

Advanced machine learning

MTH793P 2026

Dr. Hong Qi , QMUL
(adapted from slides by Prof. Martin Benning)

Introduction

hong.qi@qmul.ac.uk

Supervised vs. Unsupervised Learning

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

supervised

Goal: Use given labels to make accurate prediction

Examples: classification, regression

Input: x_1, \dots, x_n - unlabeled data

unsupervised

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

Unsupervised machine learning

- Semi-supervised binary classification with graphs
- 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



TH793P

Module Information

- **Lectures: Wednesdays, 14:00-16:00, Queens LG1 (Week 1-6, 8-11)**
- **Last lecture: Tuesday, 9:00-11:00, Queens LG1 (Week 12)**
- **IT Labs: Fridays, 15:00-16:00, Bancroft 1.15a PC Lab (Week 2-6, 8-12)**

Office Hours: Thursdays 12:00-13:00 at Learning Cafe in MB-B11



TH793P

Assessment & Available Material

- Midterm exam (week 7), 40% of final mark.
- Final project, 60% of final mark.
- Weekly* coursework (written / coding), unmarked.
- Online booklet
- Online reading list

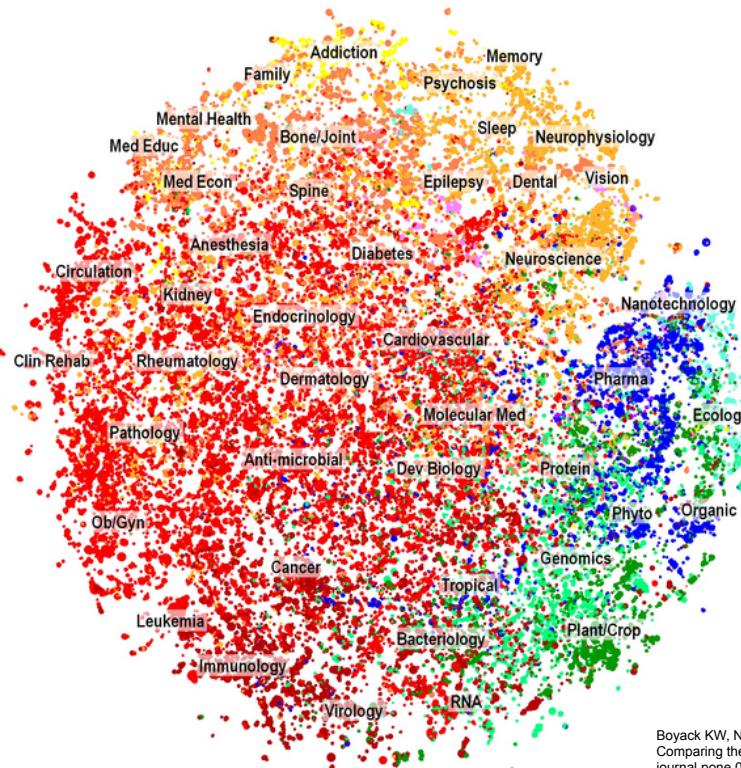


TH793P

Clustering

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

lustering of two million
biomedical publications

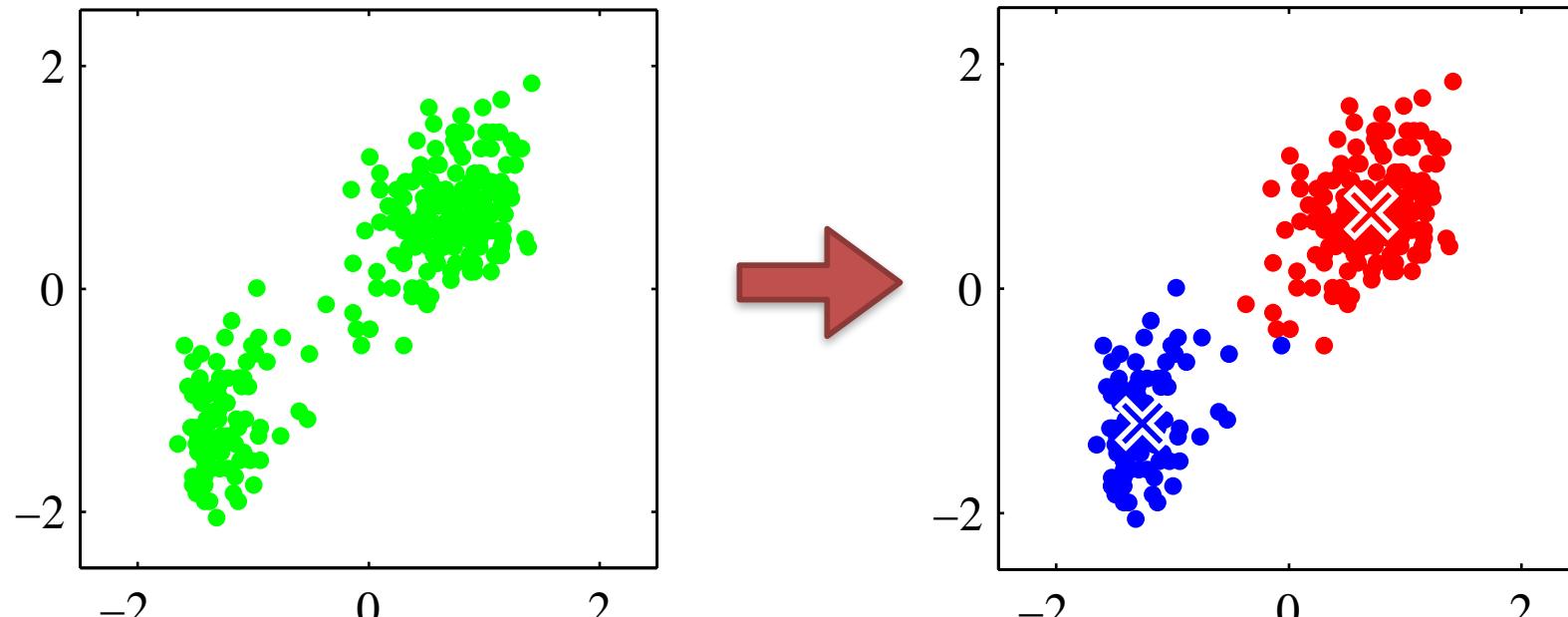


Boyack KW, Newman D, Duhon RJ, Klavans R, Patek M, et al. (2011) Clustering More than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches. PLOS ONE 6(3): e18029. <https://doi.org/10.1371/journal.pone.0018029>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018029>

TH793P

Clustering

Clustering:



From Bishop. Pattern Recognition & Machine Learning

Clustering

Another clustering example: image compression / quantisation

$K = 2$



$K = 3$



$K = 10$



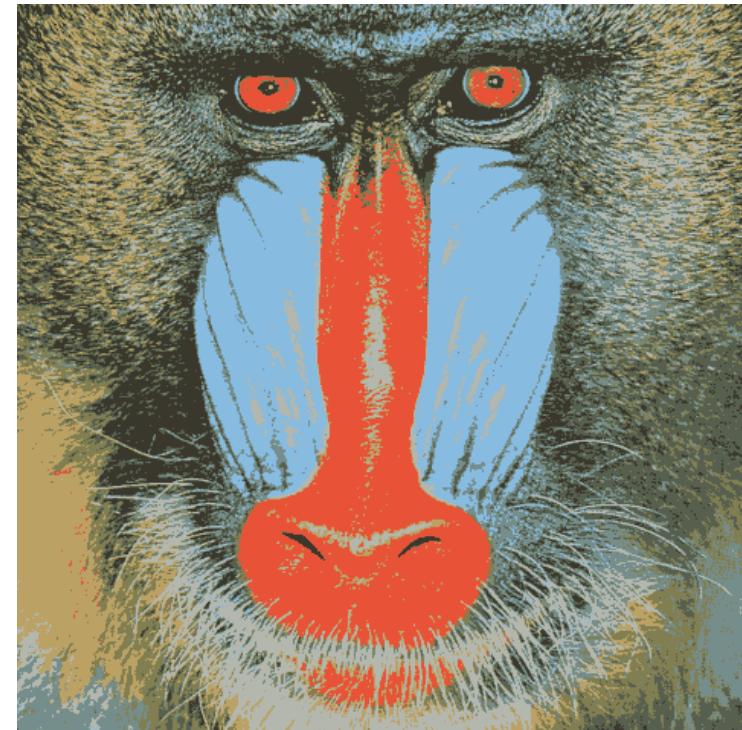
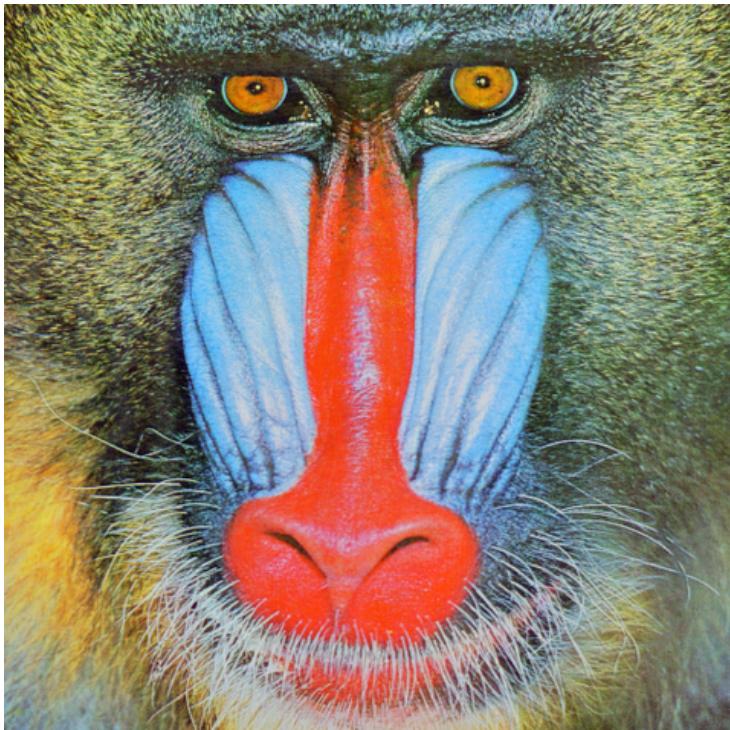
Original image



From Bishop. Pattern Recognition & Machine Learning

Clustering

Another clustering example: image compression / quantisation



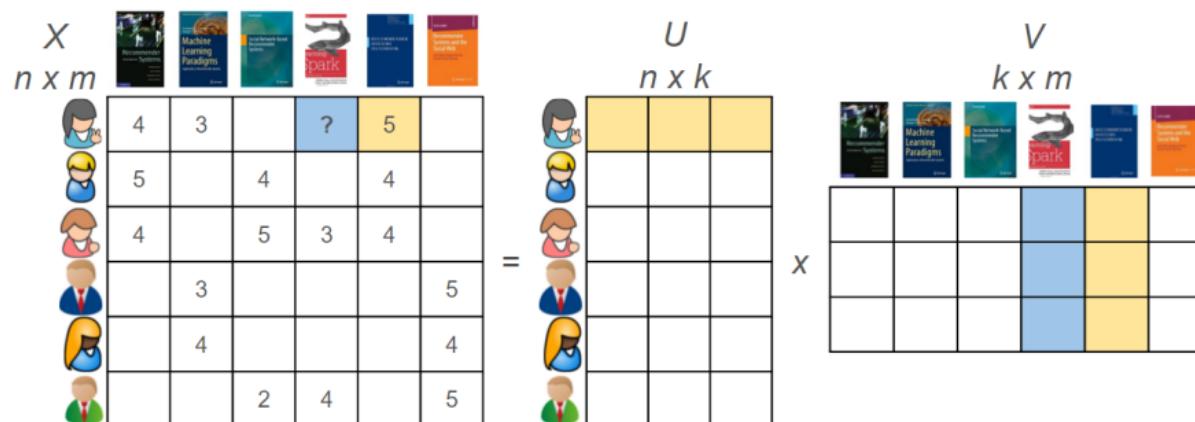
TH793P

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

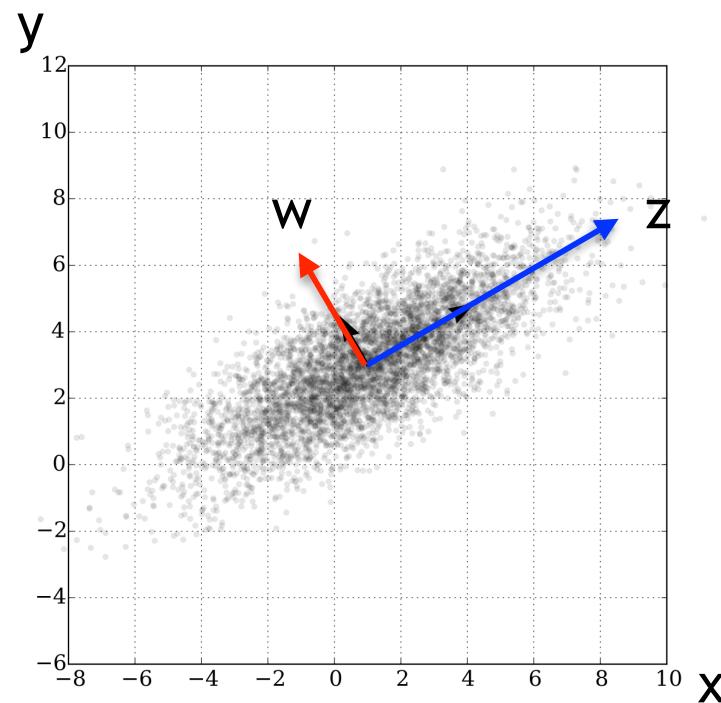


[Image Source](#)

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

Dimension Reduction

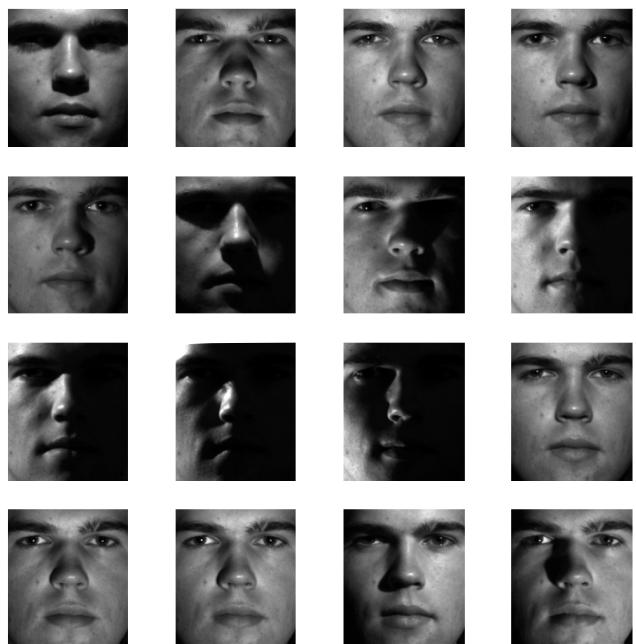
Principal Component Analysis (PCA)



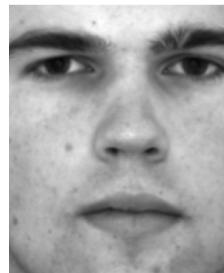
TH793P

Dimension Reduction

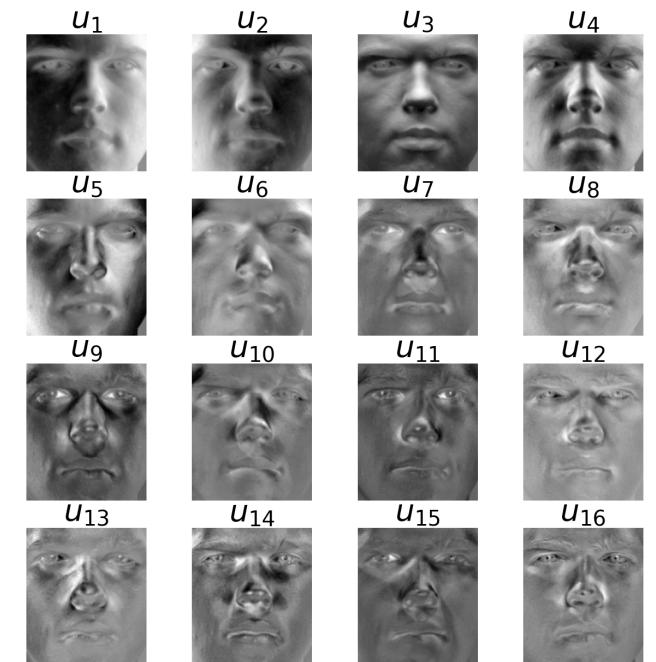
Principal Component Analysis (PCA) Example: Eigenfaces*



mean



+



TH793P

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

Dimension Reduction

Principal Component Analysis (PCA) Example: Eigenfaces*



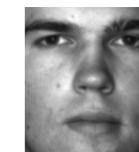
original



1 eigenface



5 eigenfaces



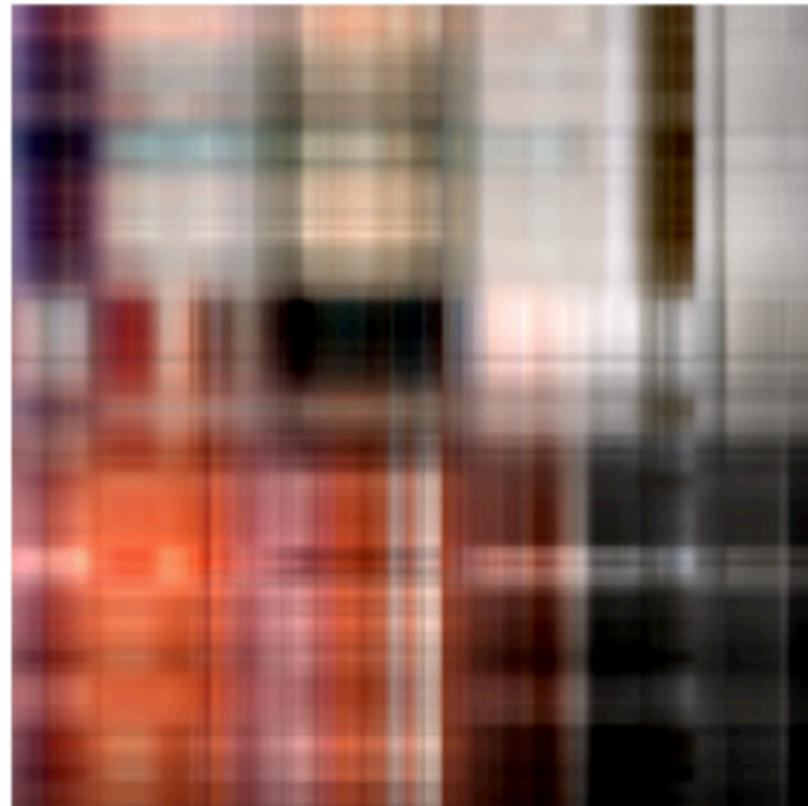
10 eigenfaces



100 eigenfac



Dimension Reduction

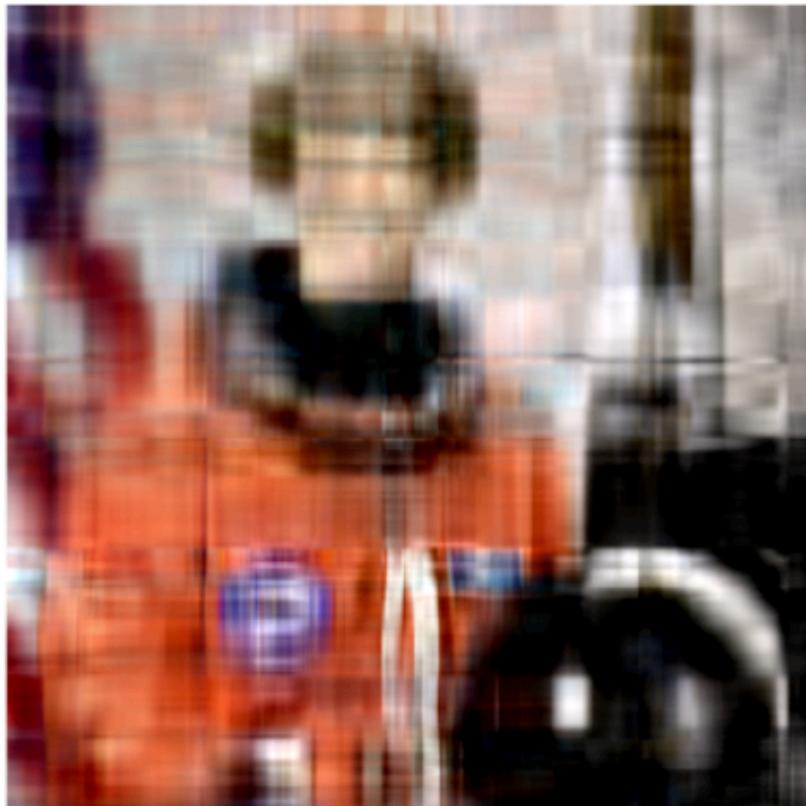


Eileen Collins. [©NASA Great Images database](#)



TH793P

Dimension Reduction



TH793P

Dimension Reduction

Robust PCA



Original face



TH793P

Approximation



Low-rank



Sparse



=

+

Dimension Reduction

Robust PCA



Can we separate vehicles
from background?



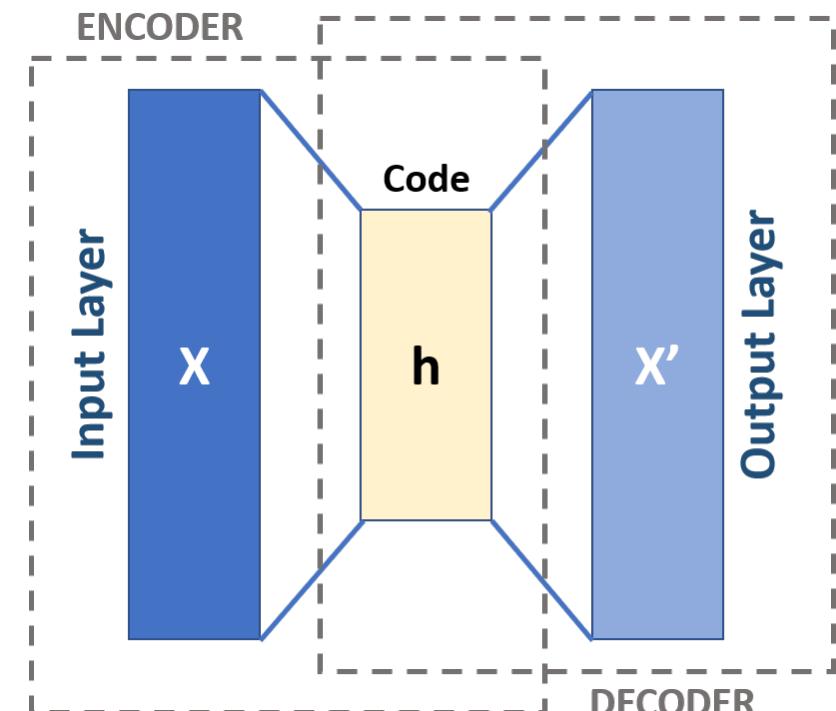
[@Met Office Informatics Lab](#)

TH793P

Autoencoders

What is an autoencoder?

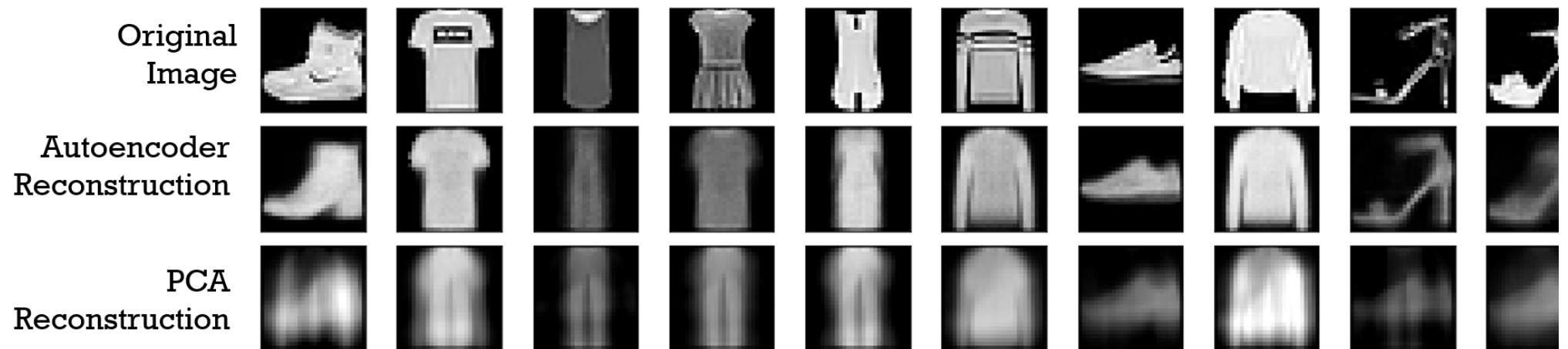
“An **autoencoder** is a type of artificial neural network used to learn efficient data codings in an unsupervised manner. The aim of an **autoencoder** is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal noise.”



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

[© Wikimedia commons](#)

Autoencoders



Fashion MNIST

Reconstruction of 28x28pixel images by an Autoencoder / PCA
by Michaela Massi, [CC BY-SA 4.0](#)

Questions?



TH793P