Lecture 18: Dimensionality Reduction
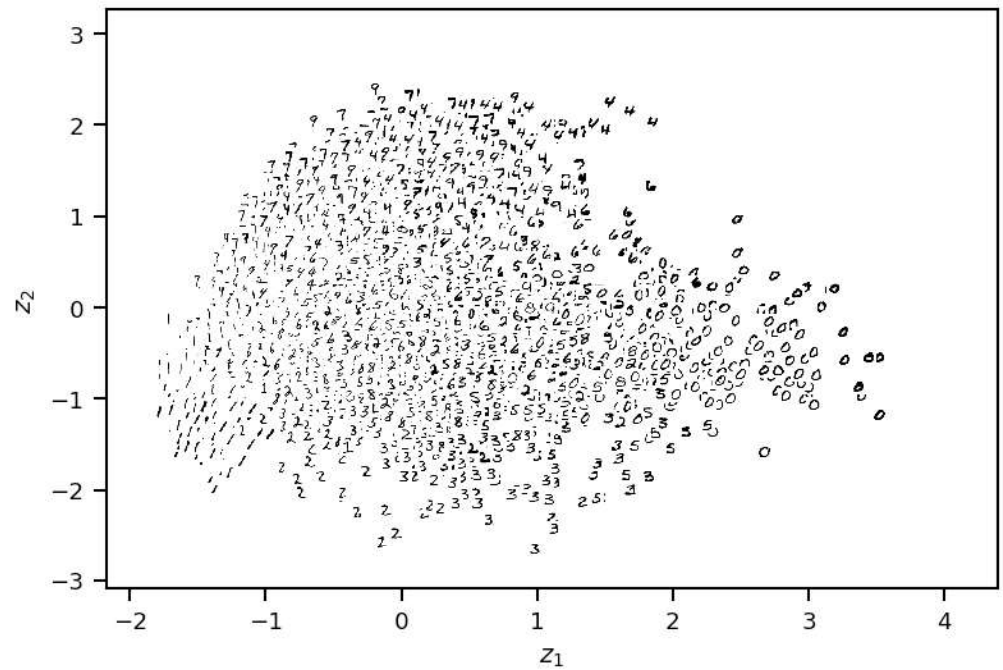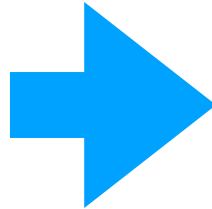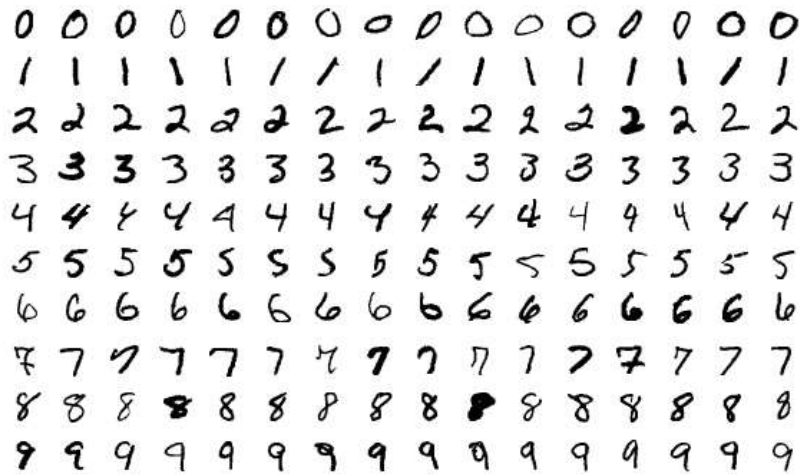
Professor Ilias Bilionis

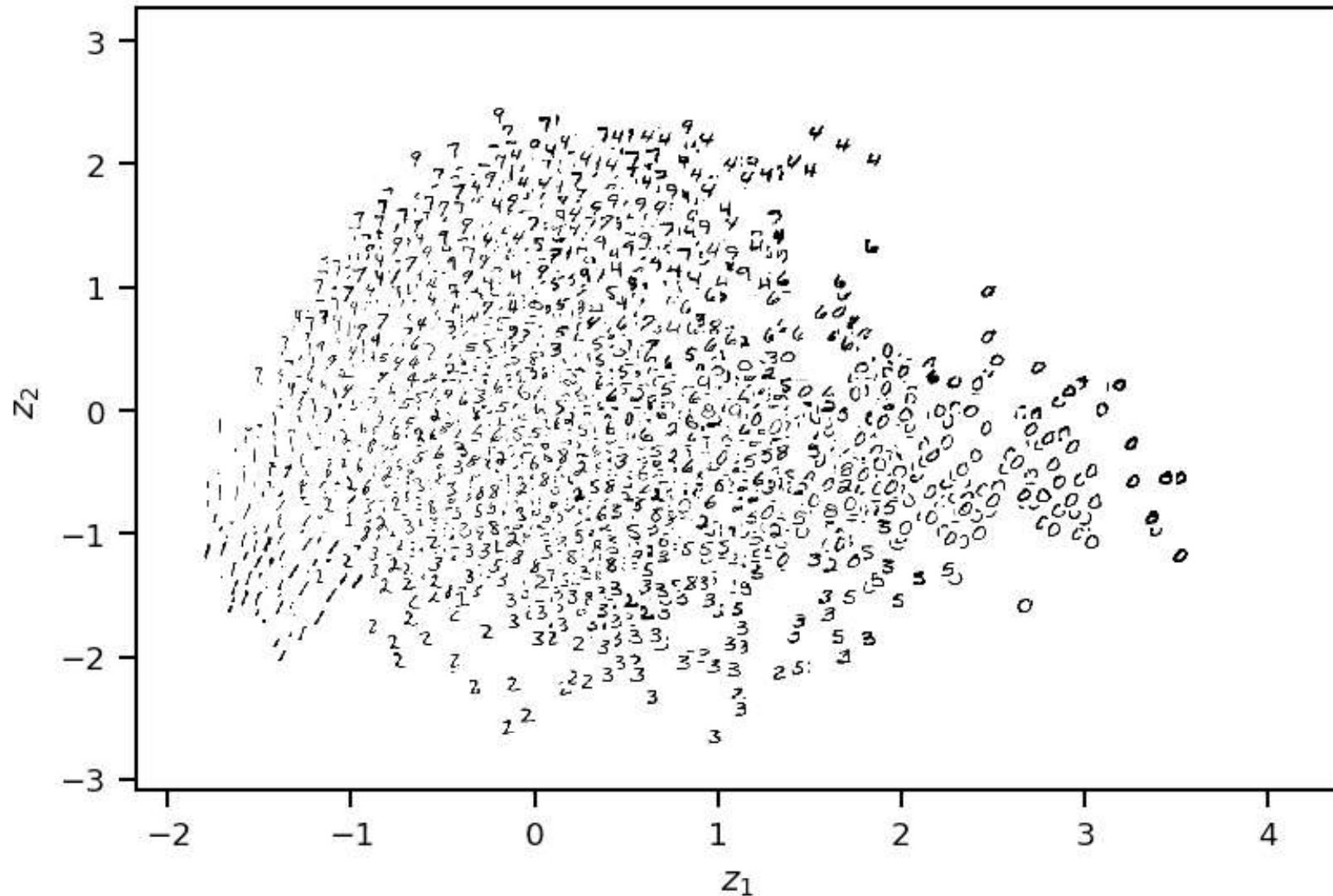# Clustering high-dimensional data

Clustering high-dimensional data

- K-means directly on high-dimensional data does not work very well…

- IDEA:

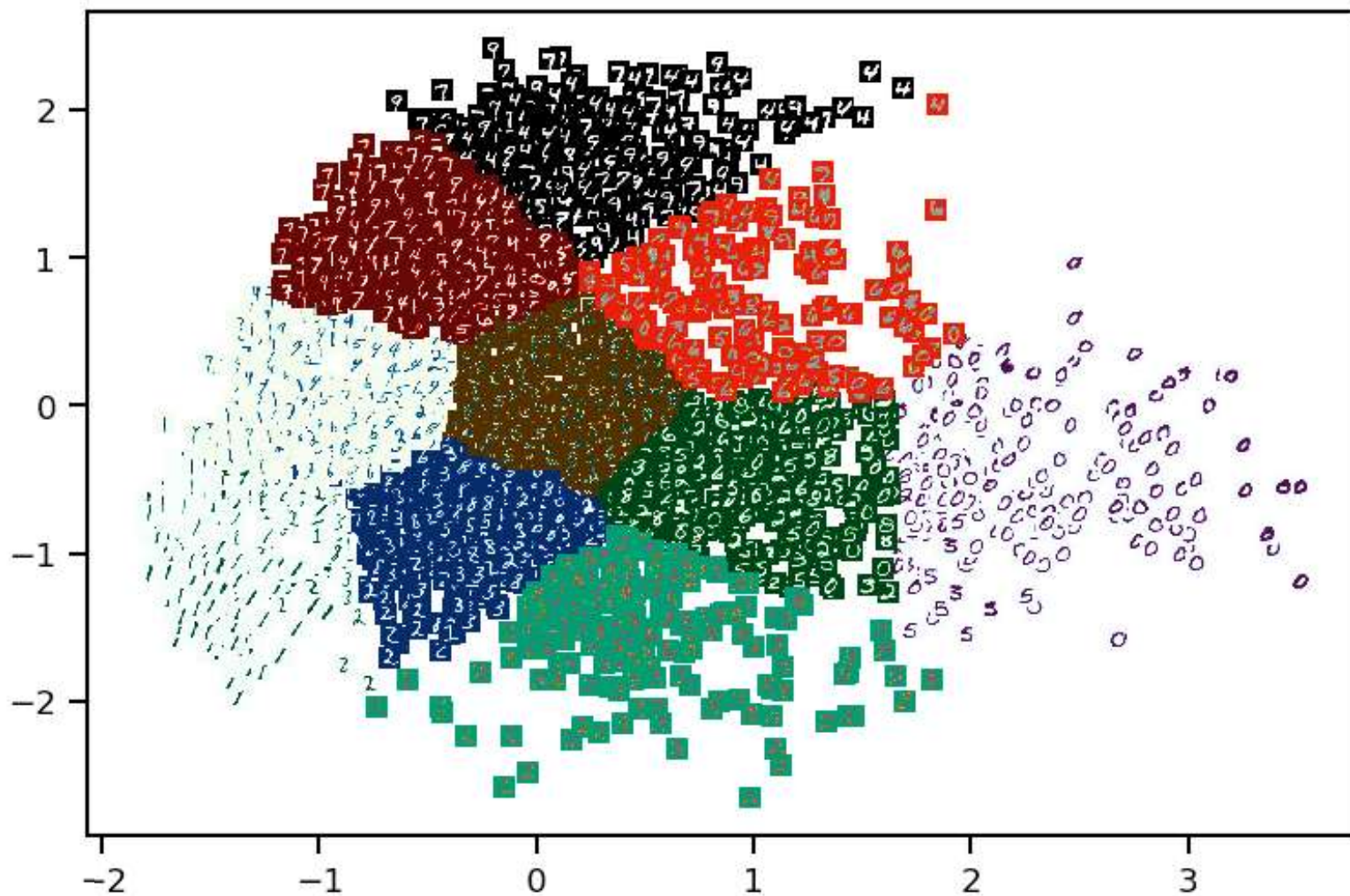 - First reduce dimensionality using PCA.

  - Then apply K-means.

**PREDICTIVE
SCIENCE LABORATORY**

Applying PCA

Showing 2 of 87 dimensions

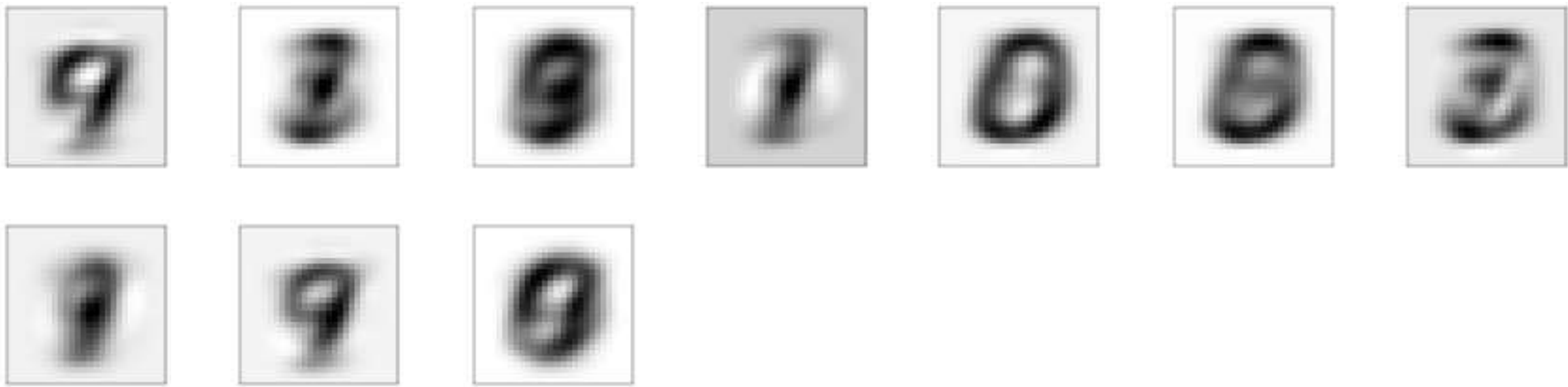**PREDICTIVE SCIENCE LABORATORY**

25

K-means on projected dataset

K-means on projected dataset

What are the cluster centers?

It looks like for this data set non-linear dimensionality reduction would be better.

sometimes PCA may not work as expected

PREDICTIVE SCIENCE LABORATORY