Advanced machine learning

MTH793P 2024

Omer Bobrowski, QMUL
(adapted from slides by Prof. Martin Benning)

Introduction

Supervised vs. Unsupervised Learning

supervised

Input: $(x_1, y_1), \ldots, (x_n, y_n)$ - labeled data

Goal: Use given labels to make accurate prediction

Examples: classification, regression

unsupervised

Input: $x_1, \ldots, x_n$ - unlabeled data

Goal: "Blindly" discover hidden patterns in the data

Examples: clustering, dimension reduction

This semester, the plan is to cover the following topics:

Supervised machine learning

- Semi-supervised binary classification with graphs

Unsupervised machine learning

- Unsupervised binary classification with graphs
- K-means clustering
- Spectral clustering
- PCA
- Manifold learning
- Matrix factorisation & completion
- Autoencoders

All relevant information is or will be made available on the QM+ module page

Module Information

- **In-person lectures:** Mondays 12:00-14:00
- **In-person tutorials:** Wednesdays 13:00-14:00
- **Office hour:** Wednesdays 14:00-15:00 (MB-111)

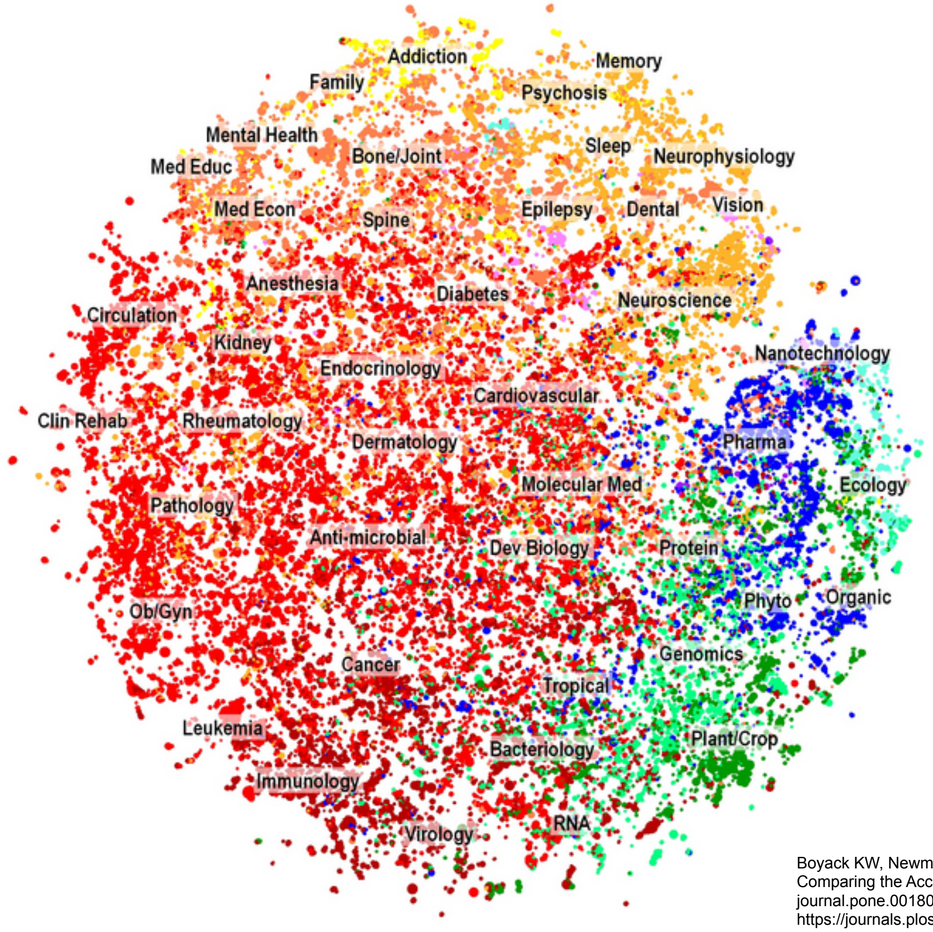
Assessment & Available Material

- Midterm exam (week 11), 40% of final mark.
- Final project, 60% of final mark.
- Weekly* coursework (written / coding), unmarked.
- Online booklet
- Online reading list

Clustering

Two-dimensional map of the PMRA cluster solution, representing nearly 29,000 clusters and over two million articles.

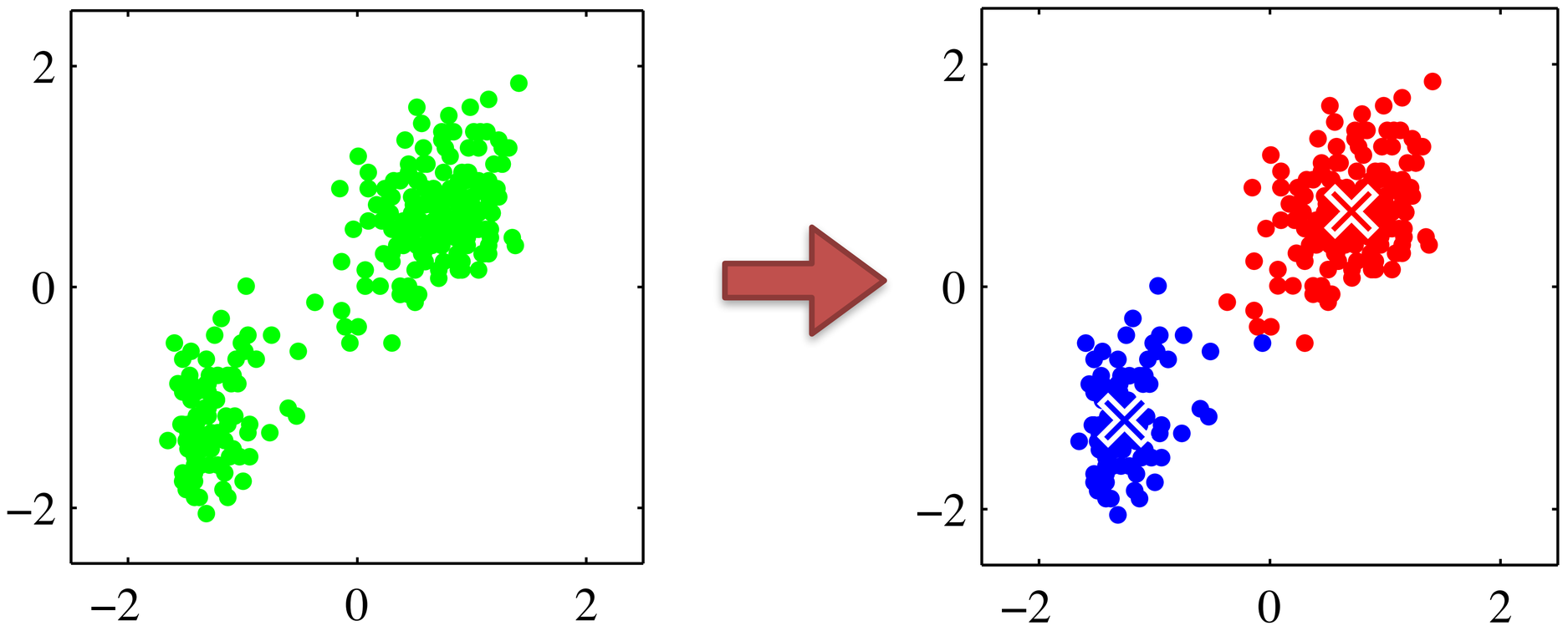
Clustering of two million biomedical publications



Boyack KW, Newman D, Duhon RJ, Klavans R, Patek M, et al. (2011) Clustering More than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches. PLOS ONE 6(3): e18029. https://doi.org/10.1371/journal.pone.0018029
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018029

Clustering

Clustering:



From Bishop. Pattern Recognition & Machine Learning

Clustering

Another clustering example: image compression / quantisation

$K = 2$

$K = 3$

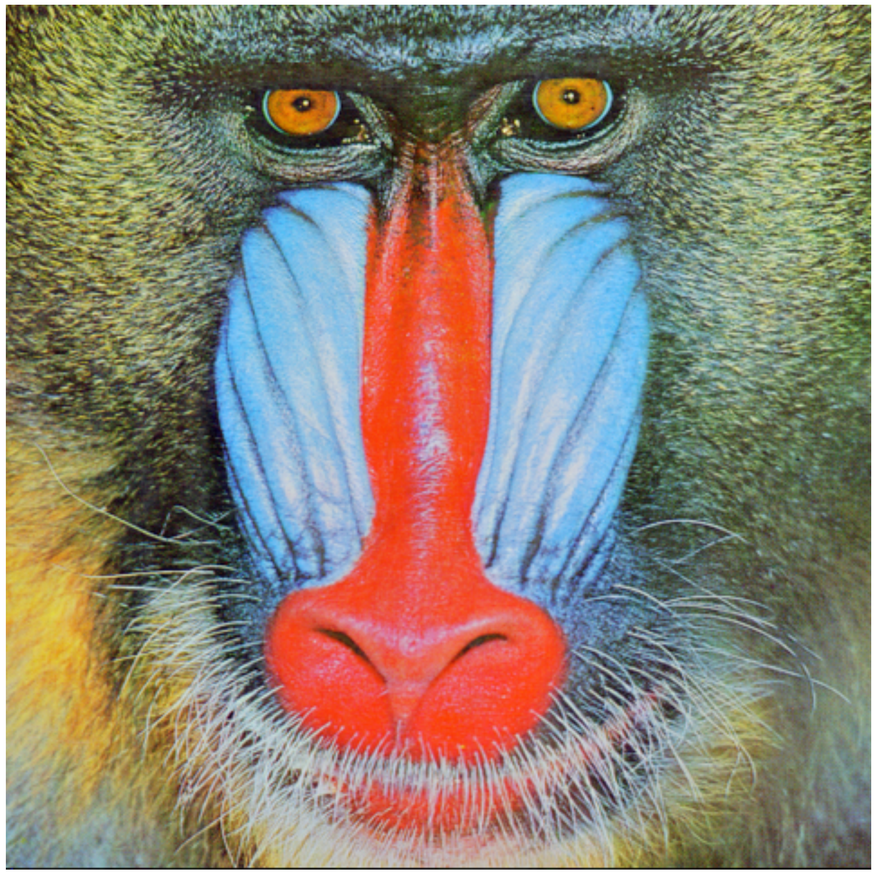
$K = 10$

Original image

From Bishop. Pattern Recognition & Machine Learning

Clustering

Another clustering example: image compression / quantisation



Matrix Factorisation

Recommender systems

Example: given some book-ratings per reader, predict remaining ratings

Input: x_{ij} = book rating for book i by reader j

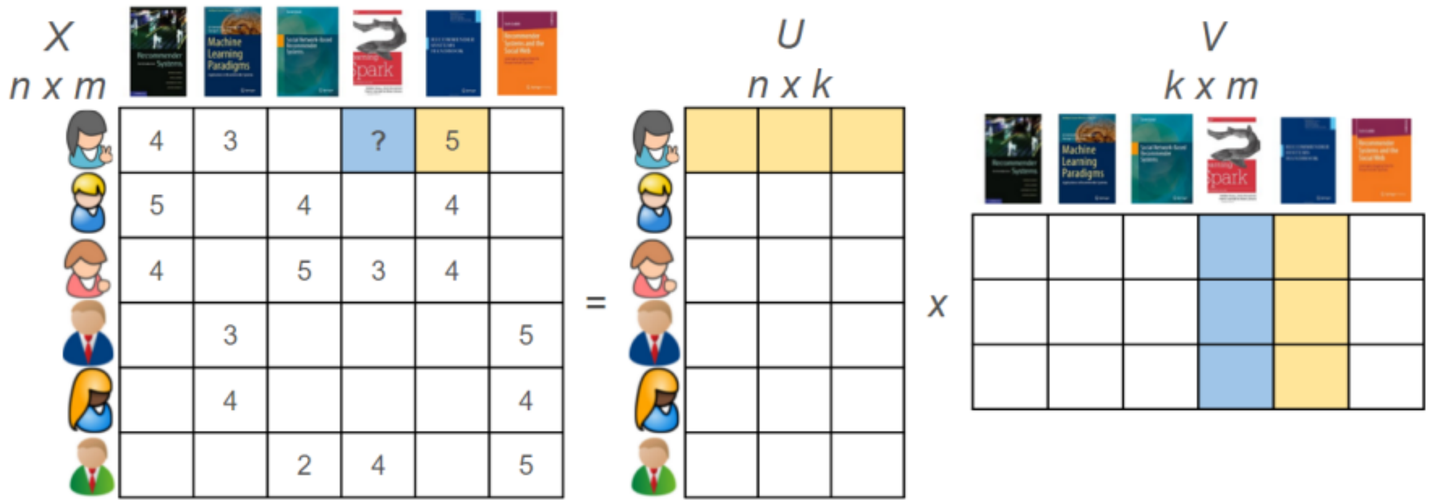
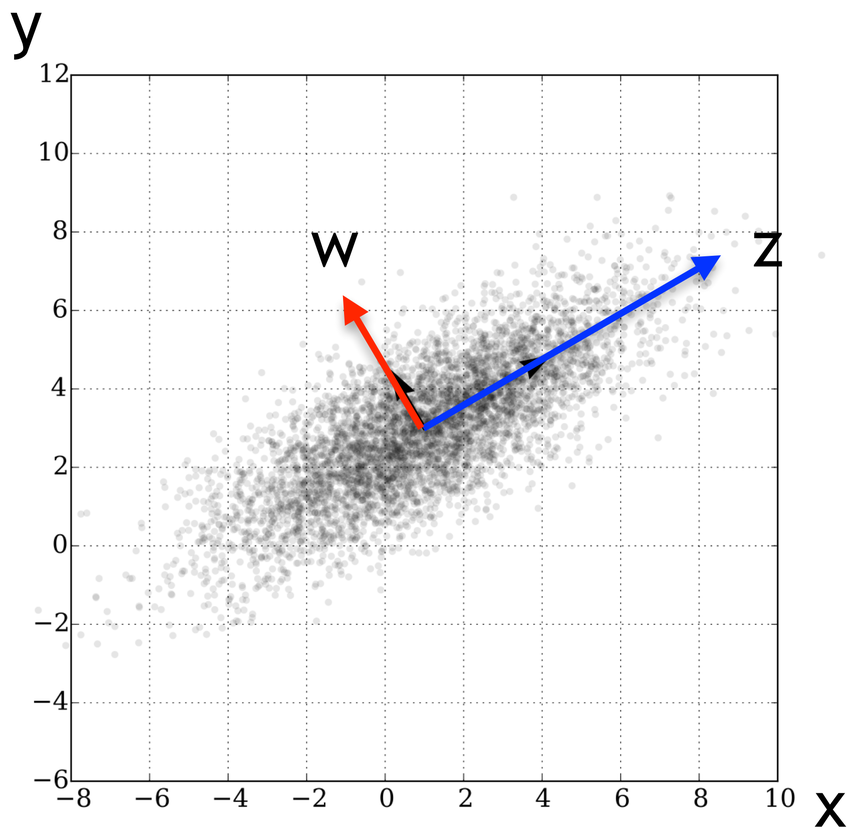


Image Source

Principle applies to other categories, *e.g.* movies 🍿

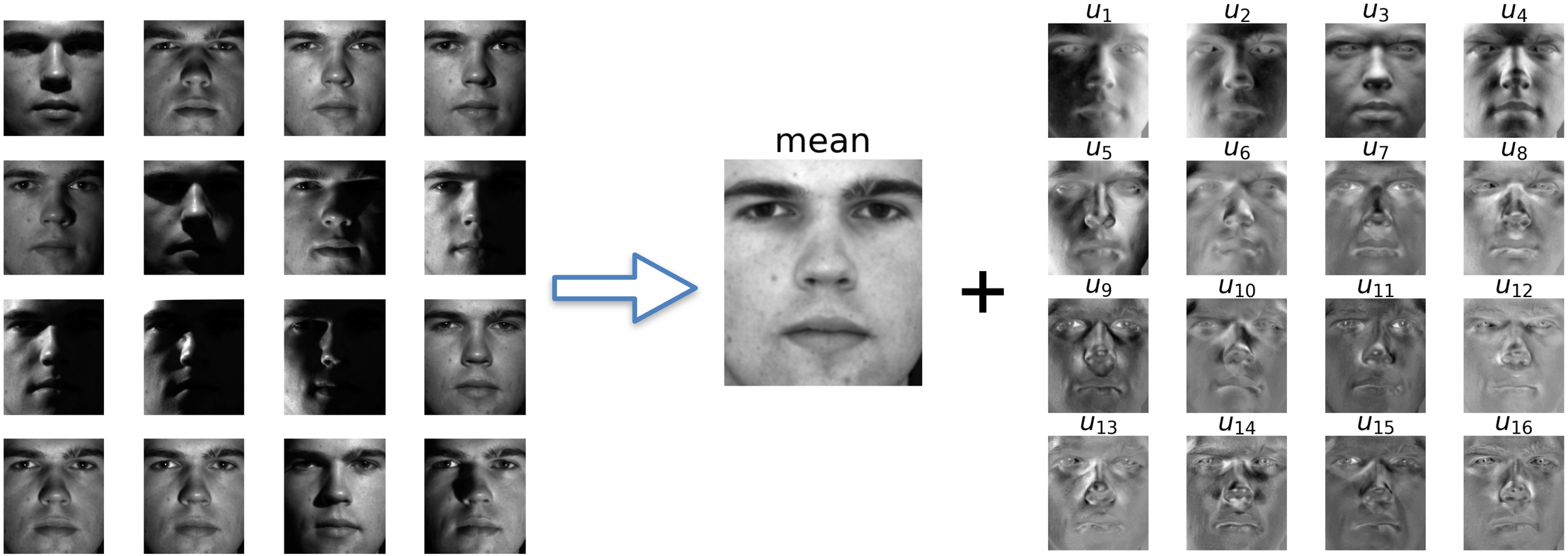
Dimension Reduction

Principal Component Analysis (PCA)



Dimension Reduction

Principal Component Analysis (PCA) Example: Eigenfaces*



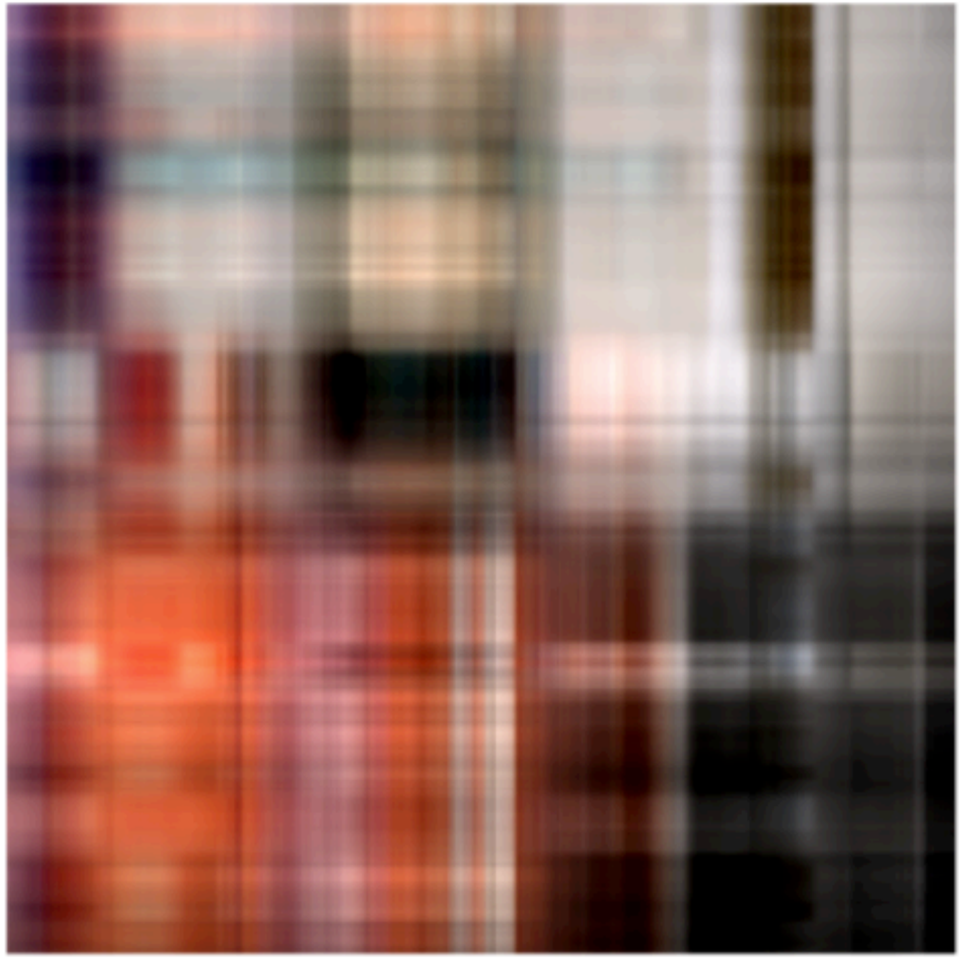
*from Understanding machine learning by Shalev-Schwartz and Ben-David

Dimension Reduction

Principal Component Analysis (PCA) Example: Eigenfaces*

*from Understanding machine learning by Shalev-Schwartz and Ben-David

Dimension Reduction



Eileen Collins. ©NASA Great Images database

Dimension Reduction

Dimension Reduction

Robust PCA

Original face

Approximation

=

Low-rank

+

Sparse

Dimension Reduction

Robust PCA

©Met Office Informatics Lab

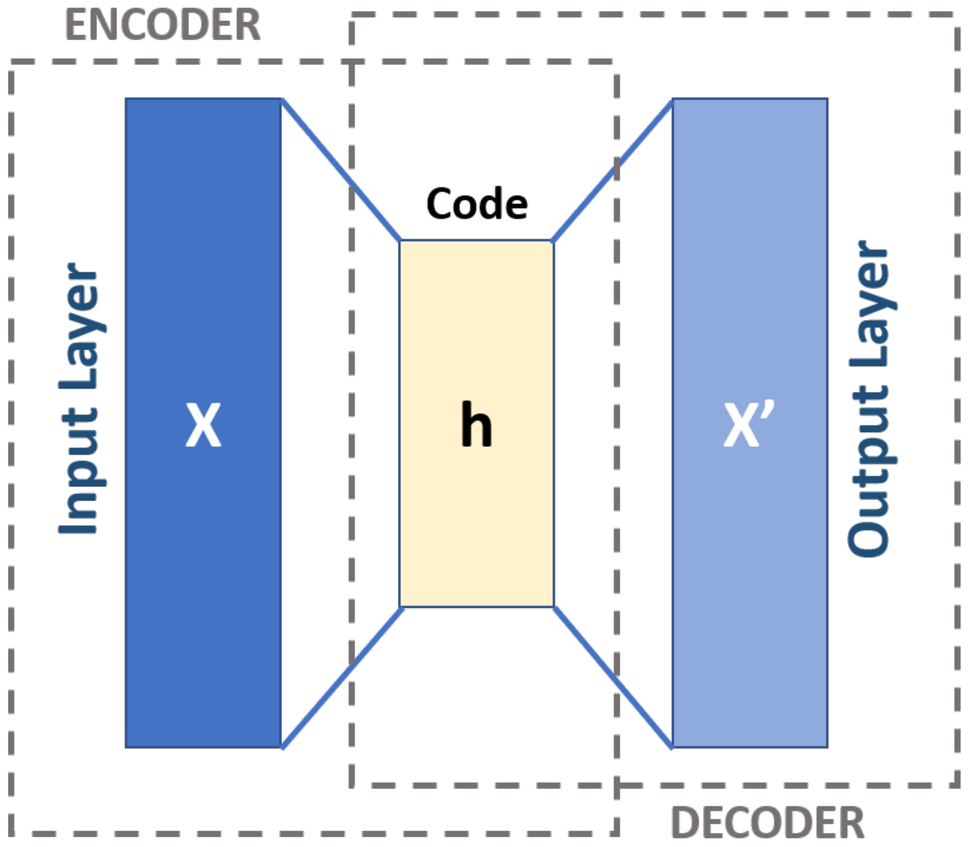
Can we separate vehicles from background?

Autoencoders

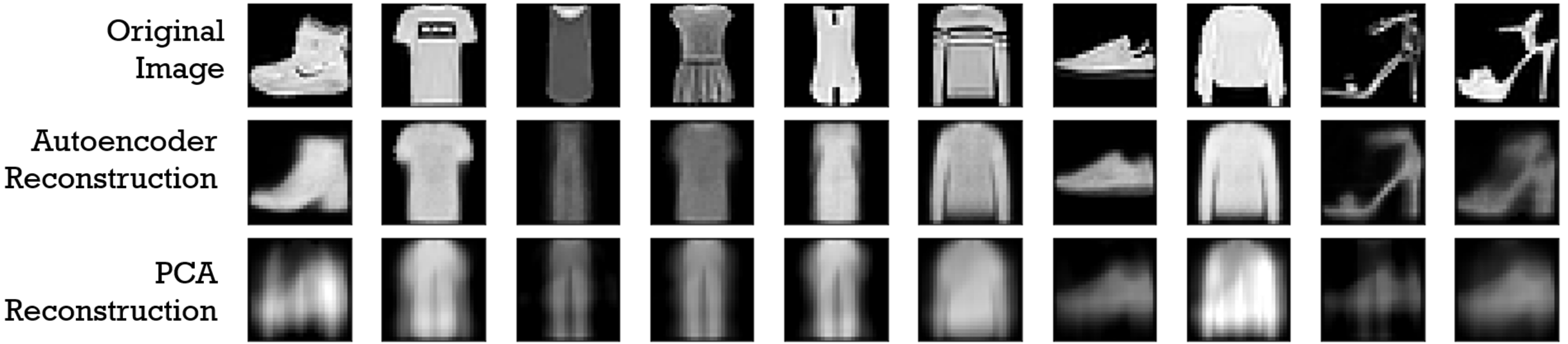
What is an autoencoder?

"An **autoencoder** is a type of artificial neural network used to learn efficient data codings in an unsupervised manner. The aim of an **autoencoder** is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal noise."

Source: Wikipedia, https://en.wikipedia.org/wiki/Autoencoder



Autoencoders



Fashion MNIST

Reconstruction of 28x28pixel images by an Autoencoder / PCA

by Michaela Massi, CC BY-SA 4.0

Questions?