 \\ 1.26 & 1.80 & 1.26 \end{bmatrix}$$

Compute Covariance Matrix

$$S = \frac{Z^T Z}{n-1}$$

Computed using numpy

$$S = \begin{bmatrix} 1.00 & 0.99 & 0.99 \\ 0.99 & 1.00 & 0.99 \\ 0.99 & 0.99 & 1.00 \end{bmatrix}$$

All features are highly correlated ( $\approx 0.99$ ), they share almost the same pattern.

Find Eigen Values & Eigen Vectors

$$(S - \lambda I) = 0$$

Eigen Values  $\Rightarrow$  How much variance each principal component explains.

~~Eigen Vectors~~  $\lambda_1 = 2.98$ ,  $\lambda_2 = 0.02$  &  $\lambda_3 = 0.00$

Eigen Vectors  $\Rightarrow$  Direction of each new principal component.

Component

pc1

pc2

pc3

Eigen Vector (direction)

$$[0.58, 0.58, 0.58]$$

$$[0.70, -0.71, 0.00]$$

$$[-0.40, -0.40, 0.82]$$

## Variance Ratio:

$$\text{Explained Variance Ratio} = \frac{\lambda_i}{\sum \lambda_i}$$

$$= [0.993, 0.007, 0.000]$$

PC1 Explains 99.3% of total variance

⇒ All 3 Variables move together in one direction.

## Transform Data:

$$\text{PCA} = Z \times \text{eigen Vectors}$$

$$\text{PC1} = 0.58 \times \text{Height}_Z + 0.58 \times \text{Weight}_Z + 0.53 \times \text{Age}_Z$$

Person	Height_Z	Weight_Z	Age_Z	PC1
A	-1.26	0.93	-1.26	-2.38
B	-0.63	-0.61	-0.63	-1.10
C	0	-0.23	0	-0.13
D	0.63	0.92	0.63	1.30
E	1.26	1.80	1.26	2.31

## Interpretation:

The First Principal Component (PC1) is a combination of Height, weight & Age.

Since it captures 99% of variance, we can represent all three features with just one variable.