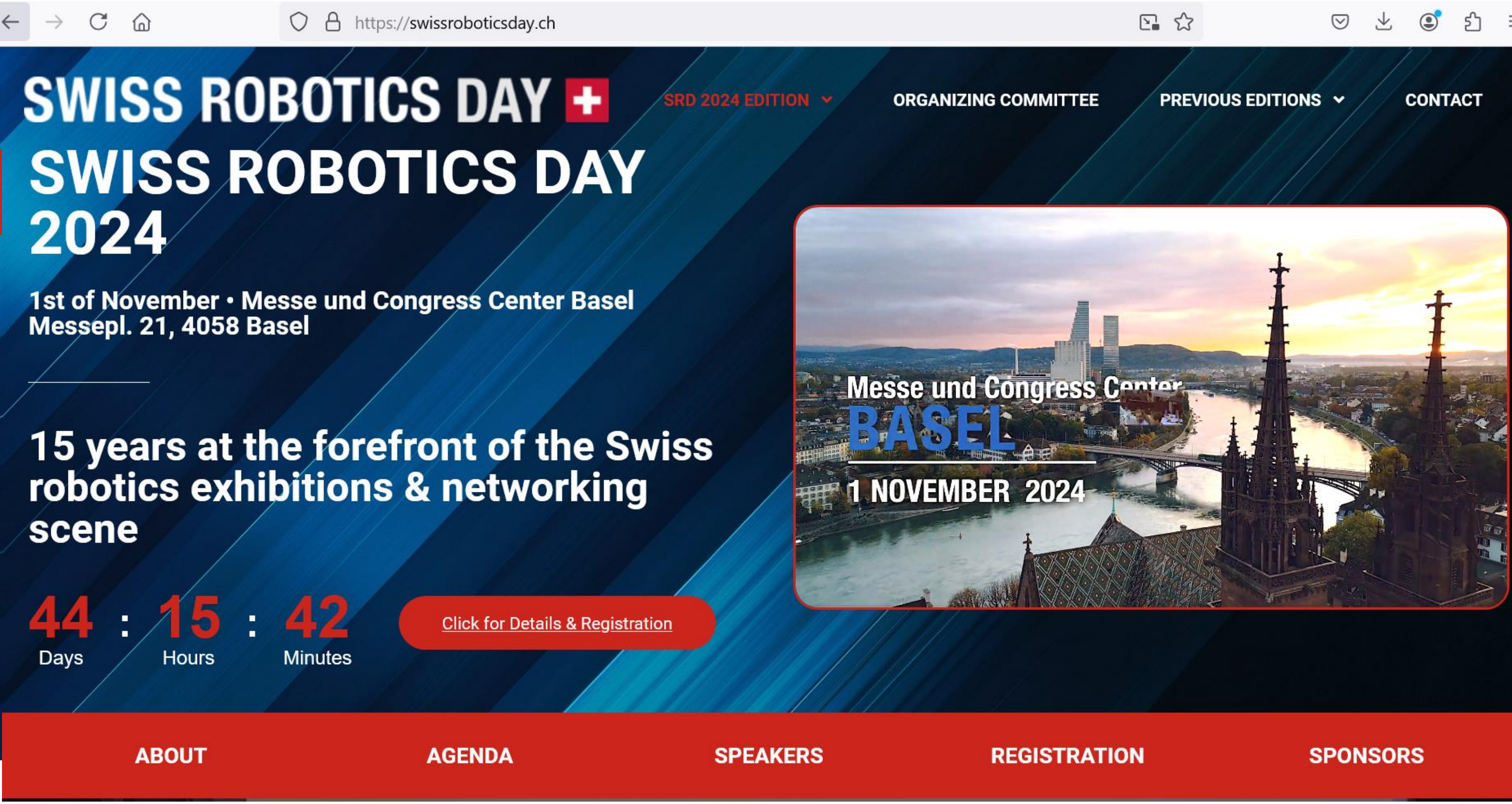


https://swissroboticsday.ch/

Discount for students on registration and train tickets

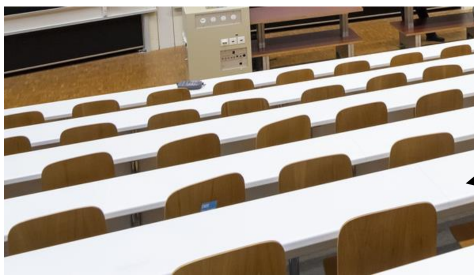
https://swissroboticsday.ch/

APPLIED MACHINE LEARNING

Principal Component Analysis (PCA)

Interactive exercise – lecture session

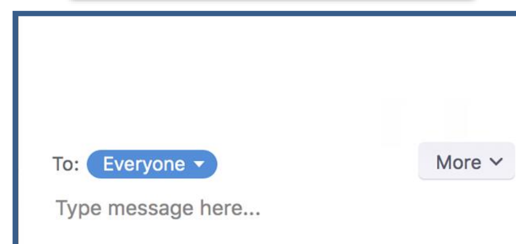
Interactions during interactive lecture



Ask live!

Type it in @ask-a-question text channel

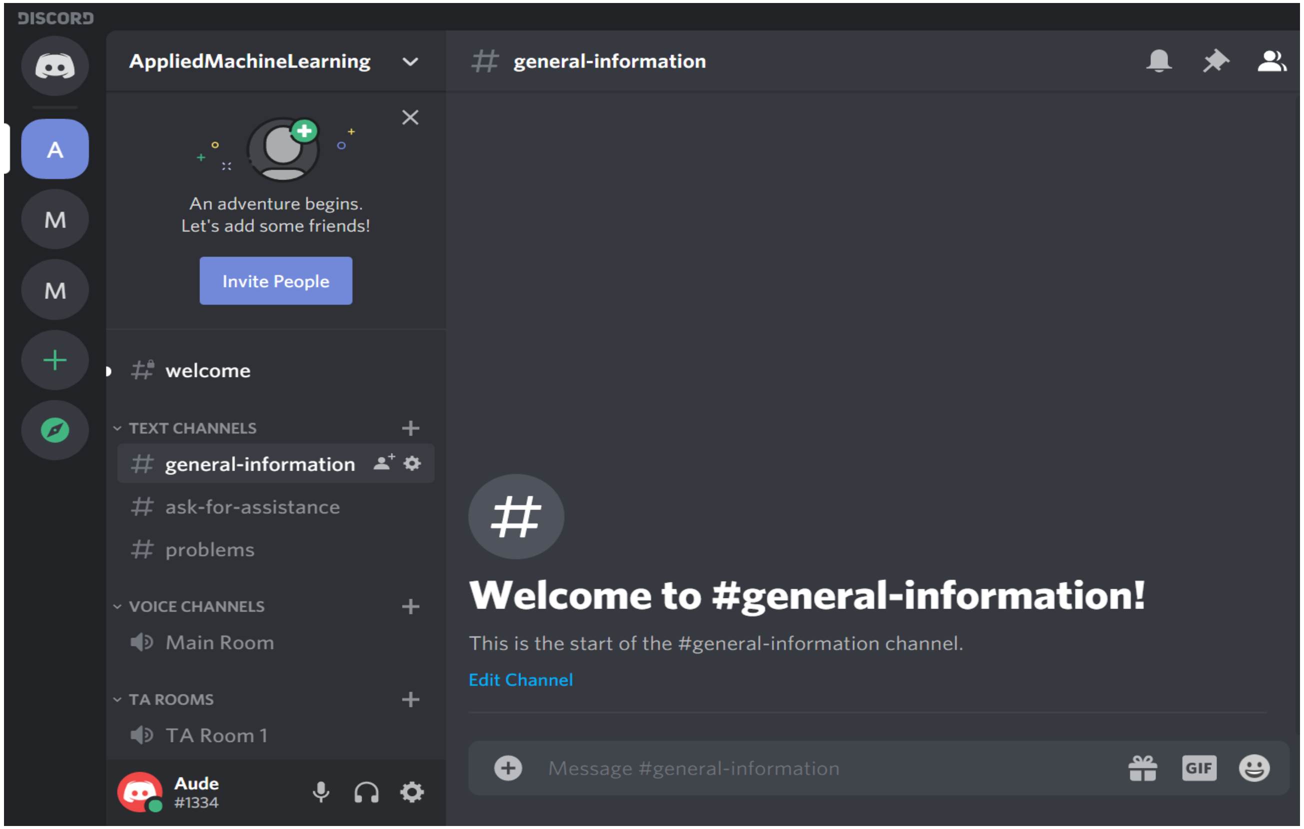
Type it in zoom chat box



Teaching assistants will reply live

Access to discord

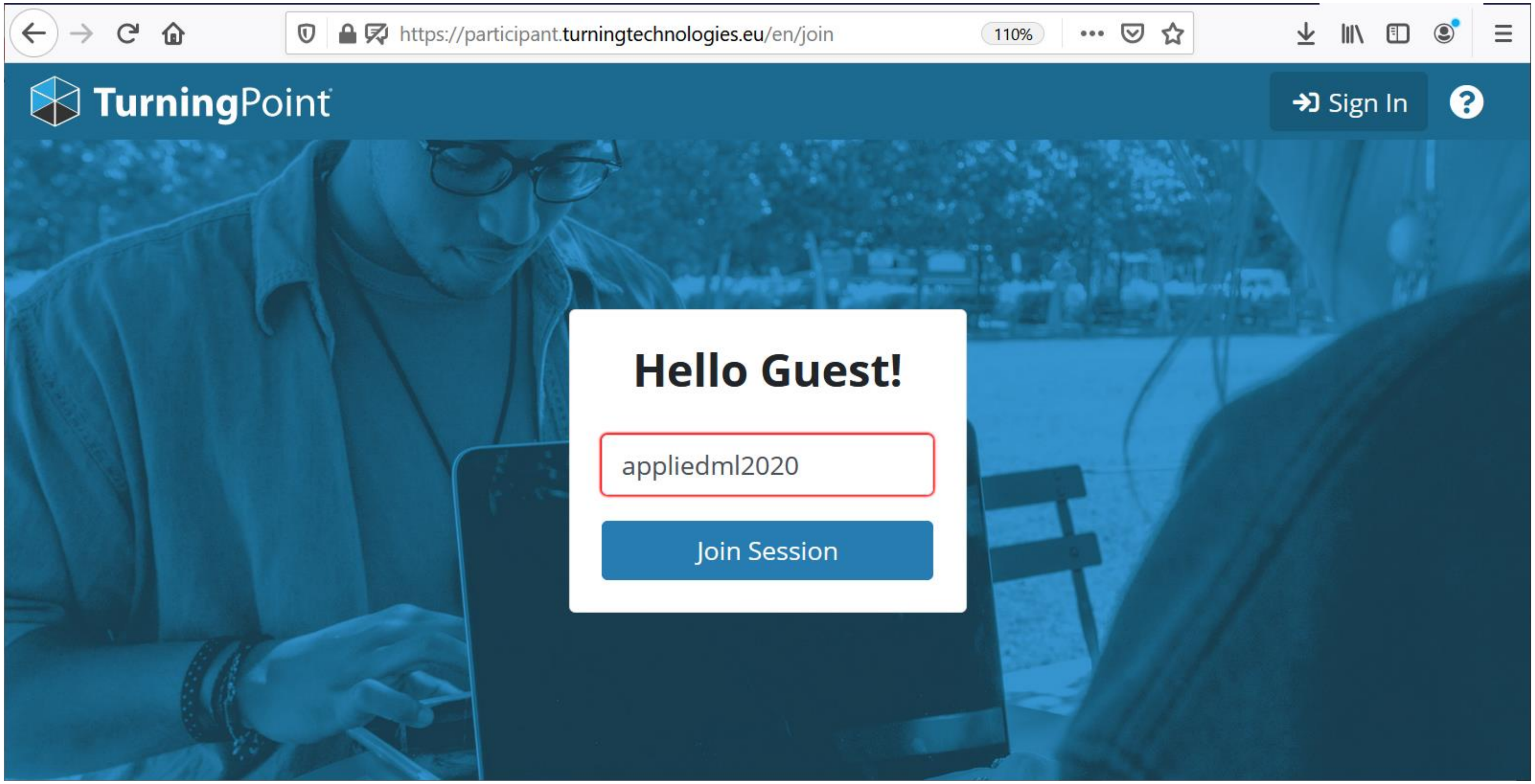
https://discord.gg/TwTfKkv



Launch polling system

https://participant.turningtechnologies.eu/en/join

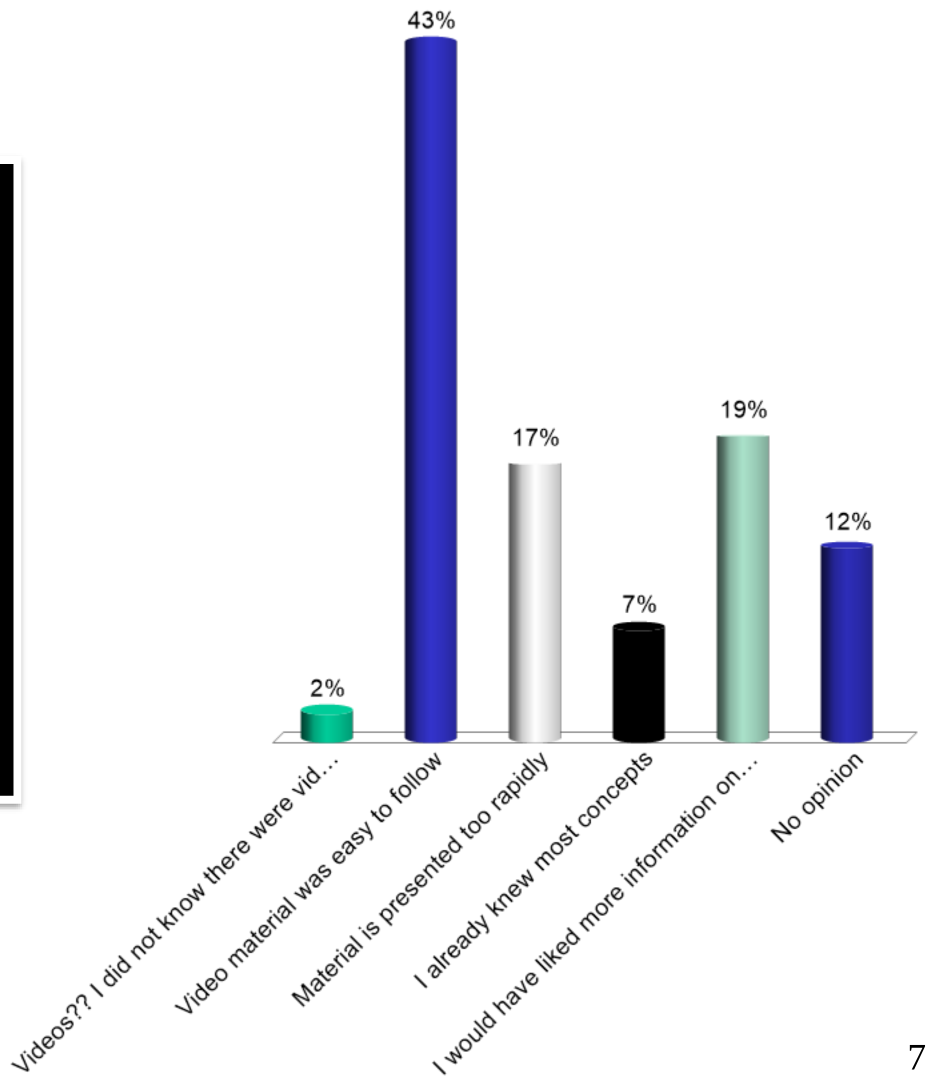
Acces as GUEST and enter the session id: *appliedml2020*



Qualify the content of the videos of the theoretical material of the class

Multiple answers possible

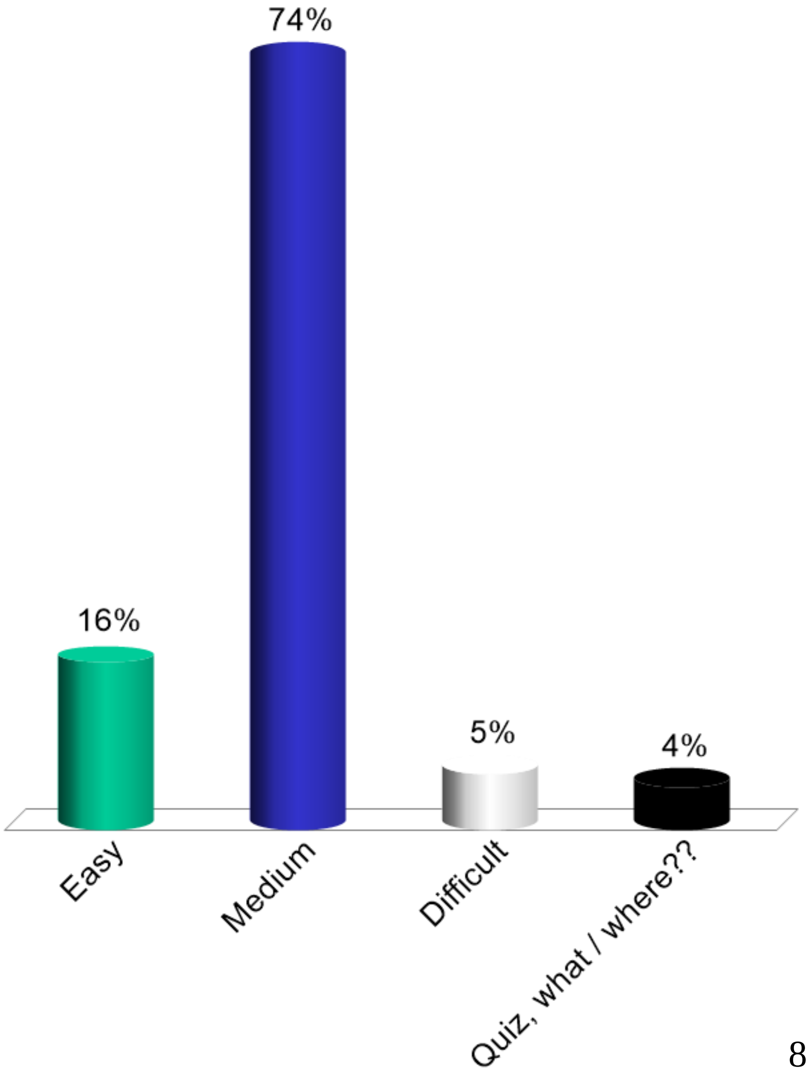
A. Videos?? I did not know there were videos!
B. Video material was easy to follow
C. Material is presented too rapidly
D. I already knew most concepts
E. I would have liked more information on some topics
F. No opinion



Qualify the content of the PCA quiz

Multiple answers possible

A. Easy
B. Medium
C. Difficult
D. Quiz, what / where??



PCA – Key Concepts

PCA has two properties:

1. It reduces the dimensionality of the data.

2. It extracts features in the data.

To achieve 1 & 2, it uses existing correlation across datapoints.

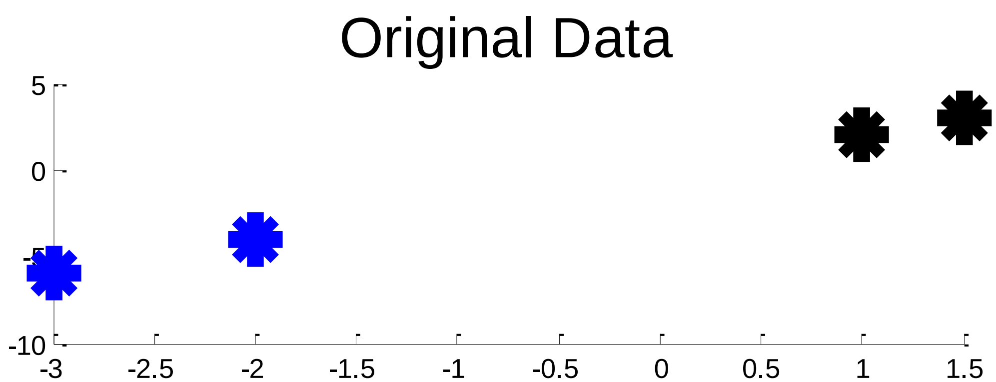
PCA can be used as:

1. Compression method for ease of data storage and retrieval.

2. Pre-processing method before classification to a) reduce computational costs, b) extract features to ease classifier's job.

Reducing the dimensionality by looking for correlations

PCA: Exercise 1 Reducing dimensionality of dataset



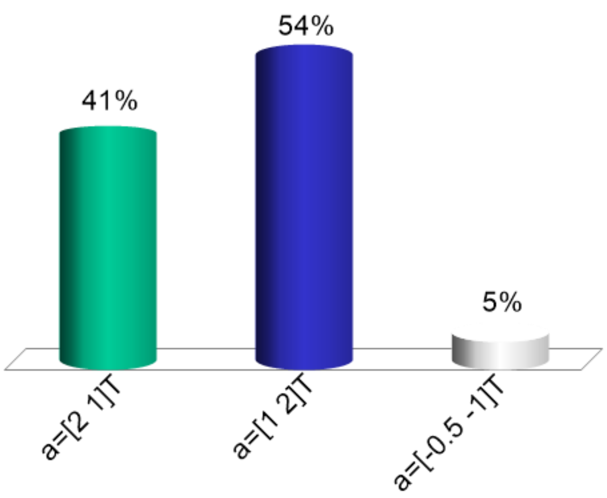
$$\text{Dataset } X = \begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2 & -4 & -6.2 & 3 \end{bmatrix}$$

Which of the projection vectors below minimizes reconstruction error?

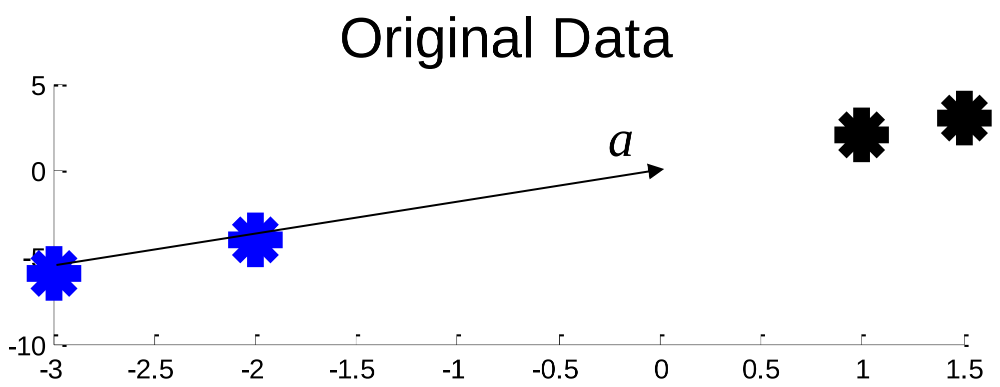
A. $a=[2\ 1]^T$

B. $a=[1\ 2]^T$

C. $a=[-0.5\ -1]^T$



PCA: Exercise 1 Reducing dimensionality of dataset

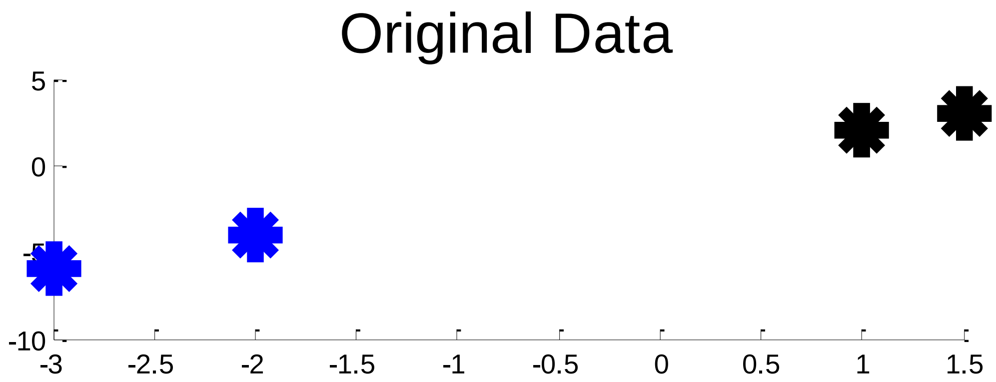


$$\text{Dataset } X = \begin{bmatrix} 1.1 & -2 & -2.9 & 1.5 \\ 2 & -4 & -6.2 & 3 \end{bmatrix}$$

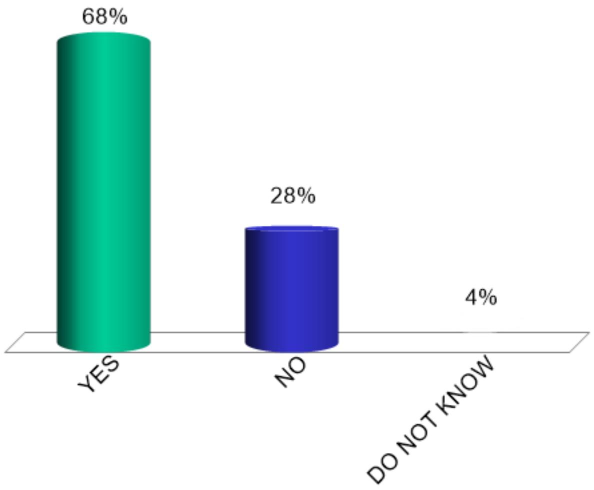
Which of the projection vectors below minimizes reconstruction error?

The projection vector $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

PCA: Exercise 2 Preprocessing for classification

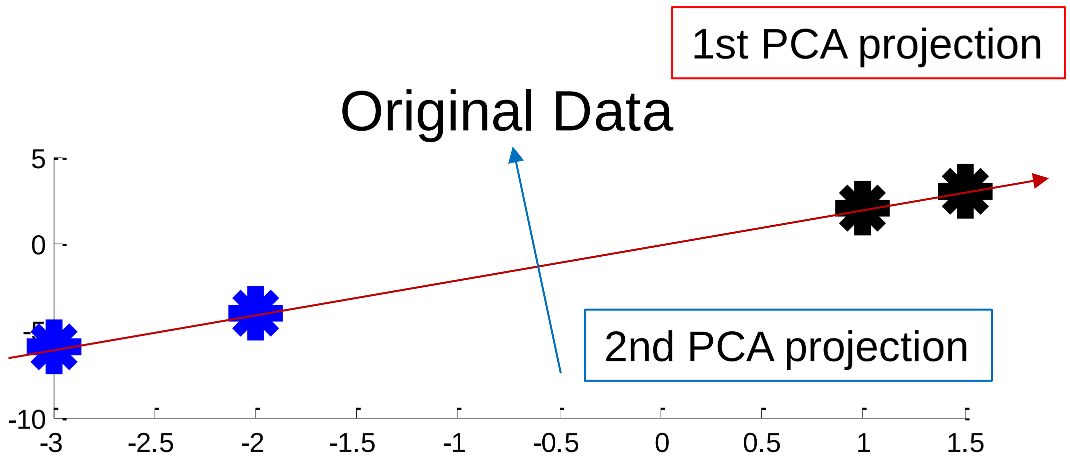


Are the two groups of points still separated once projected onto the first PCA projection (eigenvector with largest eigenvalue)?



A. YES
B. NO
C. DO NOT KNOW

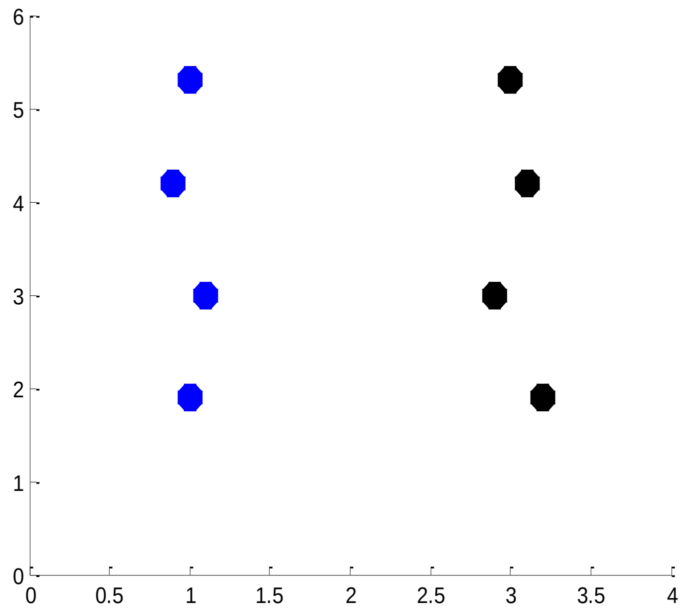
PCA: Exercise 2 Preprocessing for classification



$$\text{Dataset } X = \begin{bmatrix} 1 & -2 & -3 & 1.5 \\ 2 & -4 & -6 & 3 \end{bmatrix}$$

The first PCA projection allows to separate
the two groups of datapoints with a cut-off midway.

PCA: Exercise 3 Preprocessing for classification

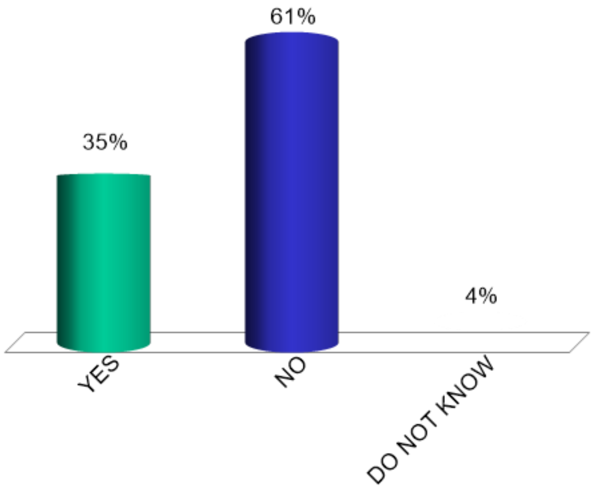


Are the two groups of points still separated once projected onto the first PCA projection (eigenvector with largest eigenvalue)?

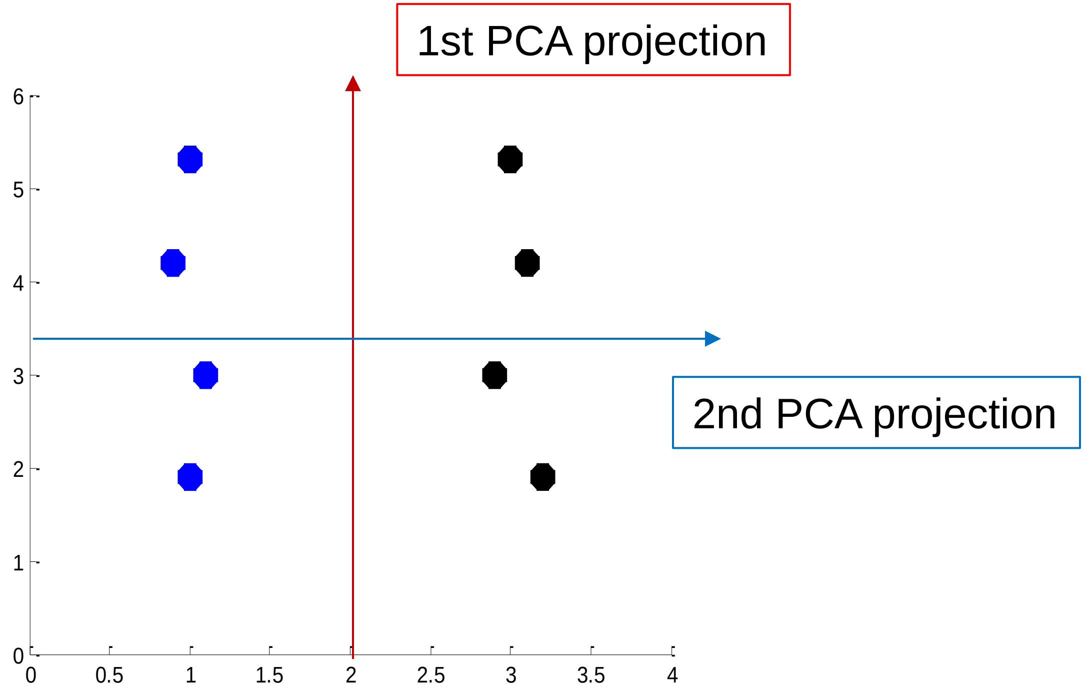
A. YES

B. NO

C. DO NOT KNOW



PCA: Exercise 3 Pre-processing prior to classification



- The first projection does not separate the data but rather merge them – the two classes get superimposed.
- On the second projection, the groups are separated.

PCA does not seek projections that make data more separable! However, among the projections, some may make data more separable.

APPLIED MACHINE LEARNING

PCA as a method to find features in data

PCA on images

23 images, each of which is of dimension 230400.

$x \in \mathbb{R}^{320 \cdot 240 \cdot 3 = 230400}$

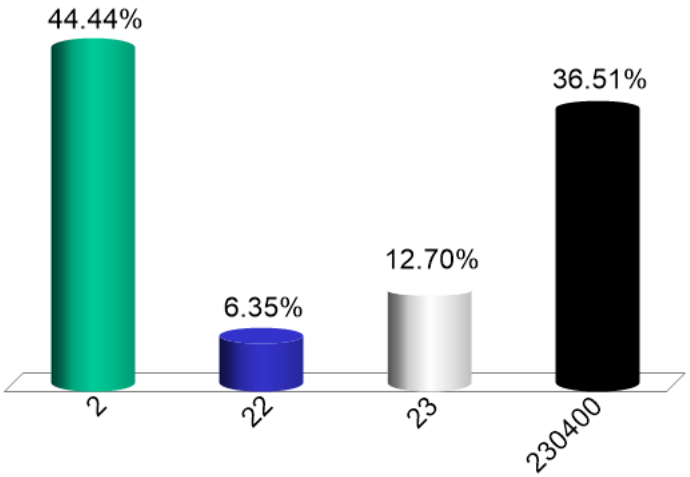
How many eigenvectors do you obtain after PCA?

A. 2

B. 22

C. 23

D. 230400



PCA on images

M=23 images, each of which is of dimension N=230400.

The covariance matrix is $C = XX^T$.

If X is $N \times M$, then C is $N \times N$.

$N = 230400$.

The eigendecomposition of C generates 230400 eigenvectors.

Only 22 of the 230400 eigenvectors are meaningful. All others have eigenvalue zero.

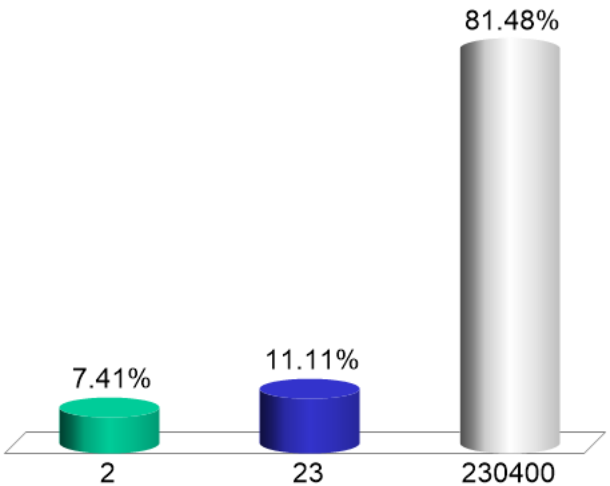
You may then choose to keep only 2 out of these 22 eigenvectors, as they may be sufficient for your task (e.g. classification).

PCA on images

M=23 images, each of which is of dimension N=230400.

What is the dimension of each eigenvector?

A. 2
B. 23
C. 230400



PCA on images

M=23 images, each of which is of dimension N=230400.

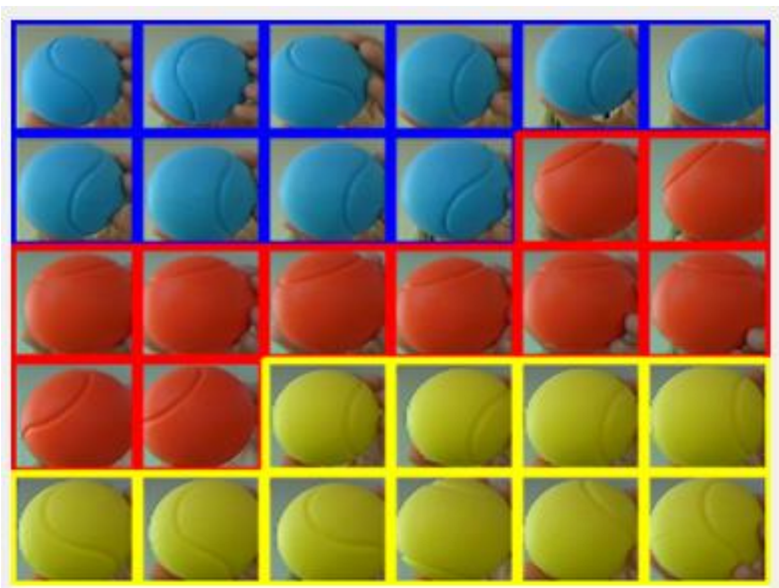
What is the dimension of each eigenvector?

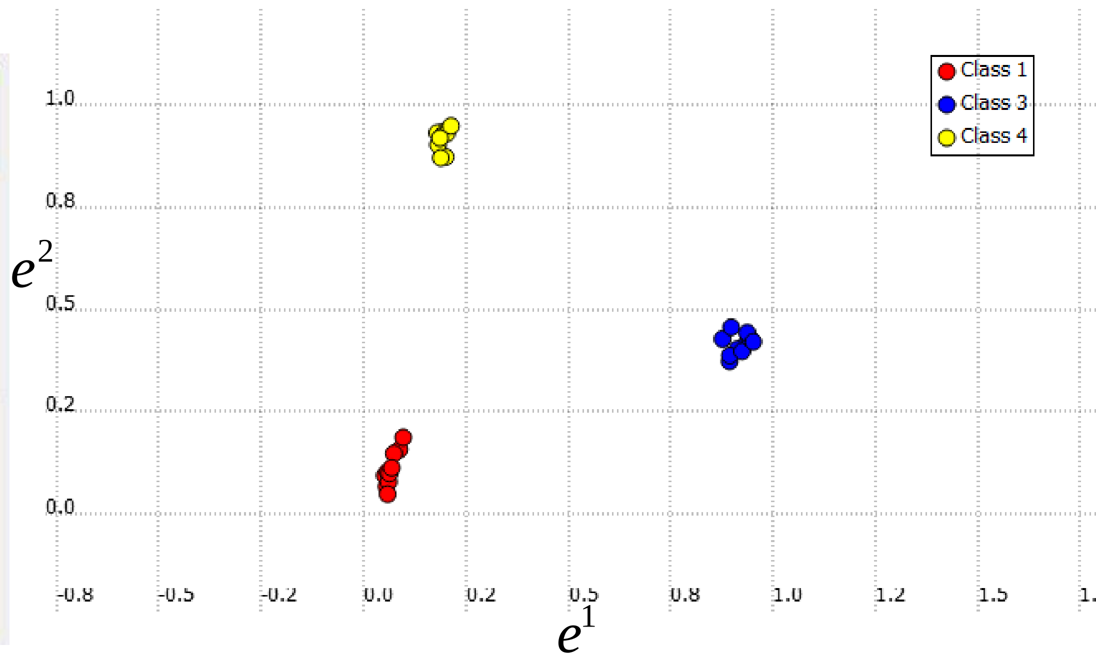
Each eigenvector is of the same dimension as the original images, i.e. N=230400.

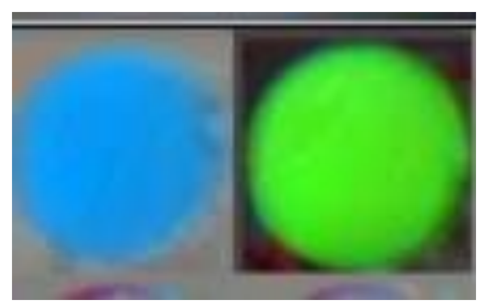
An eigenvector of a dataset of images is an image. Such an eigenvector is often referred to as *eigenface.*

Interpreting PCA projections and eigenvectors

Can you explain the “color” of the eigenvectors?



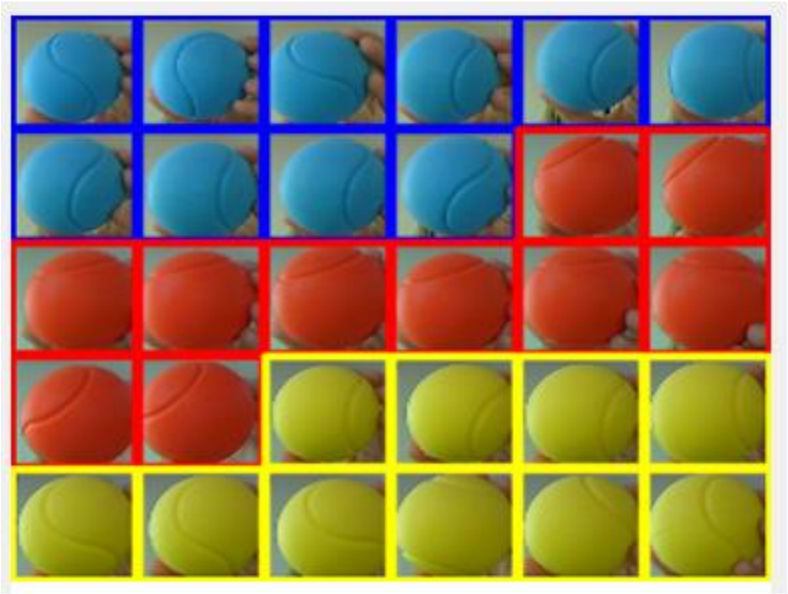


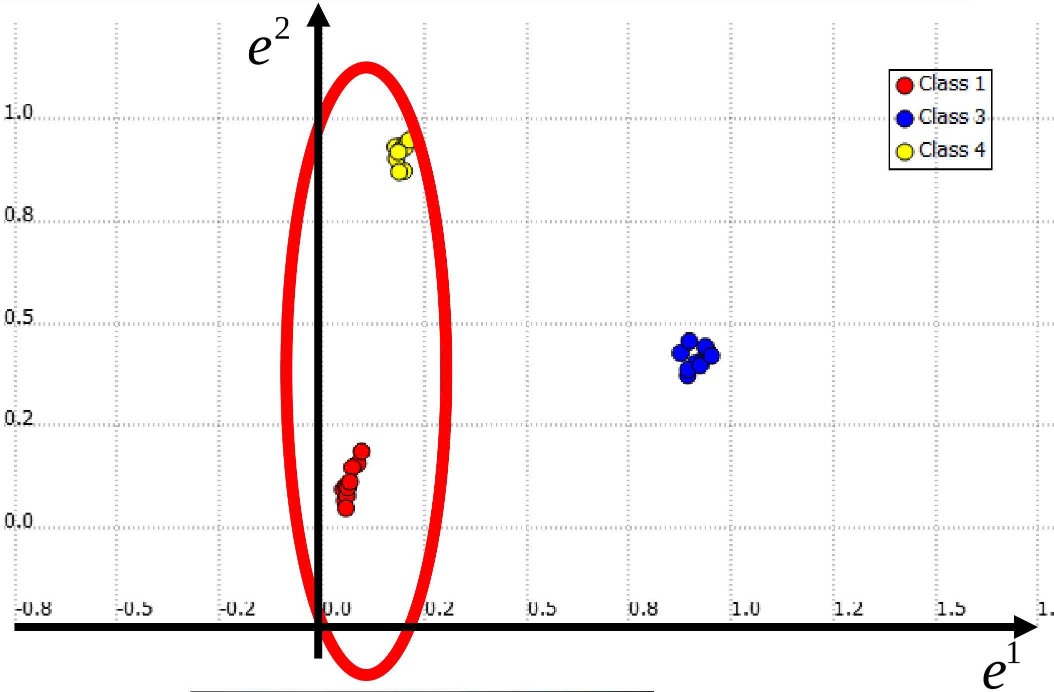


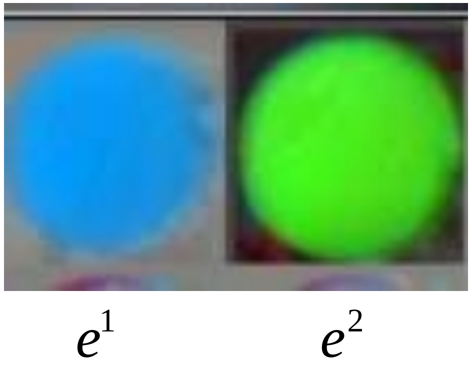
e^1 e^2

Interpreting PCA projections and eigenvectors

The red and yellow images have coordinate ~0 on the 1st eigenvector.



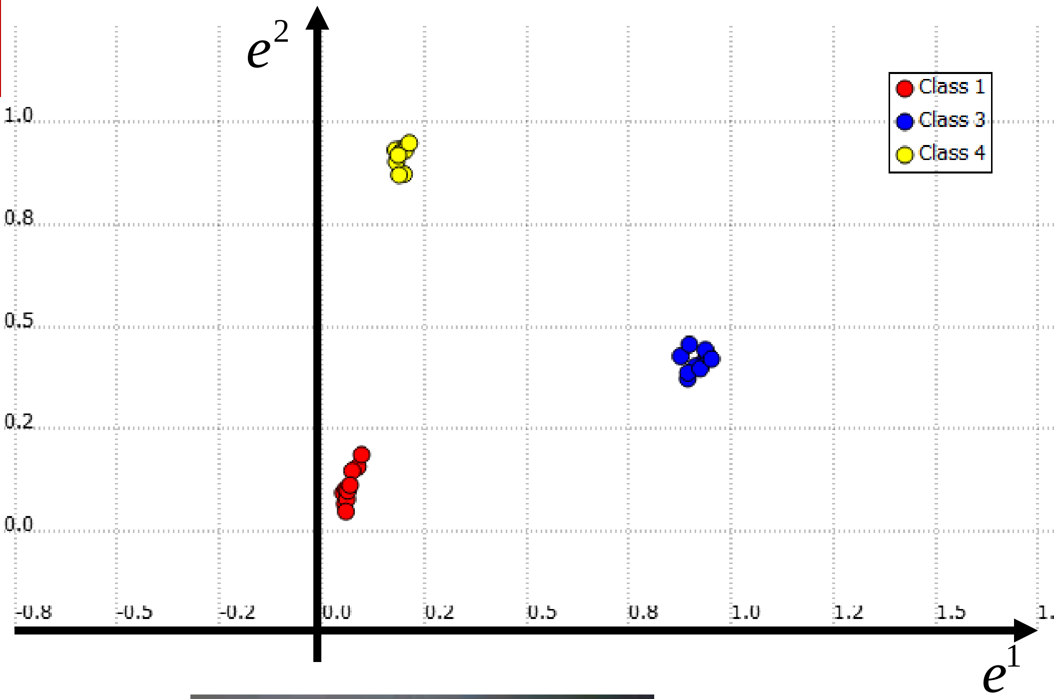




Interpreting PCA projections and eigenvectors

What do the entries of the eigenvectors look like?

$$e^1 = \begin{bmatrix} ? \\ ? \\ ? \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \qquad e^2 = \begin{bmatrix} ? \\ ? \\ ? \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

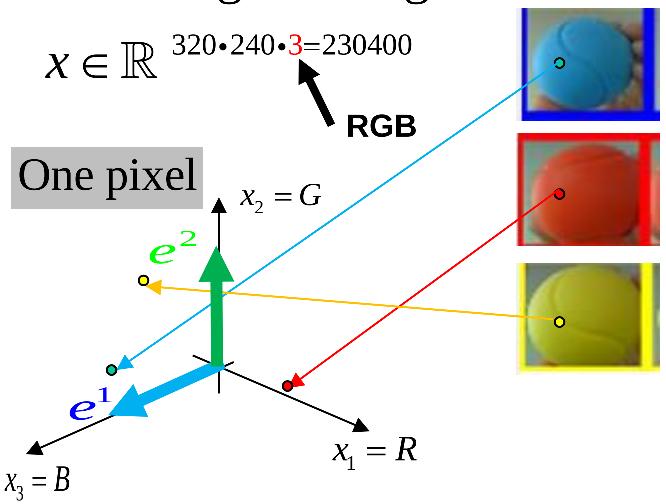


e^1 e^2

Interpreting PCA projections and eigenvectors

Each image is a high-dimensional vector

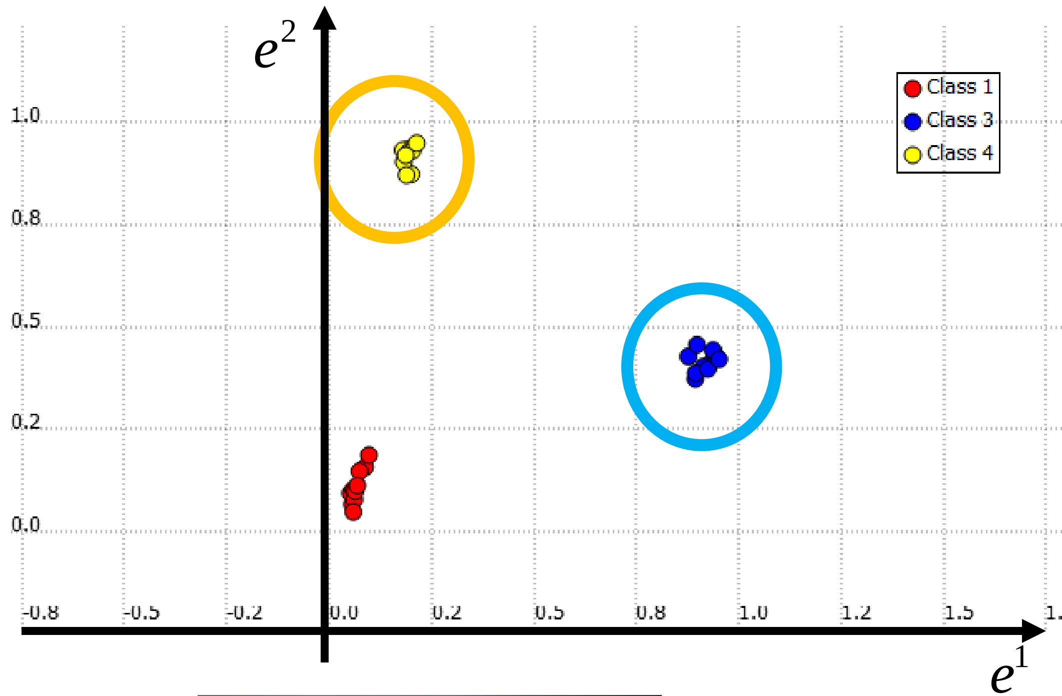
$x \in \mathbb{R}^{320 \cdot 240 \cdot 3 = 230400}$



$$e^1 = \underbrace{(x_1)^T \cdot e^1}_{\substack{\text{coordinate} \\ \text{of } e^1 \text{ onto } x_1}} x_1 + \underbrace{(x_2)^T \cdot e^1}_{\substack{\text{coordinate} \\ \text{of } e^1 \text{ onto } x_2}} x_2 + \ldots$$

$$e^1 \sim \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad e^2 \sim \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

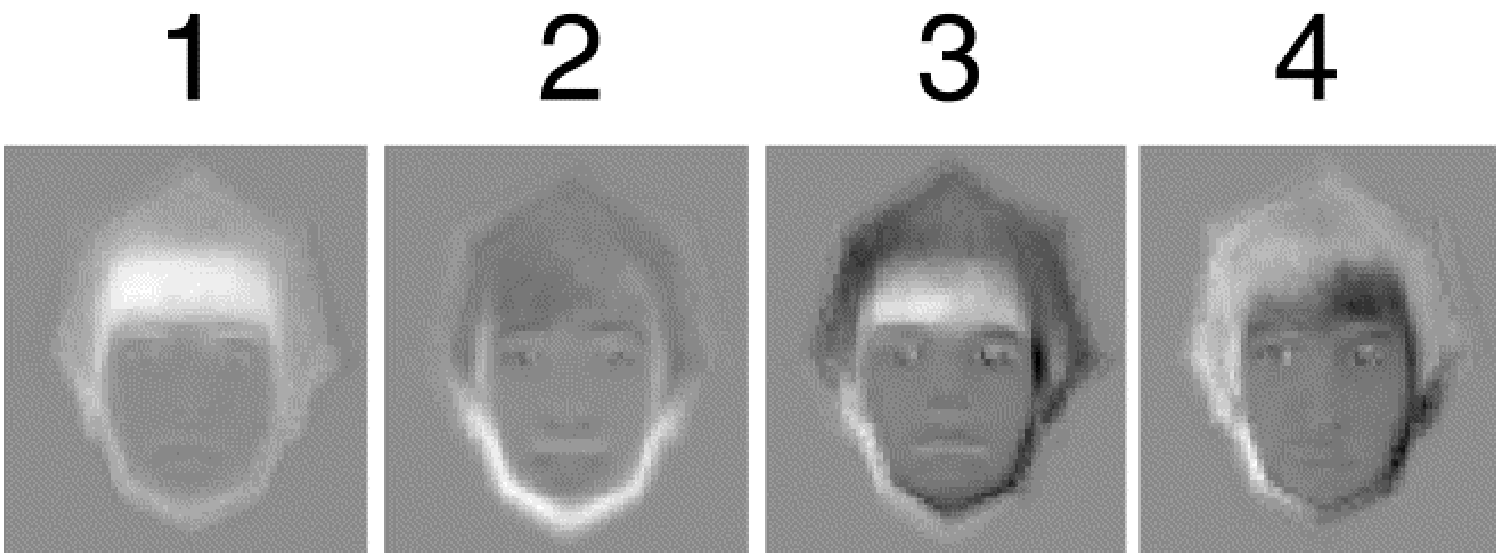
Repeats for each pixel



e^1 e^2

Eigenfaces

PCA applied to a set of 100 faces, coded in a high dimensional pixel space (54 150 dimensions),

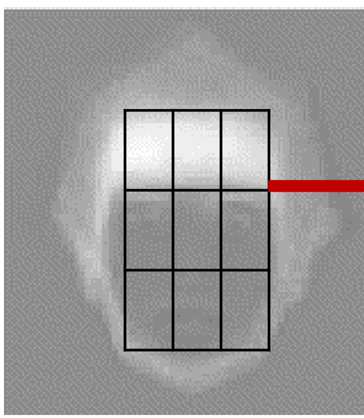


First 4 projections (4 principal components with largest eigenvalues)

Hancock, P et al (1996). Face processing: human perception and principal components analysis. *Memory and Cognition* 24 1, pp. 26–40

Interpreting PCA projections and eigenvectors

What would the entries of the first eigenvector look like?

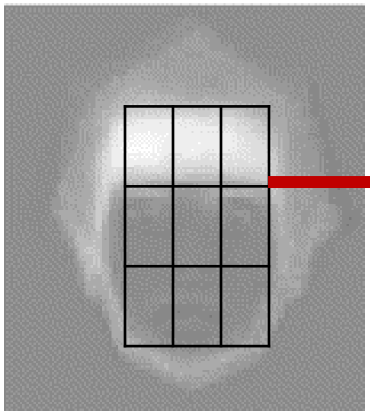


$$e^1 = \begin{bmatrix} ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \end{bmatrix}$$

Each pixel is coded with a scalar
Grey scale [0 ,1]
0 = Black
1 = White

Interpreting PCA projections and eigenvectors

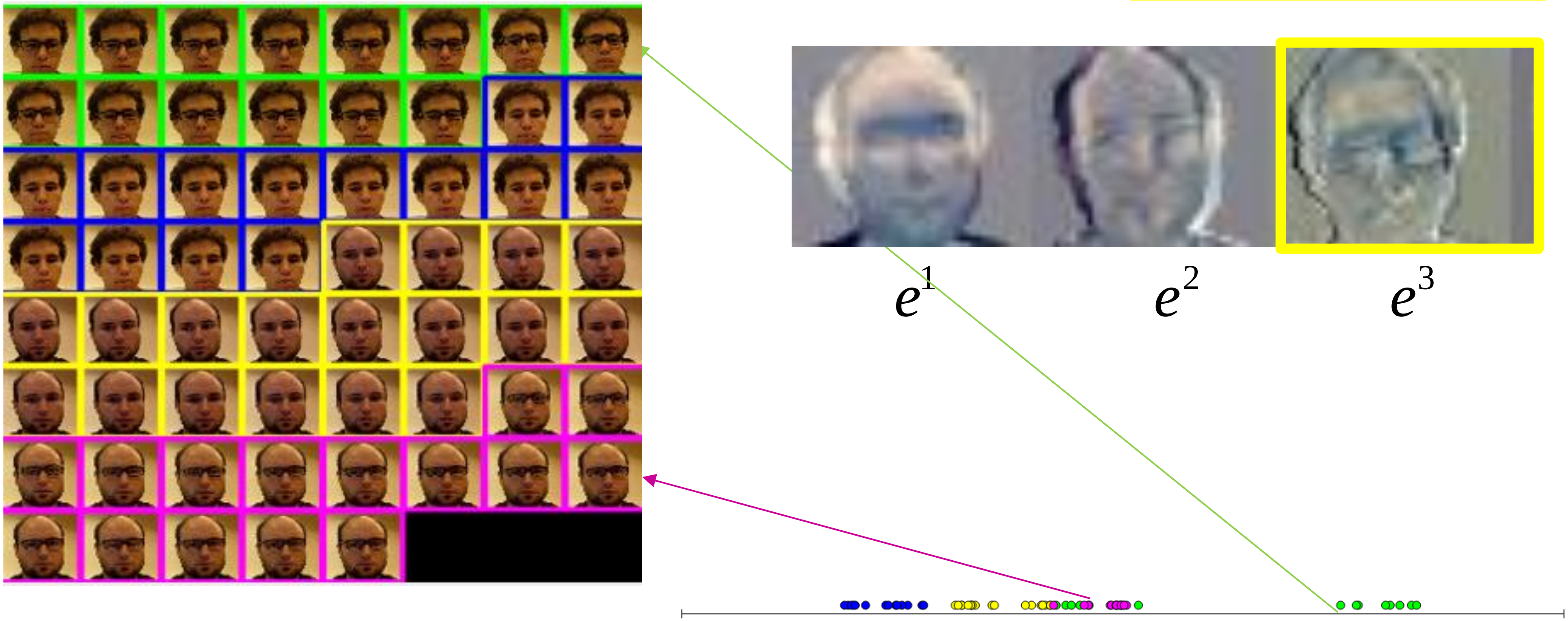
What would the entries of the first eigenvector look like?



$$e^1 \sim \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Entries not exactly zero

Groups glasses



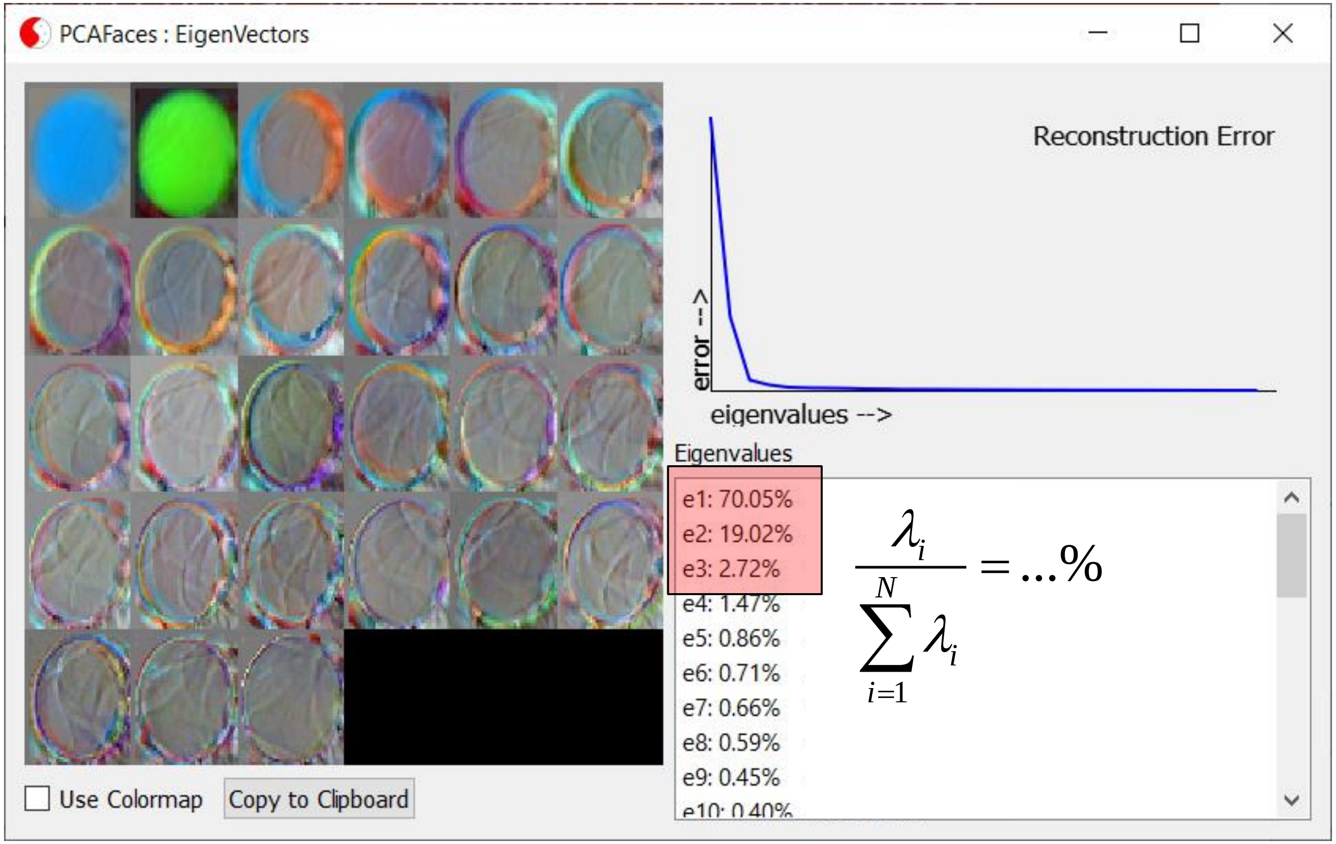
Projection along e^3

How to choose the optimal p eigenvectors?

Choose the smallest number of eigenvectors p but with smallest information loss. In the literature, you will often see that people select a subset of eigenvectors so as to incur no more than 10% information loss.

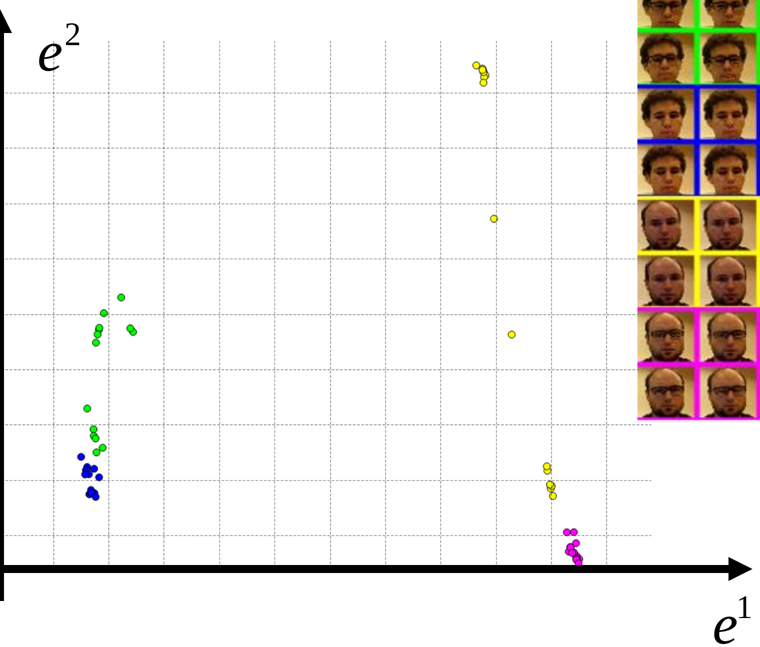
Information loss is measured as fraction of variance of data retained in the projections. Variance is measured by the eigenvalues.

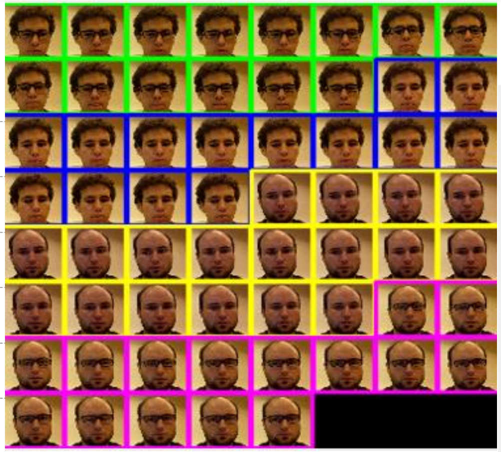
$$\frac{\sum_{j=p+1}^{N} \lambda_j}{\sum_{i=1}^{N} \lambda_i} \leq 0.1$$

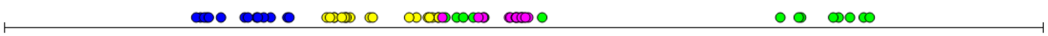


How to choose the optimal p eigenvectors?

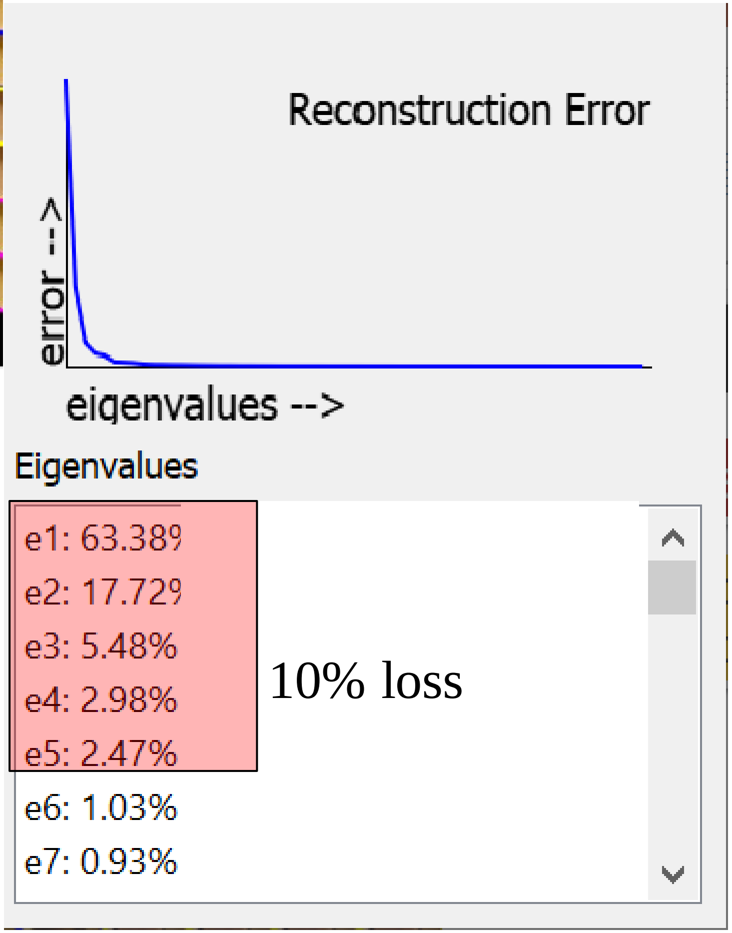
But first two eigenvectors are sufficient for separating the two faces and the third eigenvector is sufficient to extract the glasses' class.





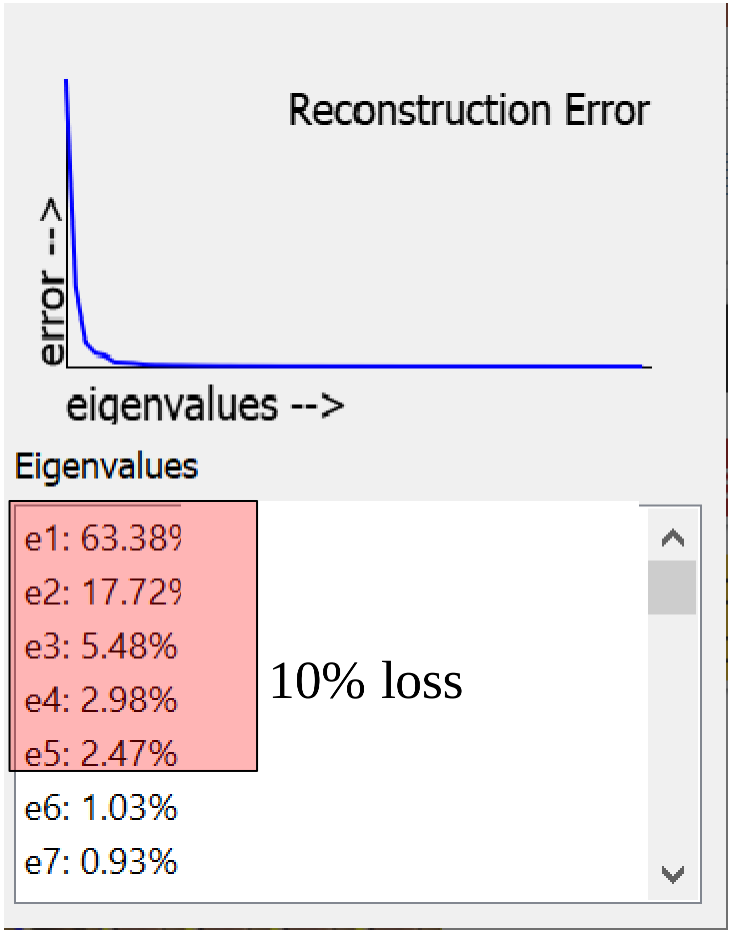


Projection along e^3



Take-Home Message

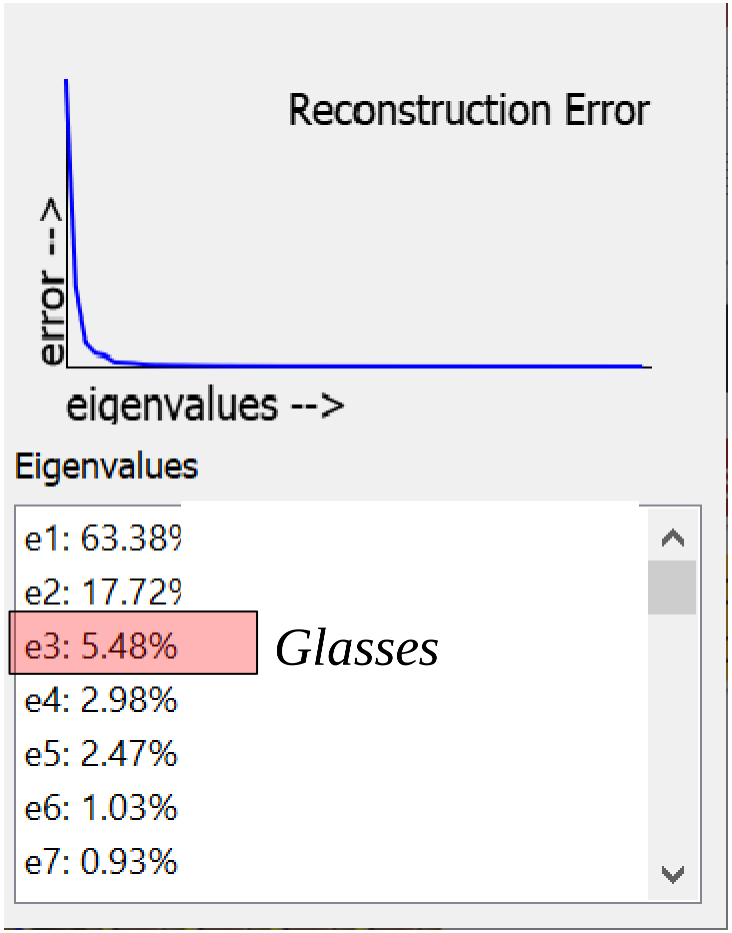
If your goal is to reduce dimensionality of the data with least deformation and information loss, pick the eigenvectors in decreasing order of their eigenvalues, until you reach the % of the variance you wish to retain.



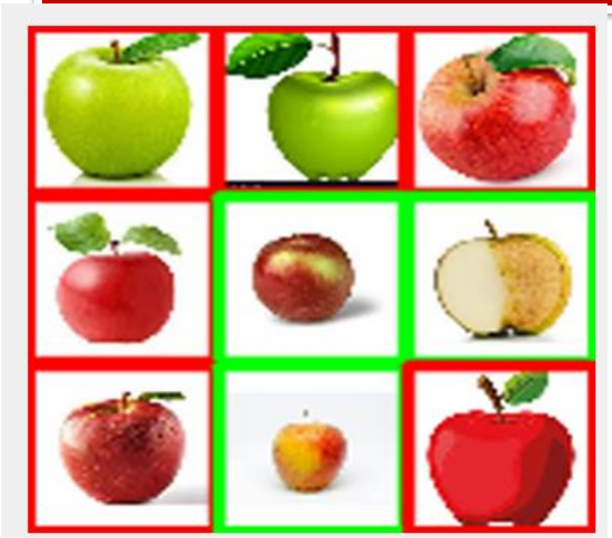
Take-Home Message

If your goal is to extract some specific information, pick the eigenvector that conveys this information.
They are not necessarily the first eigenvectors. Relevant information may be entailed in other eigenvectors, sometimes in eigenvectors with low eigenvalues.

Beware though that if you pick an eigenvector with very low eigenvalue, its statistical power will be low too and you may be picking on noise.



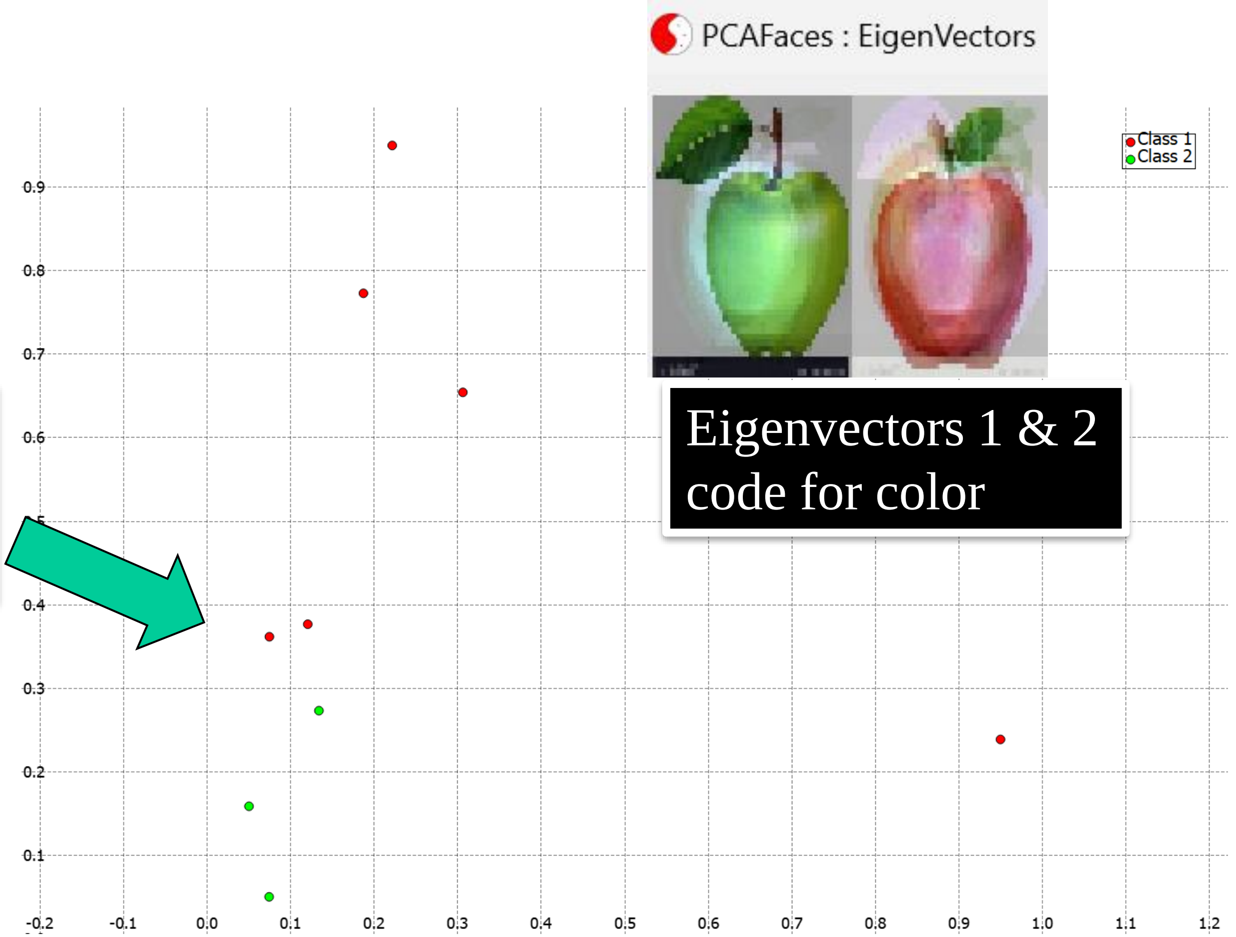
2nd Example – Fruit Dataset



Classes: Apple with a leaf vs. Apple without a leaf

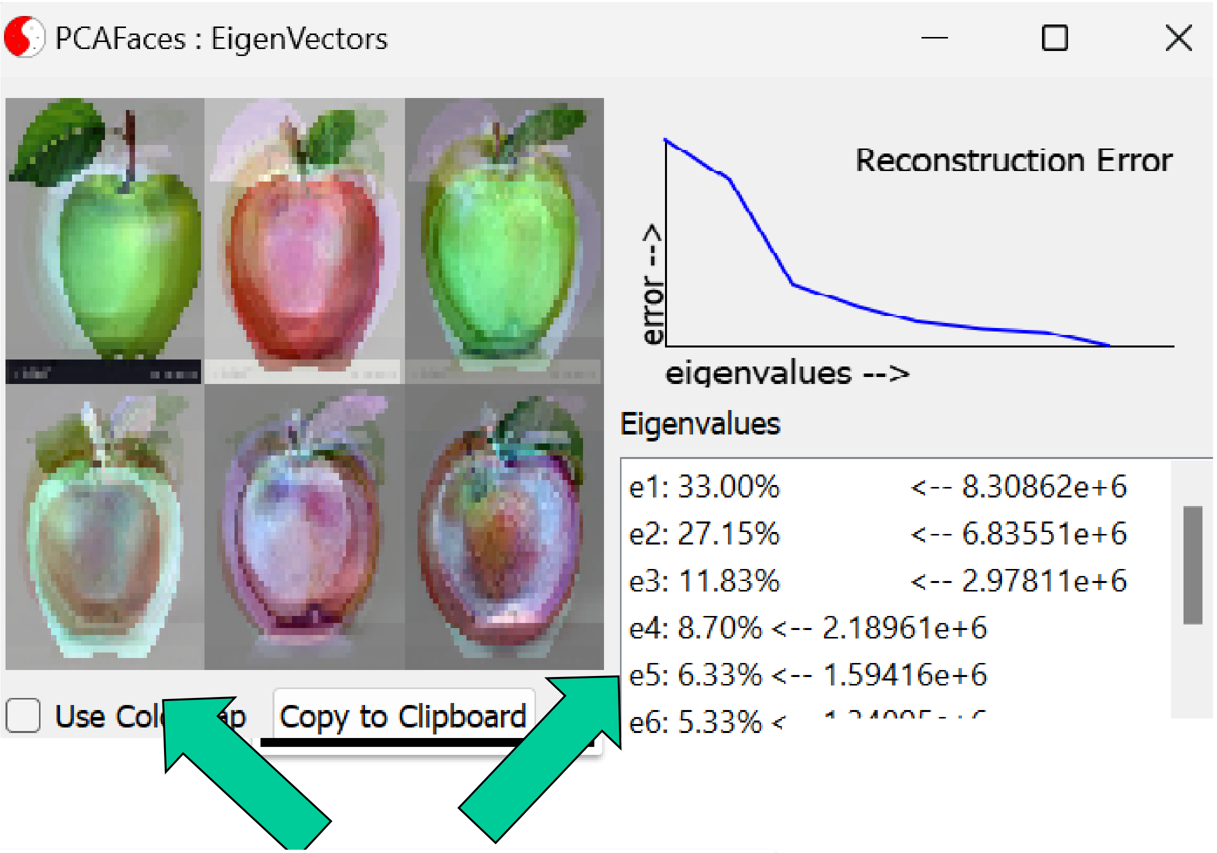
Eigenvectors 1 & 2 code for color

Difficult to separate the two classes

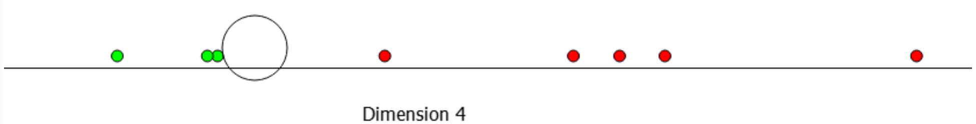


Projection along e^1 and e^2

2nd Example – Fruit Dataset



Eigenvector 4 entails information about the leafs, but it has very low statistical significance due to the fact that the leaf is encapsulated in few pixels.

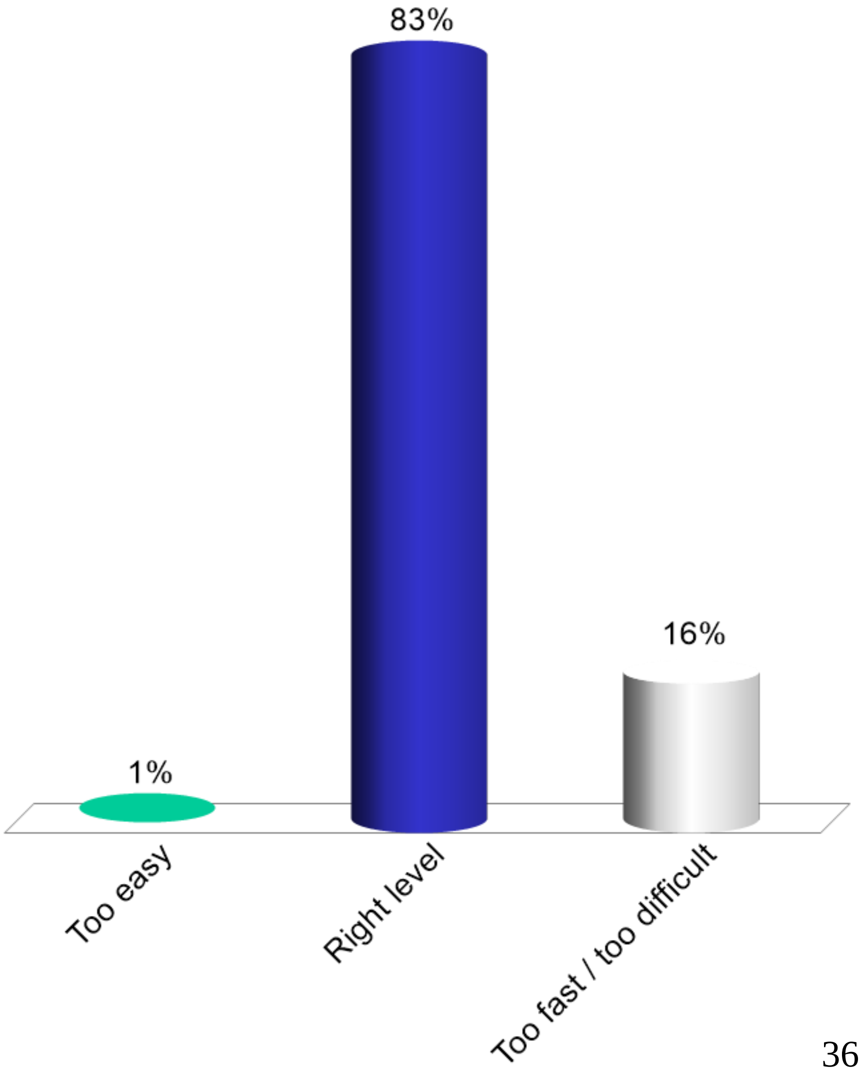


Projection along e^4

Qualify the content of the interactive class

Multiple answers possible

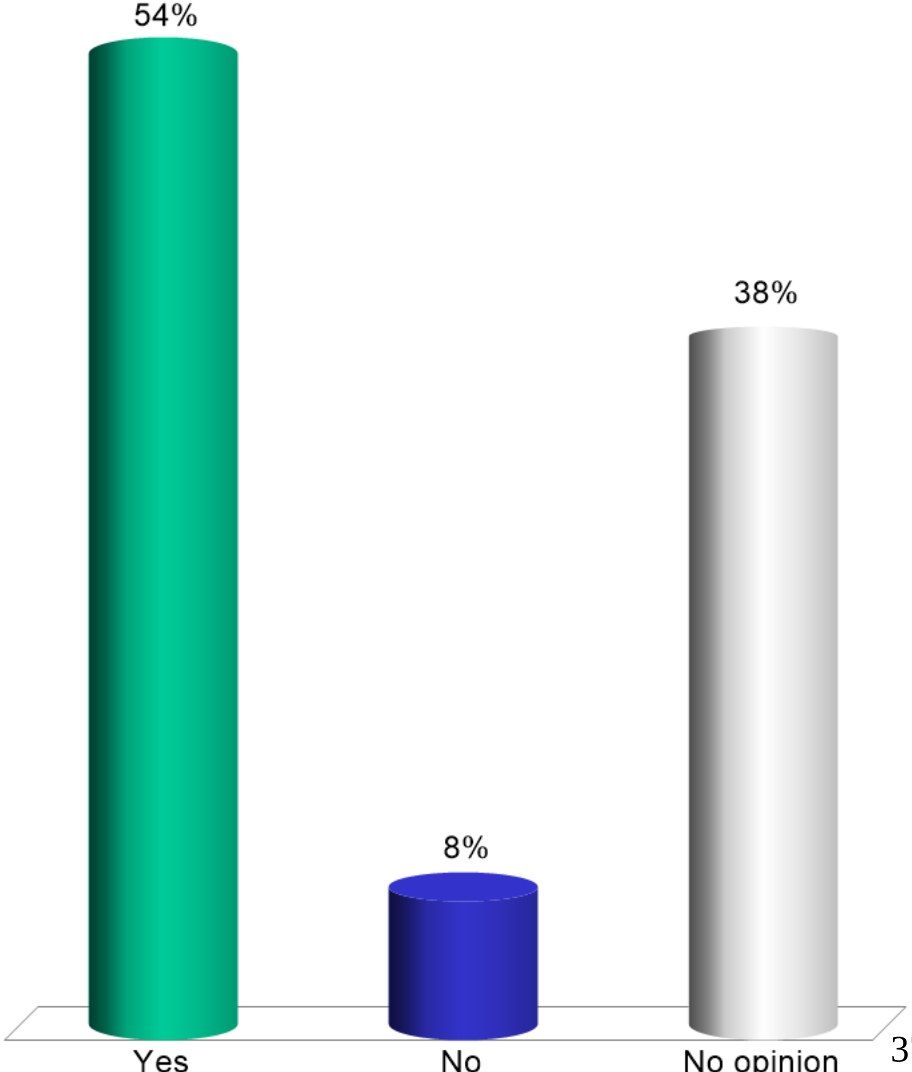
A. Too easy
B. Right level
C. Too fast / too difficult



For those of you following class on-line:

Did you find it easy to interact with TA on and zoom during the interactive exercise sessions?

A. Yes
B. No
C. No opinion



NEXT WEEK!

Class starts at 9h15am → 13h00

Computer-based practice session on PCA

On site

In classrooms

BC 07 – 08, CM 1103

On-line

Exercise session

Launch DISCORD: https://discord.gg/TwTfKkv but stay on zoom too!

Download the exercises from moodle

Raise hand and wait for a TA available

Zoom will be used only for live presentations of solutions

Students can work individually or in group → create their own room

Ask for TA who will join your room (3 min. max)