Taking *Linear Algebra* 

… to *Data Science*

Instructor: Neeraj Sharma
MF School Data Science & AI
IIT Guwahati

About Me

- Assistant Professor
- Signals are everywhere, waiting to be processed!
  - Have worked with speech, music, and EEG signals
- Think, Capture, Analyze, Visualize, Create

- Also, enjoy:
  - sports - badminton, swimming (beginner), popular science, world cinema …

- Your instructor for three lectures!

Linear Algebra (LA)

- الجبر *al-jabr:* The root word for “algebra”, thanks to a Persian Mathematician (Ref: wiki) (discuss)

- Relates to solving a system of equations (discuss)

- What is “linear” here? (discuss)
  - Limits the allowed operations
  - Helps keep the chain of operations simple
  - Thus, makes math “easy”

Ingredients of LA

- Scalars: let’s understand them

Can you give an estimate of Murali’s

- Age
- Weight
- Height

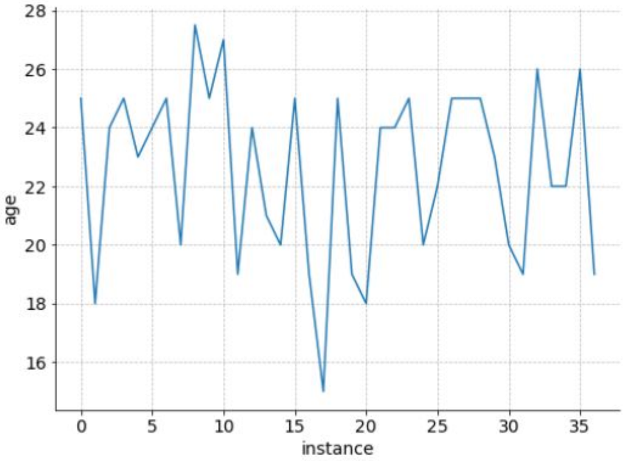
Go to url: shorturl.at/AINOS (note the CAPS)

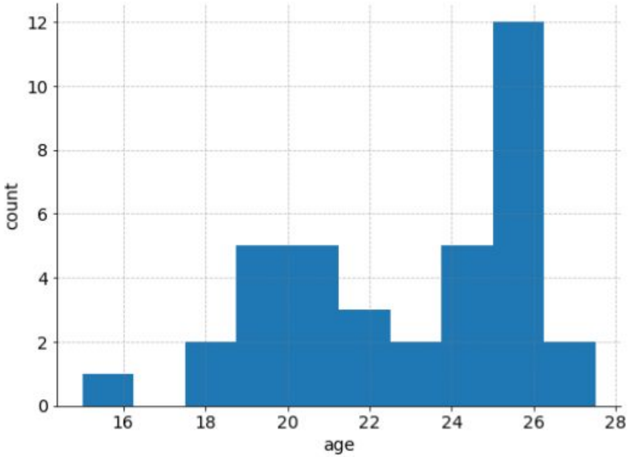
Murali

Let’s get the data you filled in

- Download the CSV file
- Read CSV file into python - jupyter notebook
    - Data cleaning - remove bad values
    - Plot the data: We got 36 3-D data points

Age:

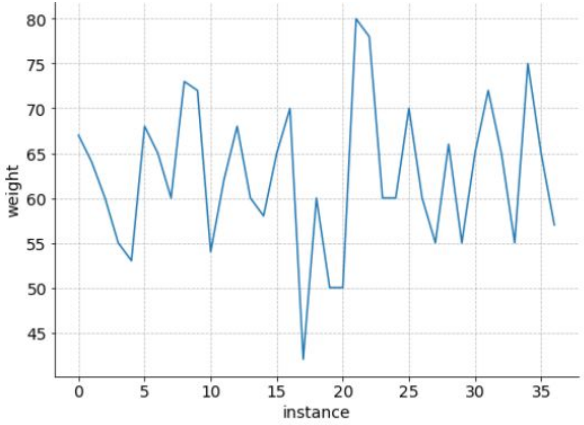


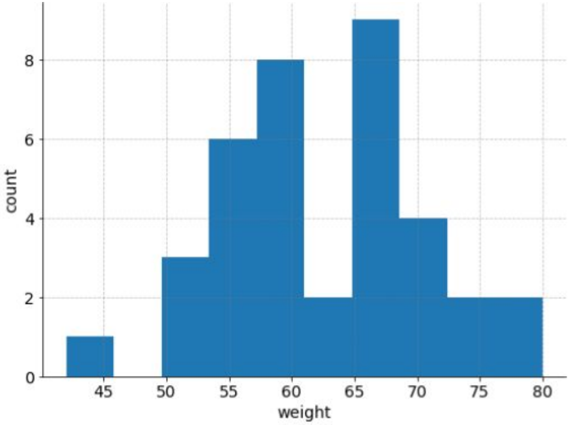


Let's get the data you filled in

- Download the CSV file
- Read CSV file into python - jupyter notebook
    - Data cleaning - remove bad values
    - Plot the data: We got 36 3-D data points

Weight:

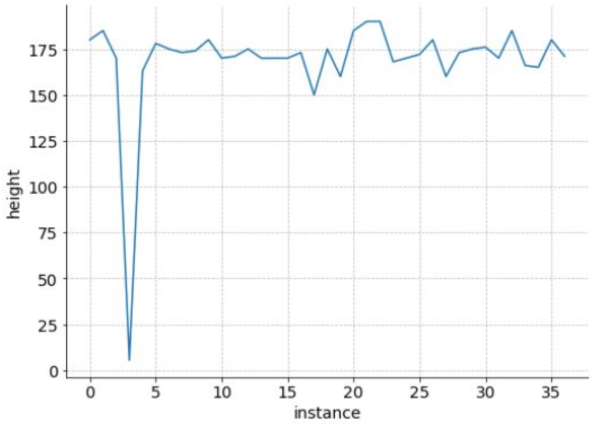


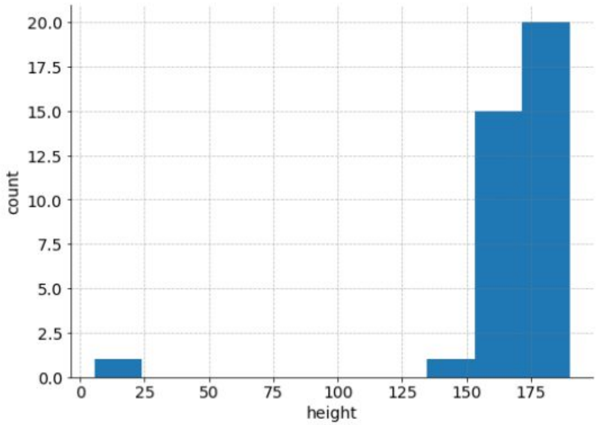


Let’s get the data you filled in

- Download the CSV file
- Read CSV file into python - jupyter notebook
  - Data cleaning - remove bad values
  - Plot the data: We got 36 3-D data points

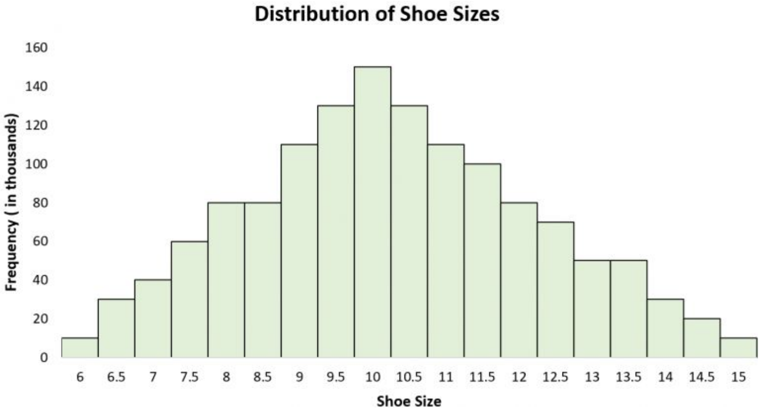
Height:

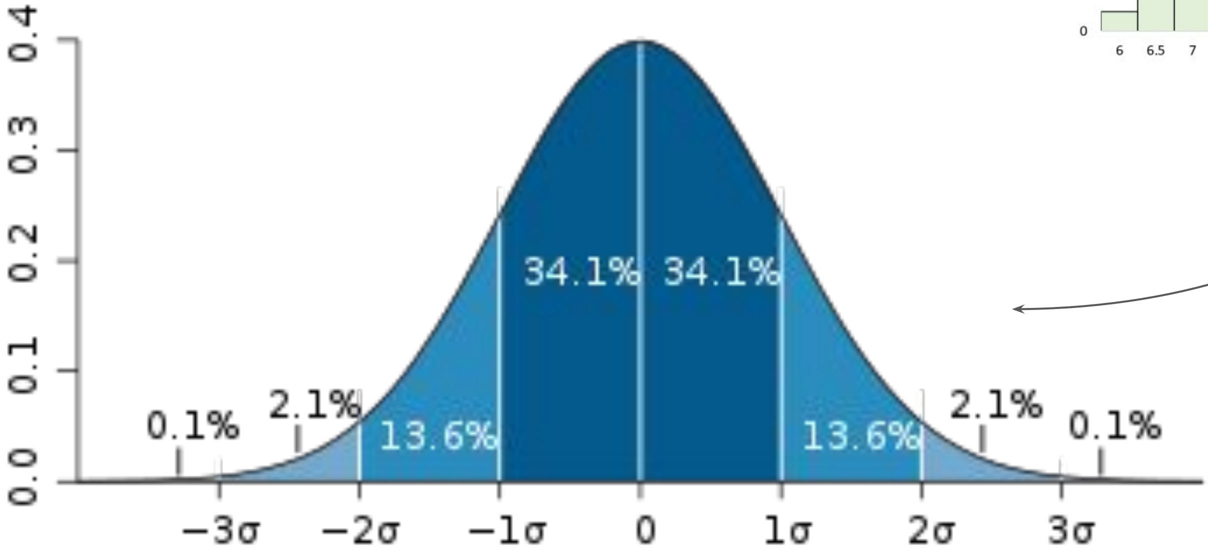




“Normal” histogram

- For scalar data points (1-D)
- Two parameters - mean and variance





A lot of “natural” data has a Normal distribution

- Why
- Central limit theorem (you will learn later)

“Normal” Distribution

- Key assumption:
  - Ideally a 1-D data should be normally distributed
  - Something happens (a process/transformation)
  - Mean shifts, Variance scales

Mean and variance normalization or standardization of x

$$Z = \frac{x - \mu}{\sigma}$$

- Can this assumption be generalized to higher dimensions
- Linear Algebra helps model the transformation as “linear” for N-dimensional data points

“Normal” Distribution

- Things are normal
- Something happens (a process/transformation)
- Mean shifts, Variance scales

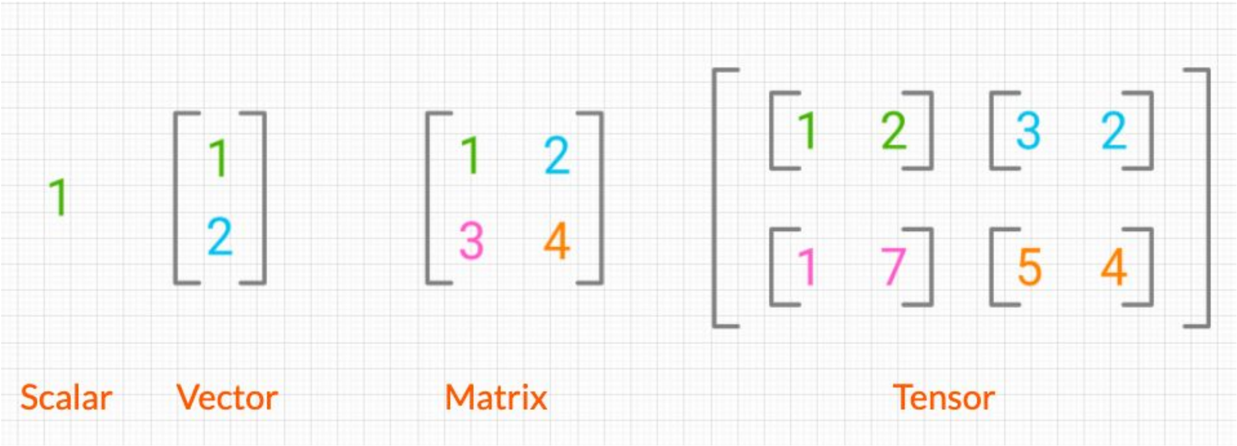
Normalization

- Make zero mean, and unit variance
- Or, mean and variance are features of **a scalar**

Ingredients of LA … going beyond scalars

- Scalars
- Vectors
- Matrices
- Tensors

A computer science view



Source: Harshit Tyagi, Towards Data Science blog post illustration (link)

Ingredients of LA … going beyond scalars

A physics view

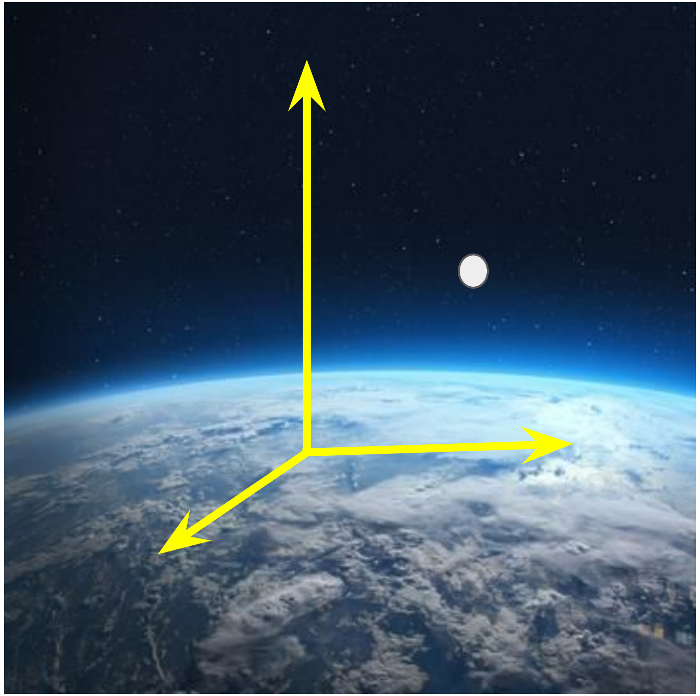
- Scalars
- Vectors
- Matrices
- Tensors

Source: Frontiers in Astronomy and Space Sciences

Ingredients of LA

- Scalars
- Vectors
- Matrices
- Tensors

A physicist view

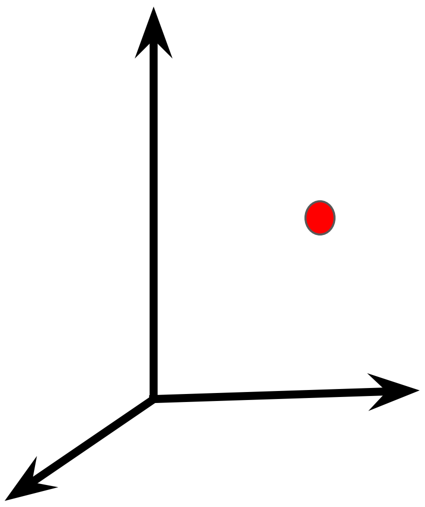


Source: Frontiers in Astronomy and Space Sciences

Ingredients of LA

- Scalars
- Vectors
- Matrices
- Tensors

A physicist view

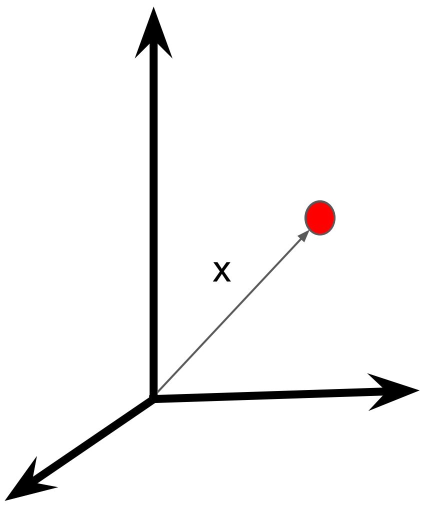


Source: Frontiers in Astronomy and Space Sciences

Vectors

- Direction
- Magnitude

(discuss)

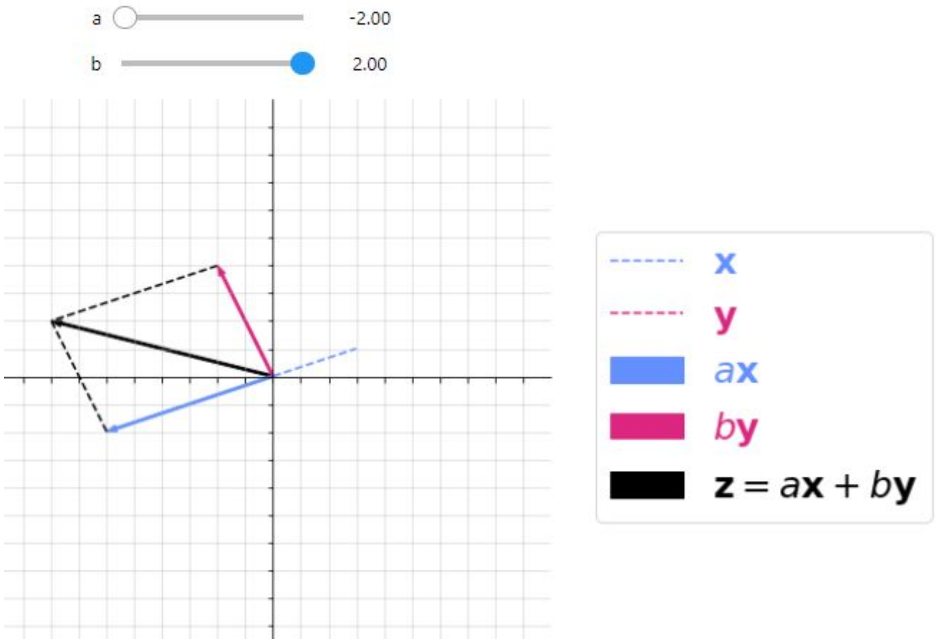


Source: Frontiers in Astronomy and Space Sciences

Vectors

- Length
- Direction
- Addition
- Subtraction
- Dot product
- Span
- Linearly independent
- Basis

(discuss)



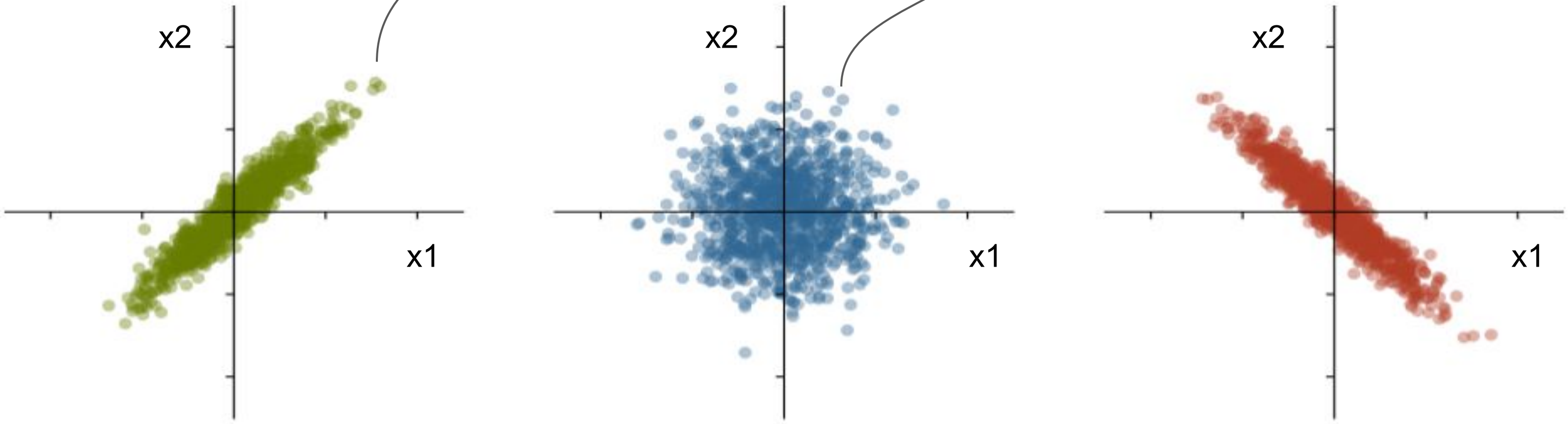
Jupyter notebook

Vectors

Given ‘x1’ you have some idea about ‘x2’

Given ‘x1’ you have no idea about ‘x2’

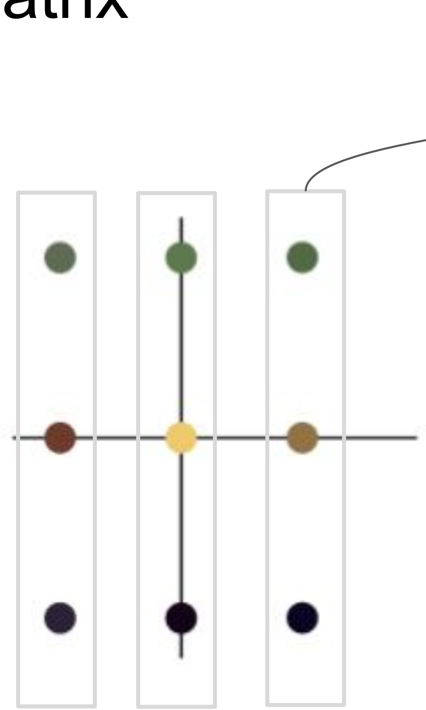
Correlation in 2-D



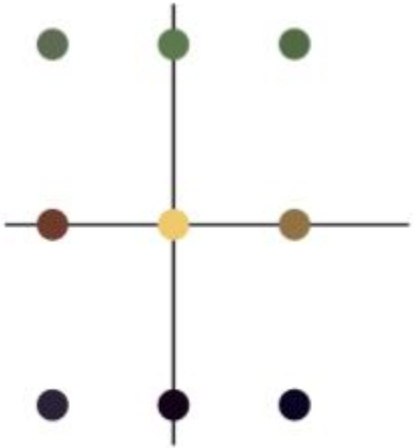
positive correlation	no correlation	negative correlation
$\begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix}$

Source: Peter Bloem https://peterbloem.nl/blog/pca-2

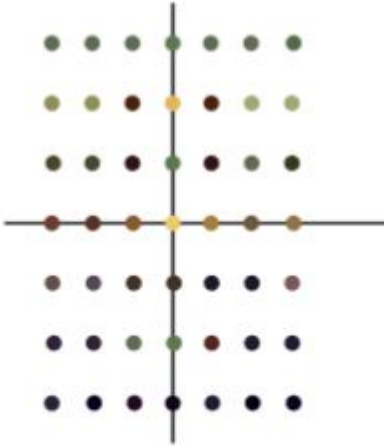
Matrix

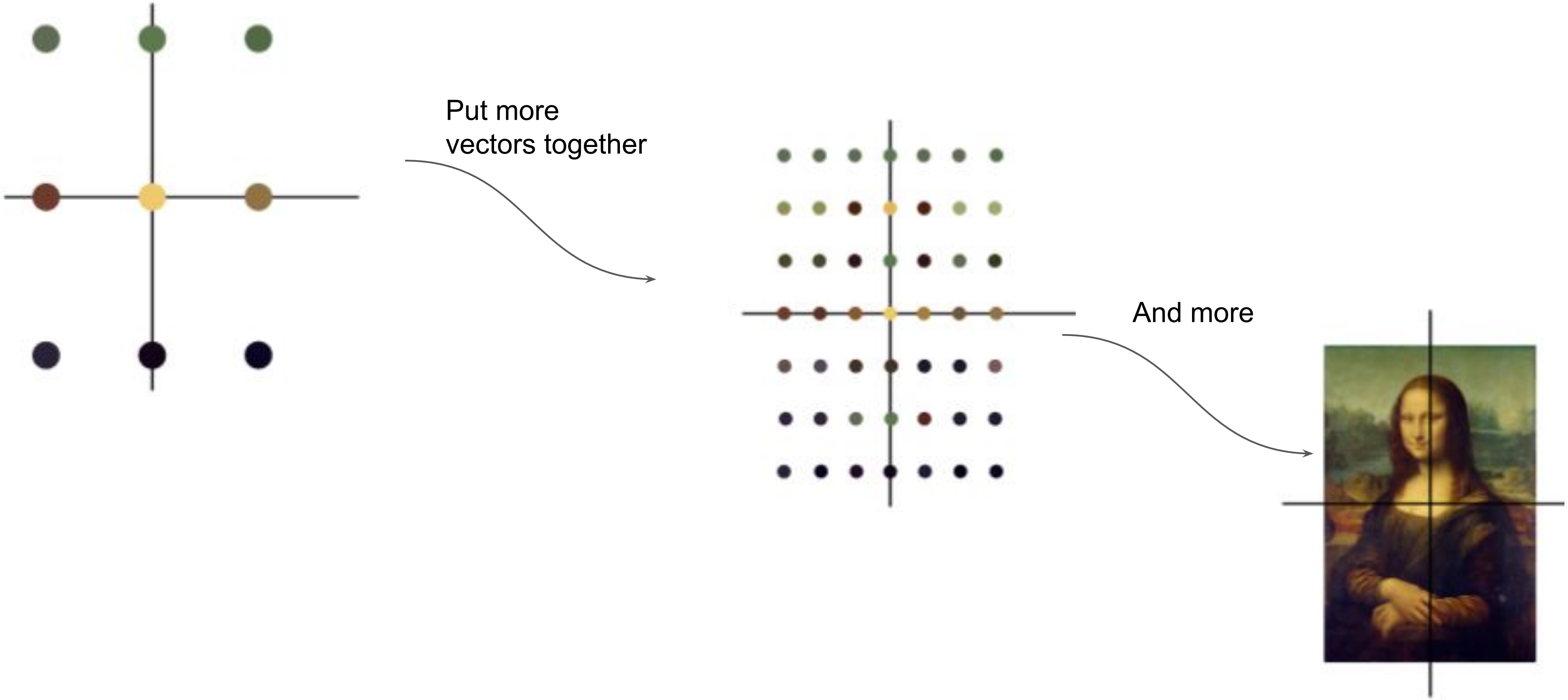


Source: Adapted from Peter Bloem https://peterbloem.nl/blog/pca-2



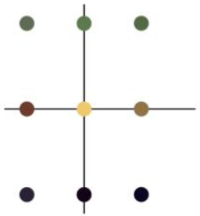
Put more
vectors together

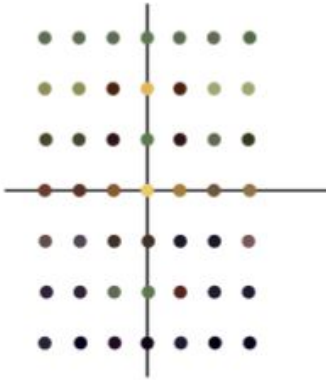




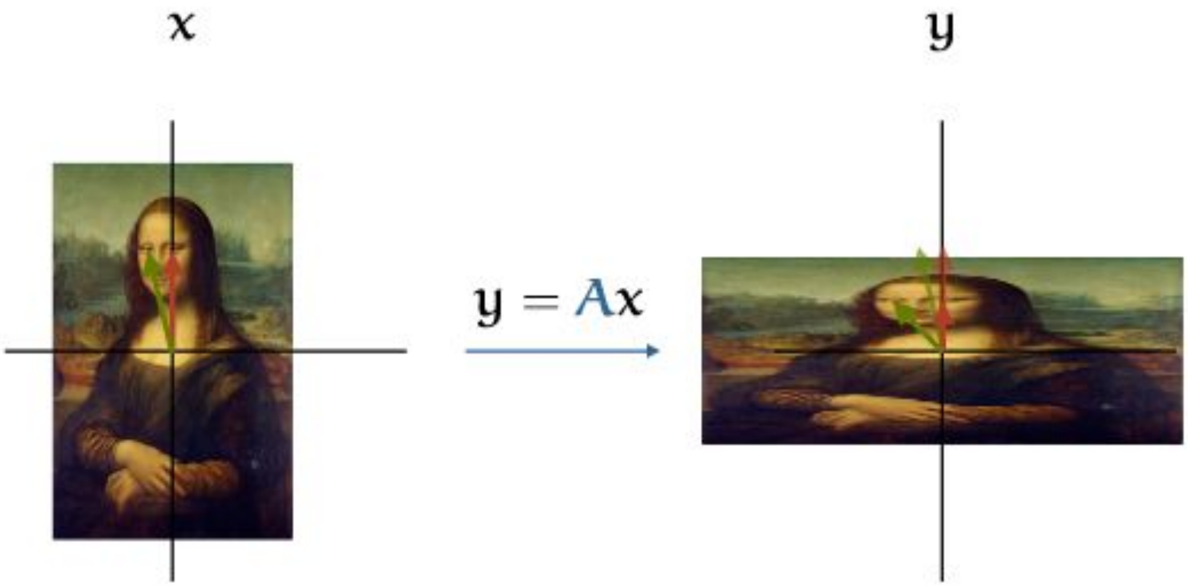
Put more
vectors together
And more

Image Matrix is a collection of vectors!





Scaling an image → a transformation



Source: Peter Bloem https://peterbloem.nl/blog/pca-2

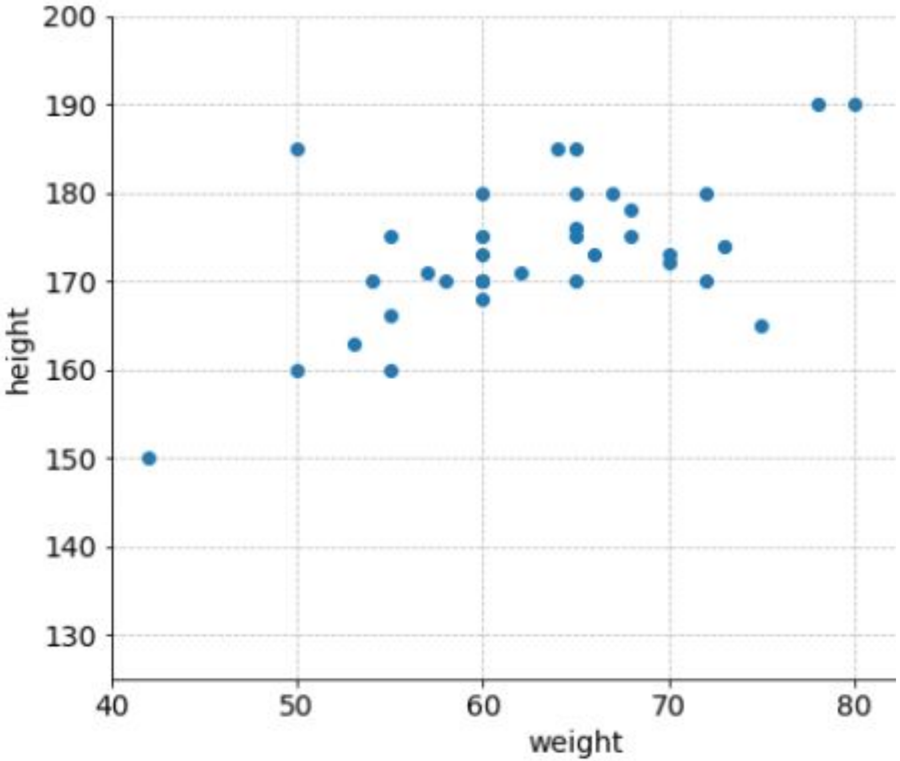
Summary

- Collected some data
- Plotted it
- Understood a data view of scalar, vector, matrix

Tomorrow

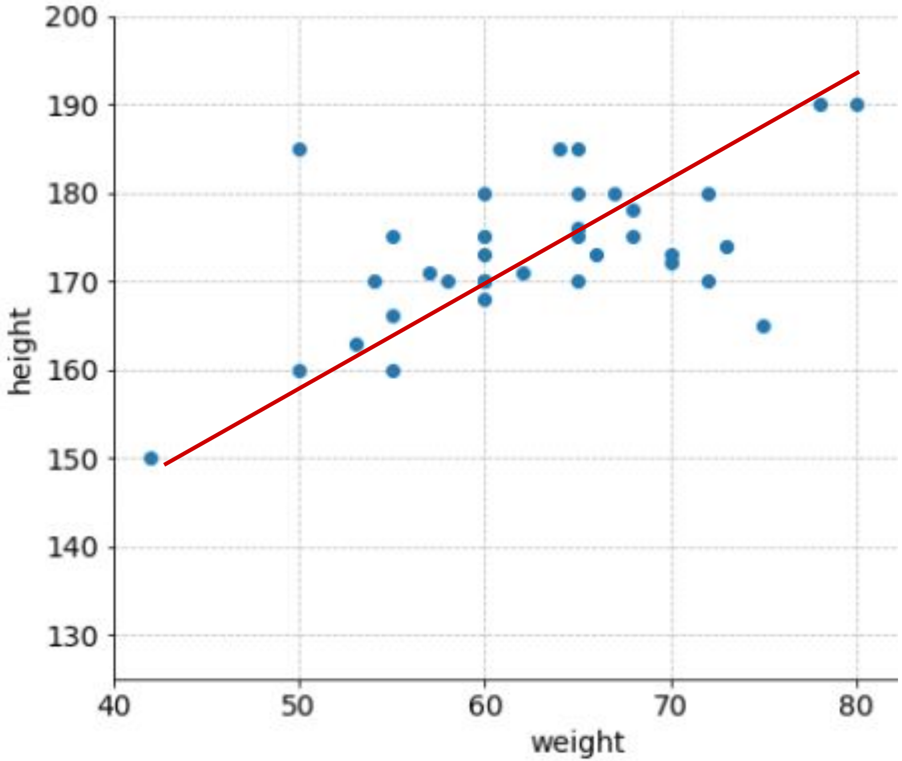
- Principal Component Analysis

Linear relations



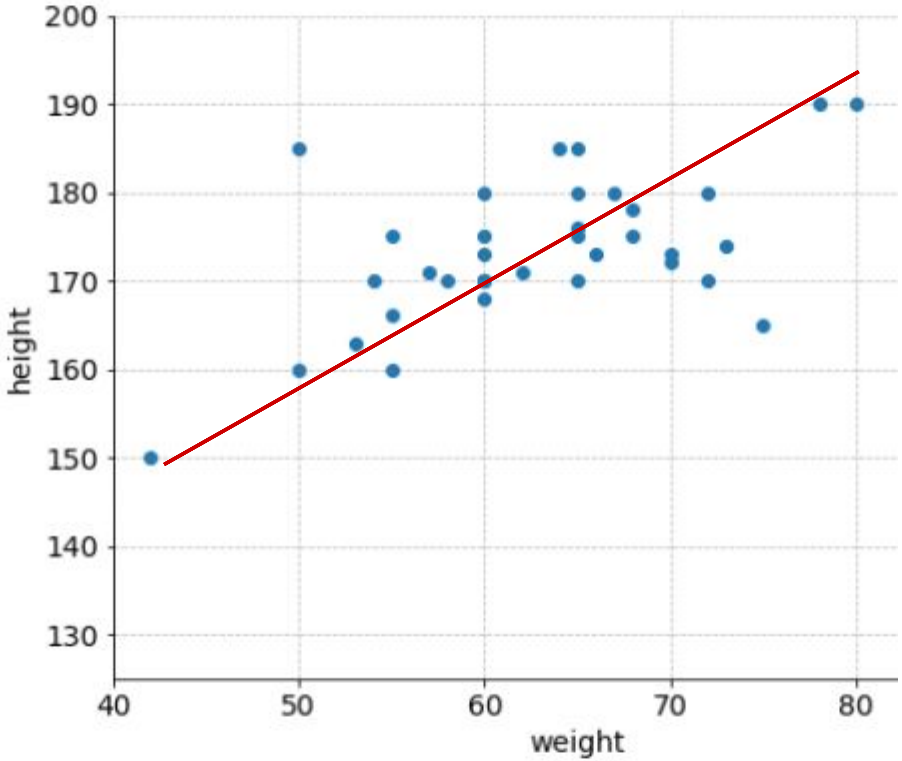
Linear relations

- Why?
  - “Looks” linear
  - May be easy to do
  - We know how to do



Linear relations

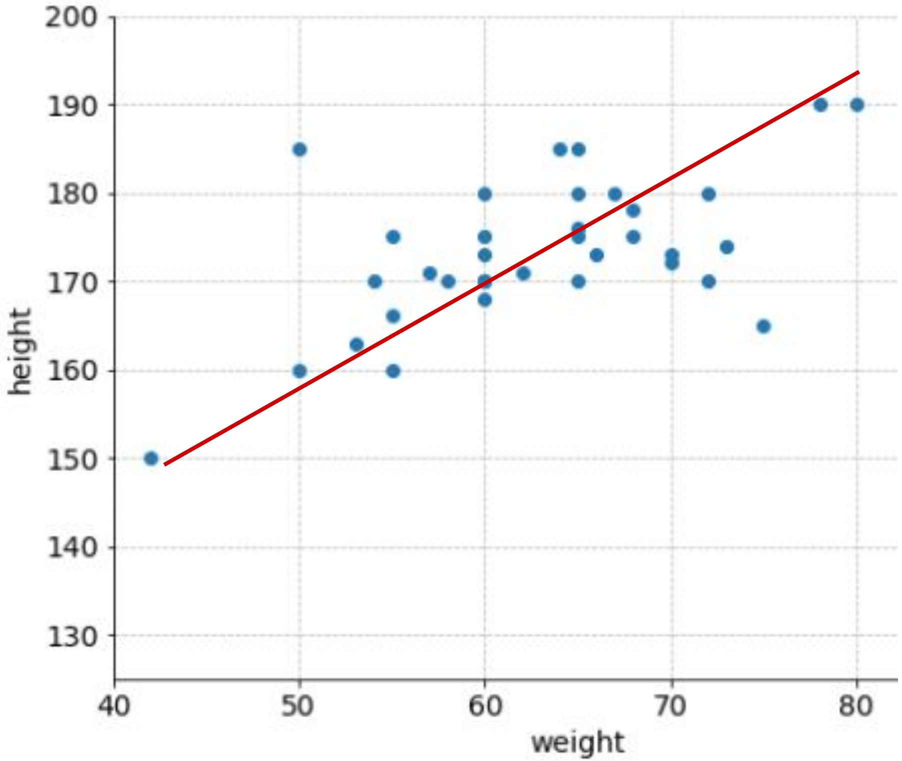
- How to do?
  - Write a linear relation or transformation
  - Estimate the parameter
    - Test the accuracy



Linear relations

- How to do?
  - Write a linear relation or transformation
  - Estimate the parameter
    - Test the accuracy

Welcome to One-dimensional PCA!

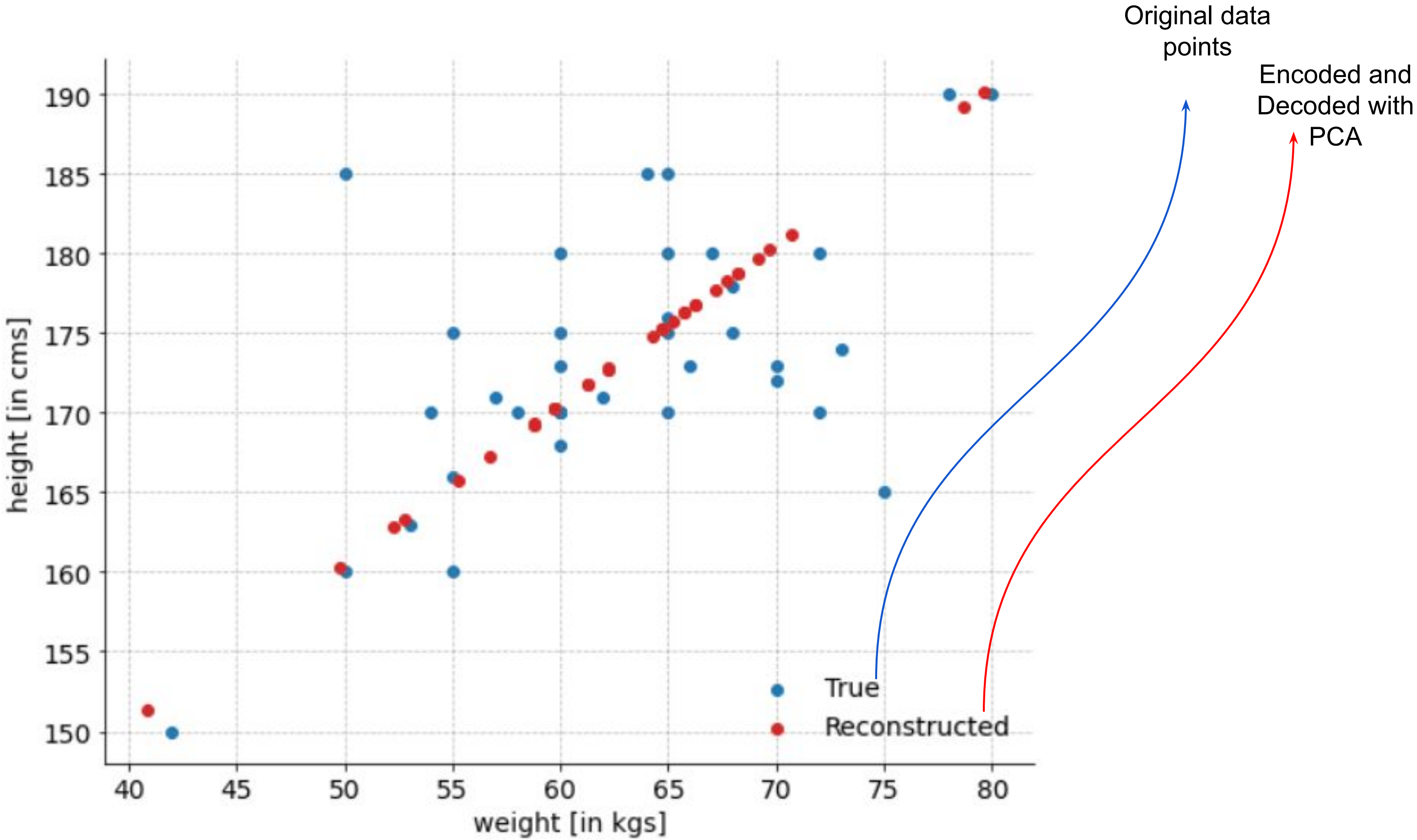


Principal Component Analysis (PCA)

- Take 2-D data points
- Project to 1-D space (Encoding)
- Reconstruct back the 2-D data (Decoding)
- Goal:
  - The “error” between original and reconstructed data should be minimum
    - Mean Square Error
  - Subject to a few additional constraints on the encoding and decoding vectors
    - Encoding and Decoding vectors are related

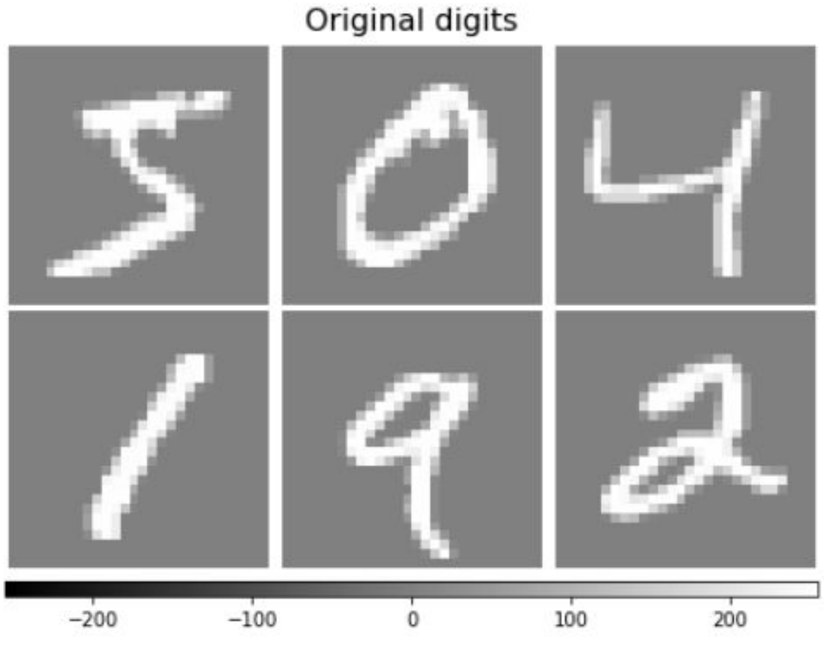
Discuss on board

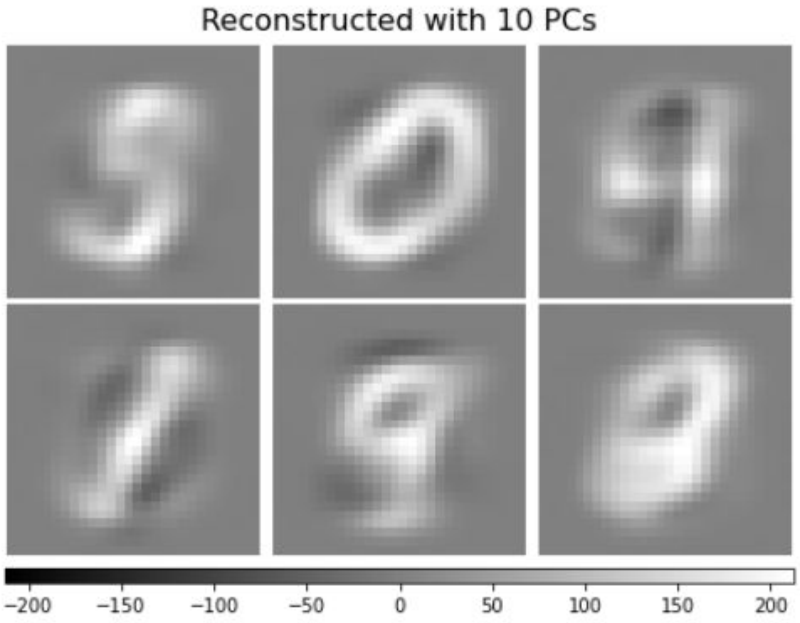
PCA: on our dataset



PCA: on MNIST Digit dataset

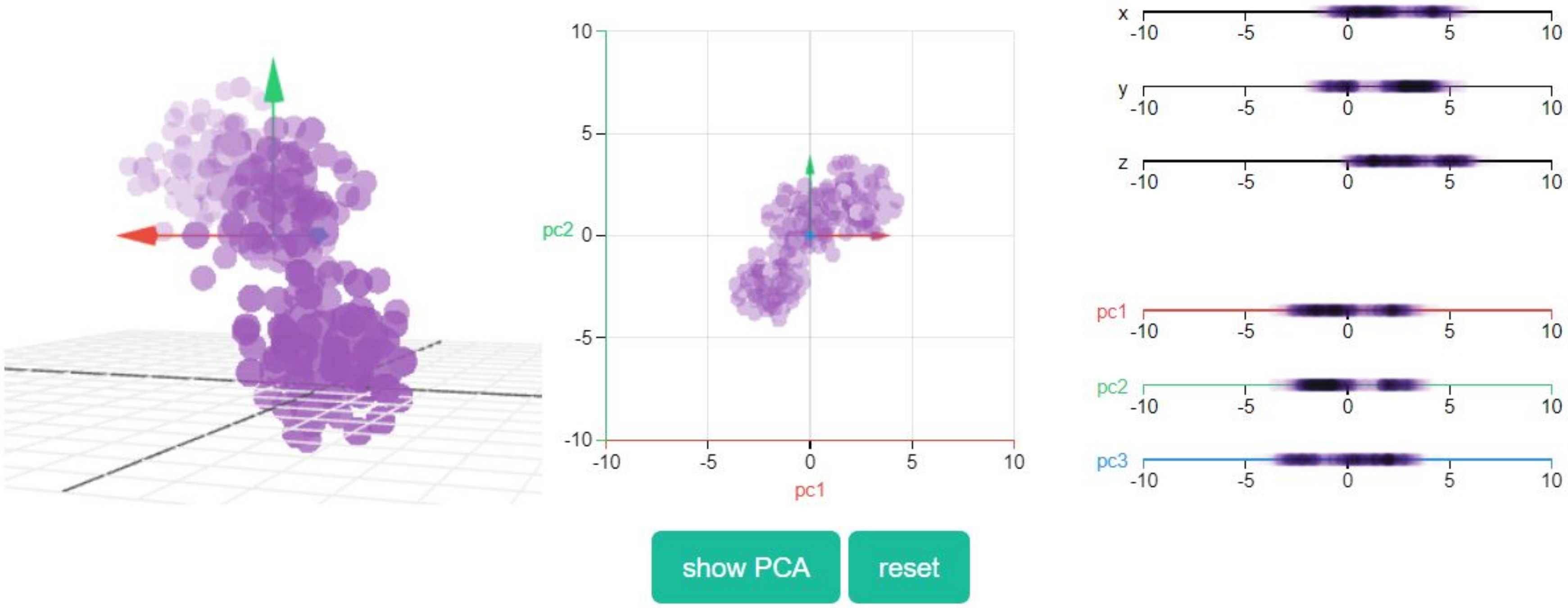
Do in Jupyter Notebook





Let's recap with an animation

https://setosa.io/ev/principal-component-analysis/



PCA: On Face images

- How does an “average” face look like?
- Are there some common features across faces?
  - Can we learn them mathematically?
- Can we reconstruct any face from these features?

PCA: On Face image dataset

5.6.1. The Olivetti faces dataset

This dataset contains a set of face images taken between April 1992 and April 1994 at AT&T Laboratories Cambridge. The `sklearn.datasets.fetch_olivetti_faces` function is the data fetching / caching function that downloads the data archive from AT&T.

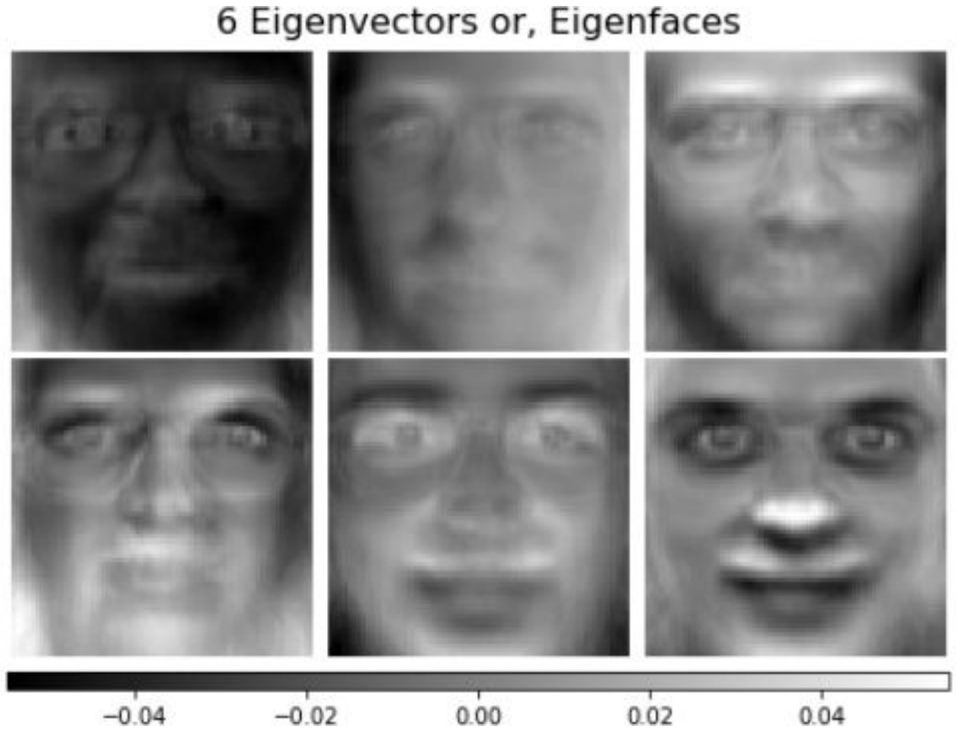
As described on the original website:

> There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

The image is quantized to 256 grey levels and stored as unsigned 8-bit integers; the loader will convert these to floating point values on the interval [0, 1], which are easier to work with for many algorithms.

Courtesy: Scikit-learn

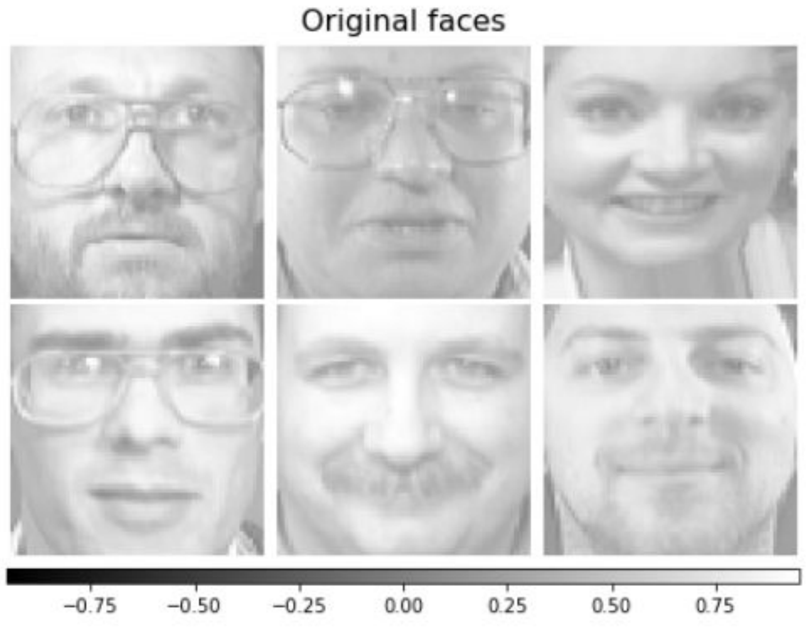
Let's go to Jupyter notebook

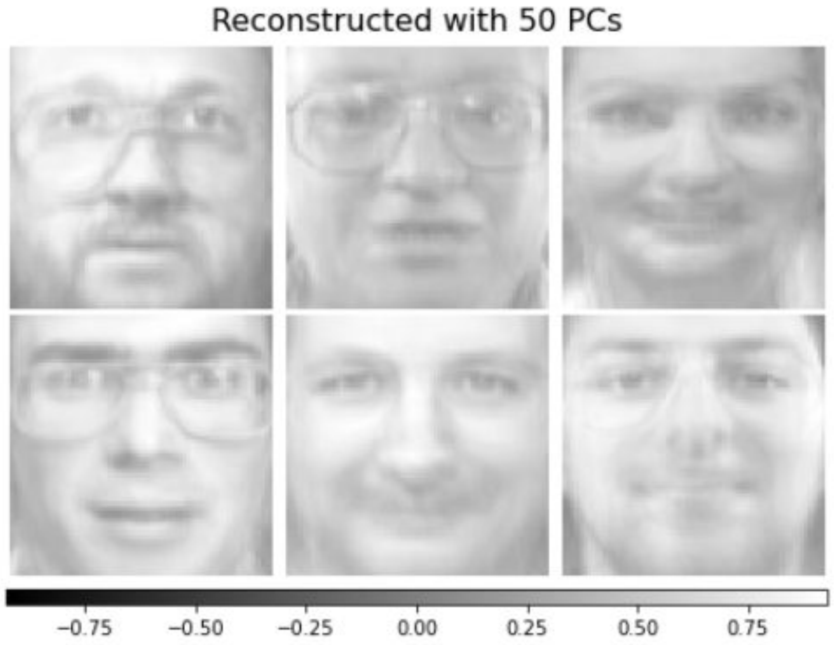


Do you notice some face-like attributes!

- Note we have plotted eigenvectors as images by reshaping.

Let’s go to Jupyter notebook

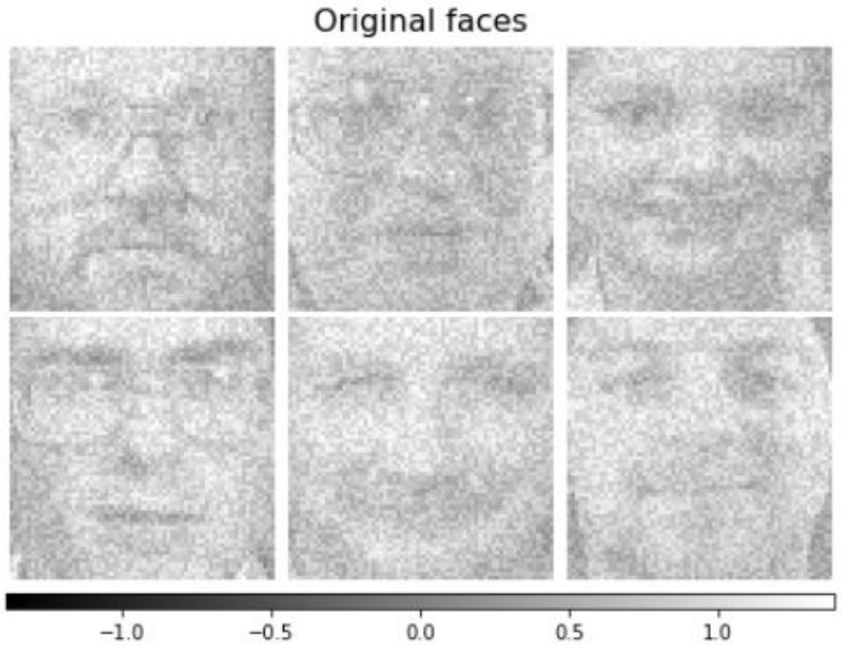


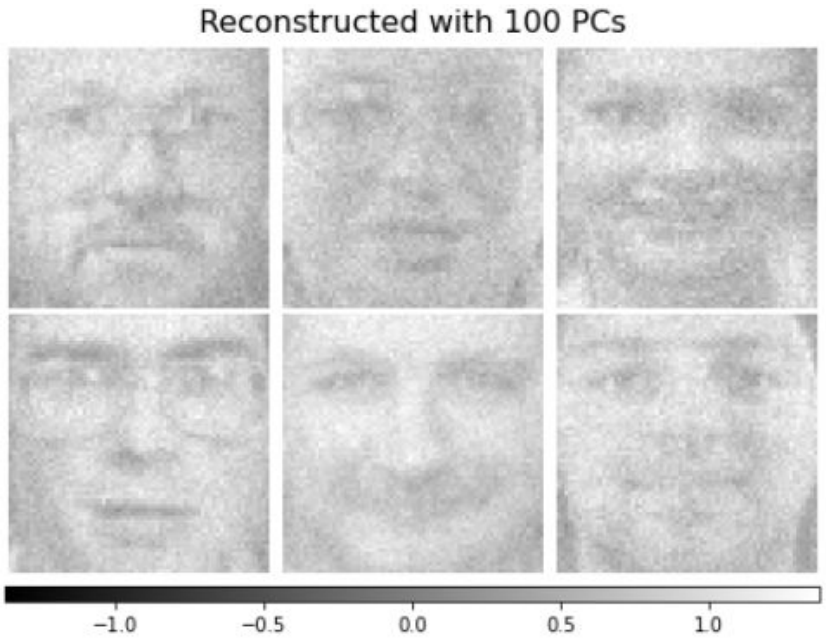


PCA: On Noisy face images

Denoising using a few of PCs for reconstruction

Let’s go to Jupyter notebook





Summary (Lecture 1,2,3)

- Ingredient of LA
  - scalars, vectors, matrices
  - Visualization and physical interpretation
  - Creating our own small dataset using responses from google form
- Correlations, Covariance matrix, and eigenvectors
- Principal Component Analysis (PCA)
  - Our own mini height-weight dataset
  - MNIST digit dataset
  - Olivetti Face image dataset
  - De-noising using PCA

Some resources

- Visualizing linear algebra
    - Video lectures by Grant Sanderson: https://www.3blue1brown.com/topics/linear-algebra
- PCA
    - Tutorial by Peter Bloem: https://peterbloem.nl/blog/pca-2
    - Book: Chap 12 in Pattern Recognition and Machine Learning, Christopher Bishop
        - https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf
- Jupyter notebooks
    - MNIST digit PCA:
        - https://colab.research.google.com/drive/1QjkThSrnmQi9jLkvQ721zKeu5ILIWKZI?usp=sharing
    - Face Images PCA:
        - https://colab.research.google.com/drive/15nD-sv_rYYxBbcPXuLSIAyUy7P-WF1DO?usp=sharing

Thanks for letting me introduce you to PCA