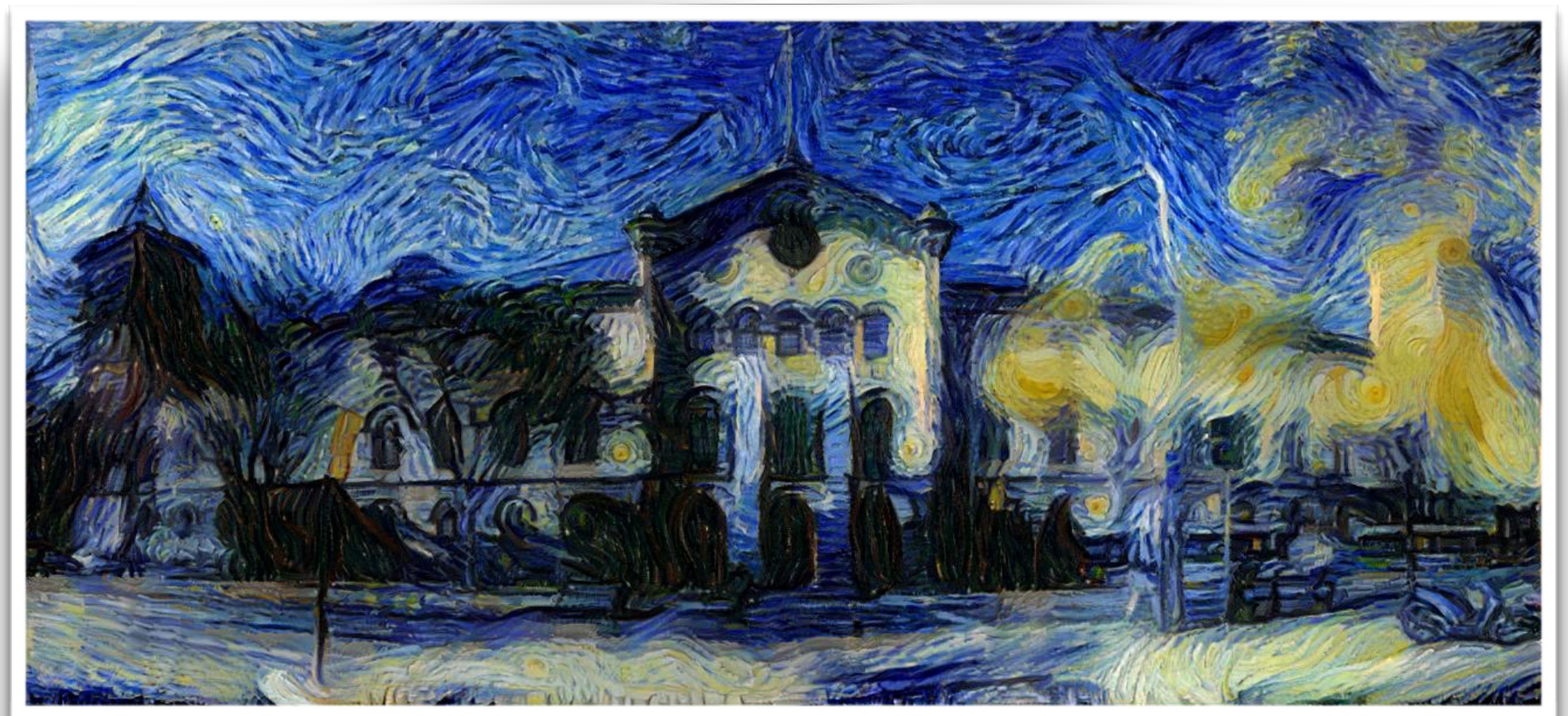
UNIVERSITAT DE BARCELONA



Unsupervised learning I: Manifold learning and deep learning

Oriol Pujol

Layout

- Autoencoders
- Learning unsupervised representations
- Sparse coding
- A manifold learning view
- Deep patient

Unsupervised learning tries to understand the properties of a particular set of data. There are different ways of doing this

- Clustering - Divide data in groups according to some notion of similarity.

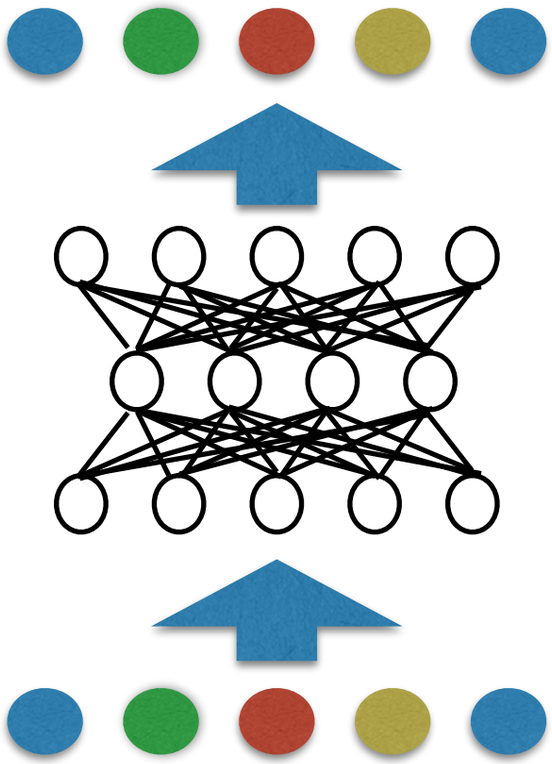
- Manifold learning - Understanding how data is distributed in the space, parameterising a manifold.

Autoencoder

- Build a network with the aim of reconstruction.

D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. In Parallel Distributed Processing. Vol 1: Foundations. MIT Press, Cambridge, MA, 1986.

Autoencoders



Reconstruction $\tilde{x}$

Loss $\mathcal{L}(\tilde{x}, x) = \sum(\tilde{x} - x)^2$

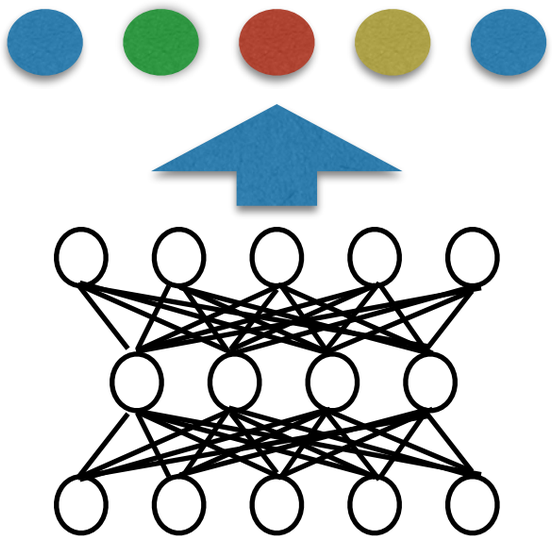
Input x

Problems

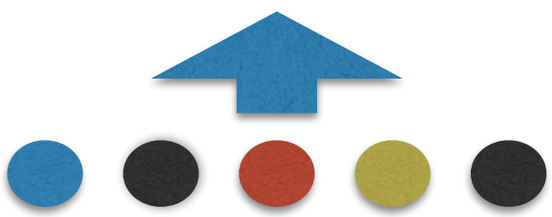
- In large networks it may learn the identity mapping rendering the auto encoder representation useless.

- In order to correct this issue and furthermore give robustness to the auto encoder, demonising auto encoders are proposed.

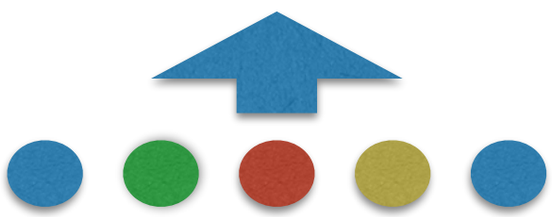
Denoising Autoencoders



Reconstruction



Noise: dropout input values



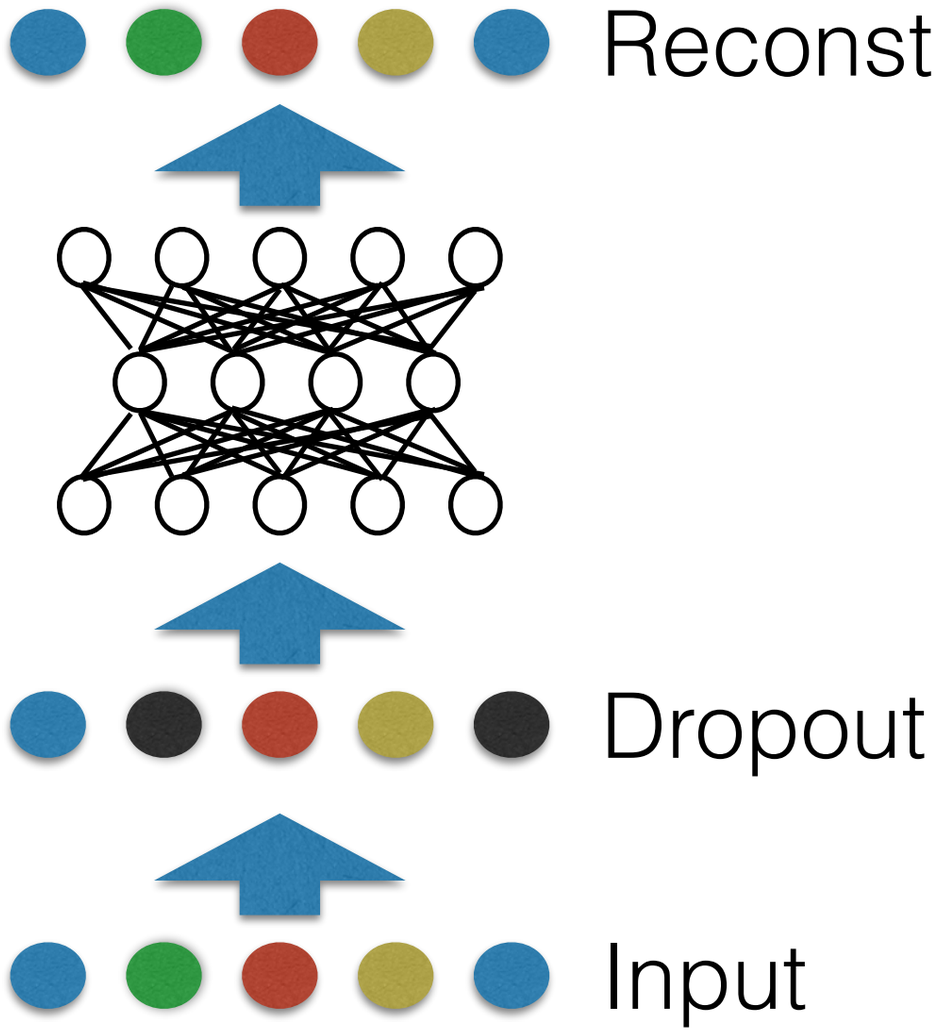
Input

Hands on

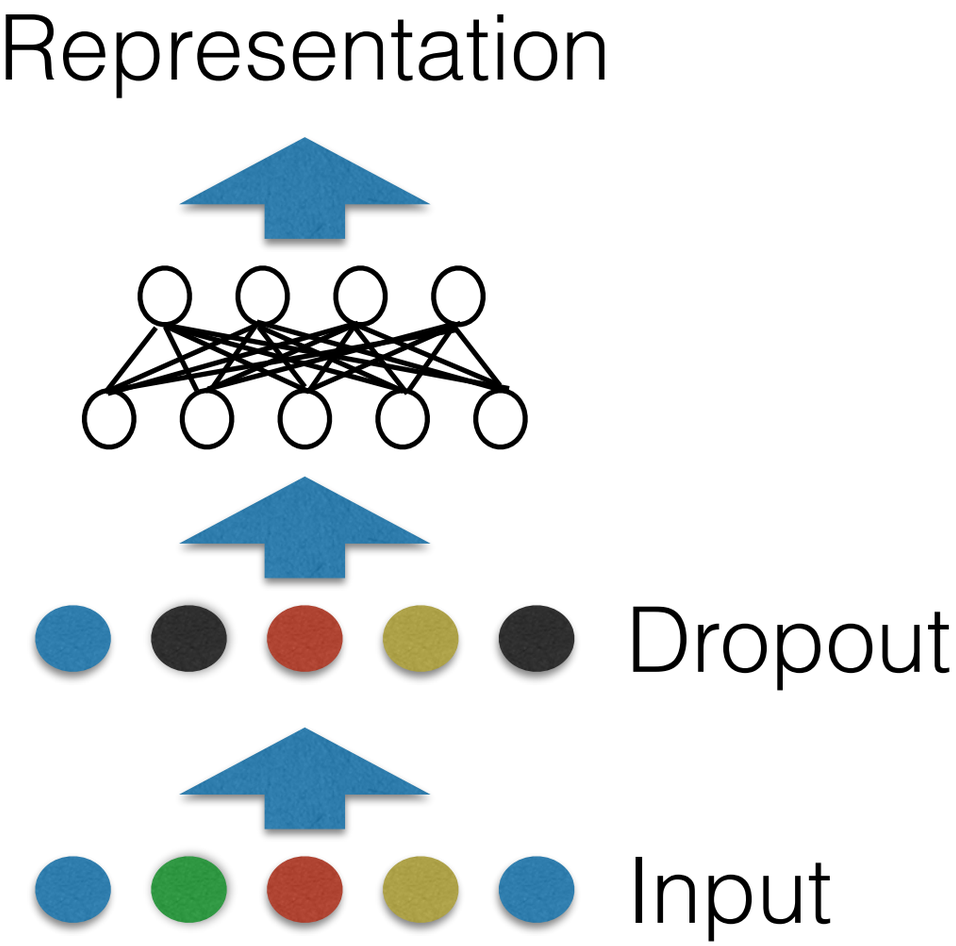
Learning representations

- What can we use these representations for?
- Transfer learning
    - Pure transfer
    - Pretraining
- Compression

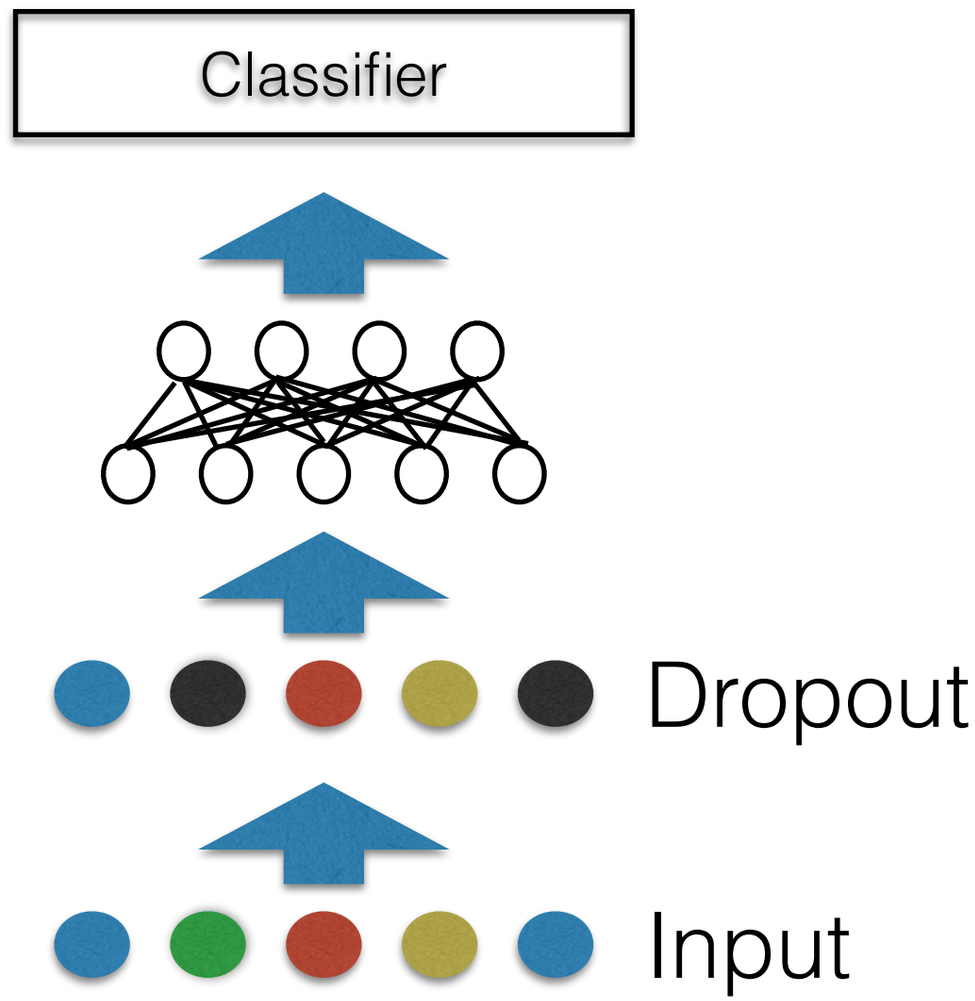
Pretraining and transfer



Pretraining and transfer



Pretraining and transfer



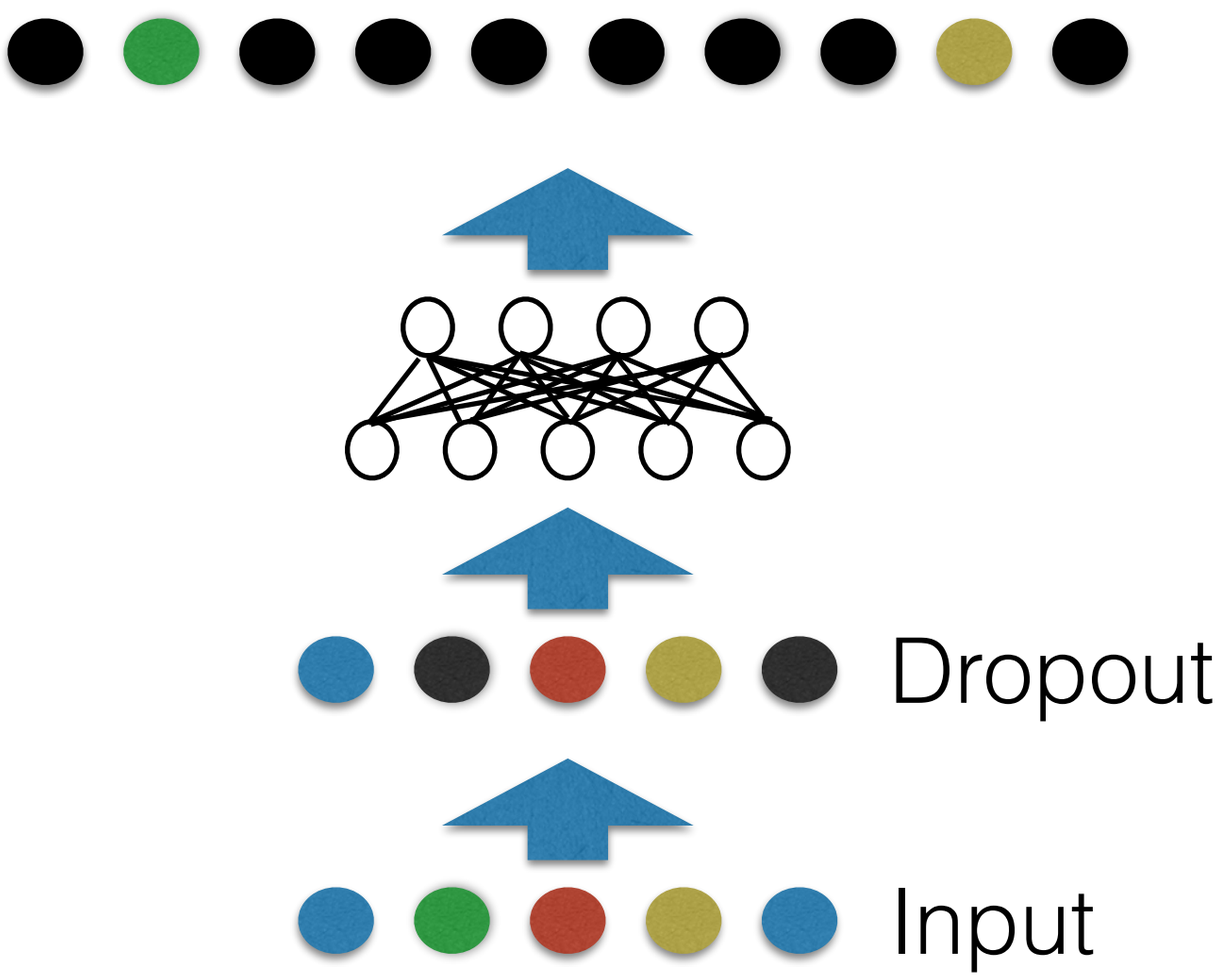
Hands on

Compression

- Sparse coding - we might enforce that inner representation is sparse. This is, the learned representation has very few active units.

- How?

Sparse coding



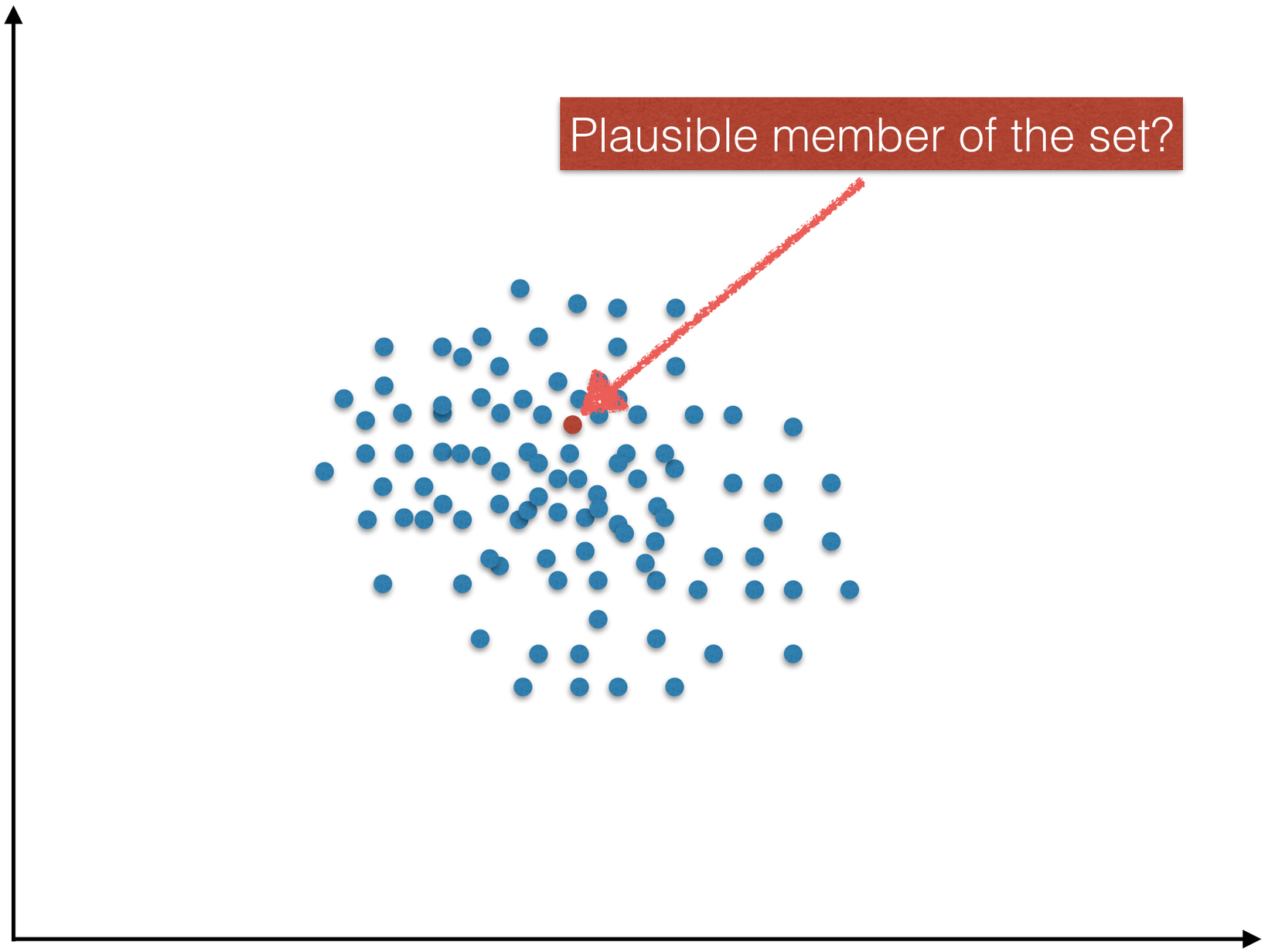
Hands on

Manifold learning

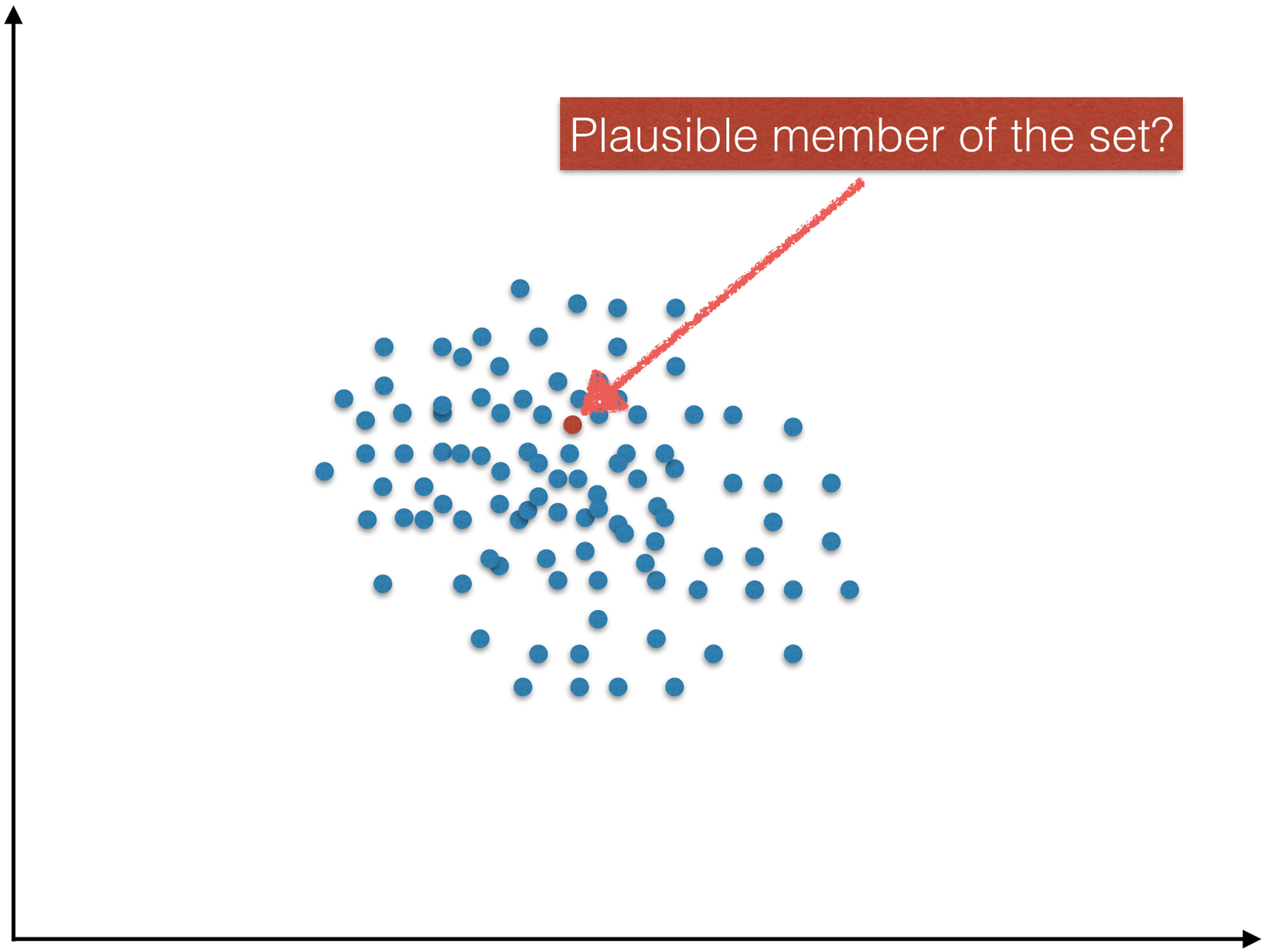
- From PCA to Non-linear dimensionality reduction

About manifolds

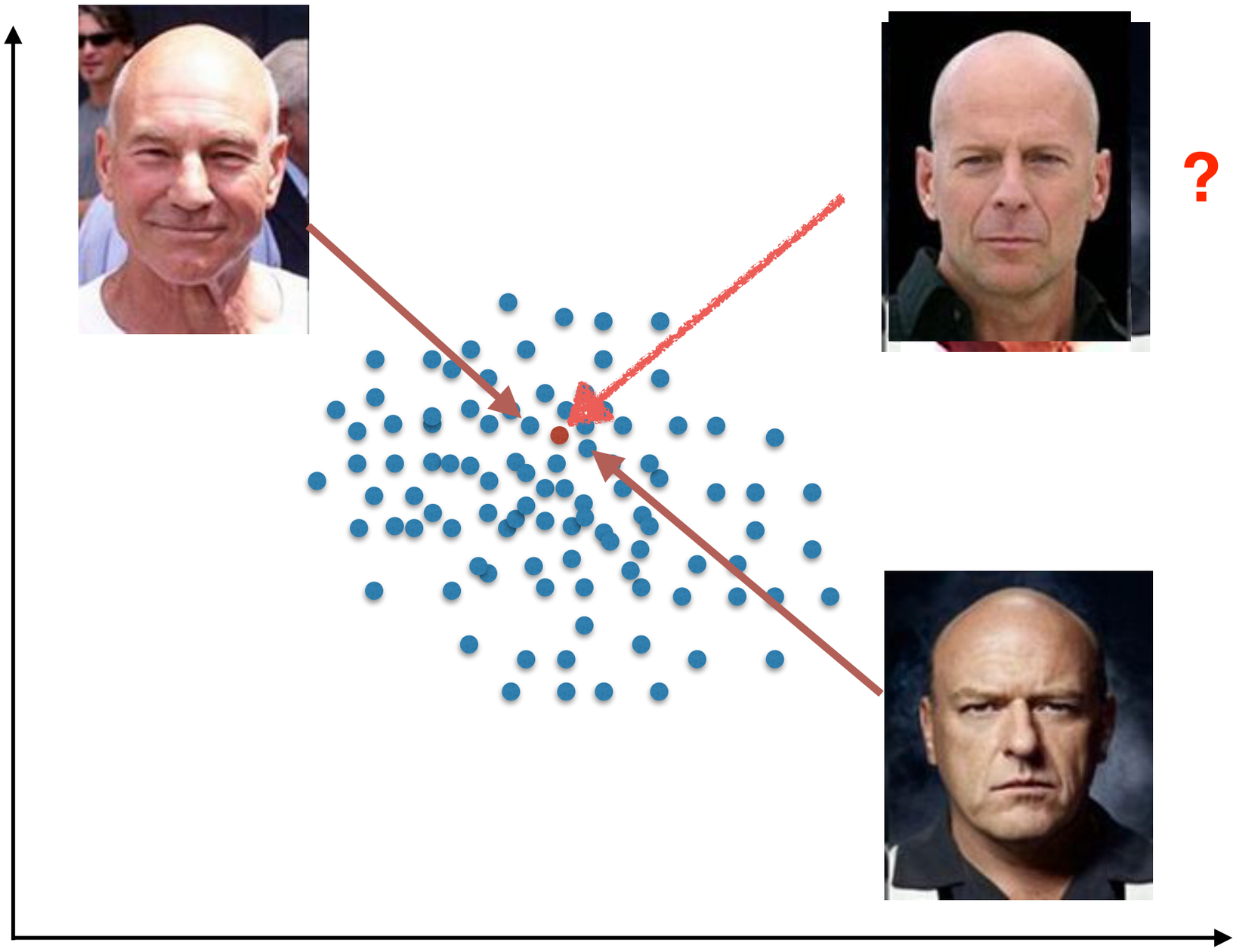
- Consider a general data space. Generally we accept that objects in this space that are close to the samples we have are together

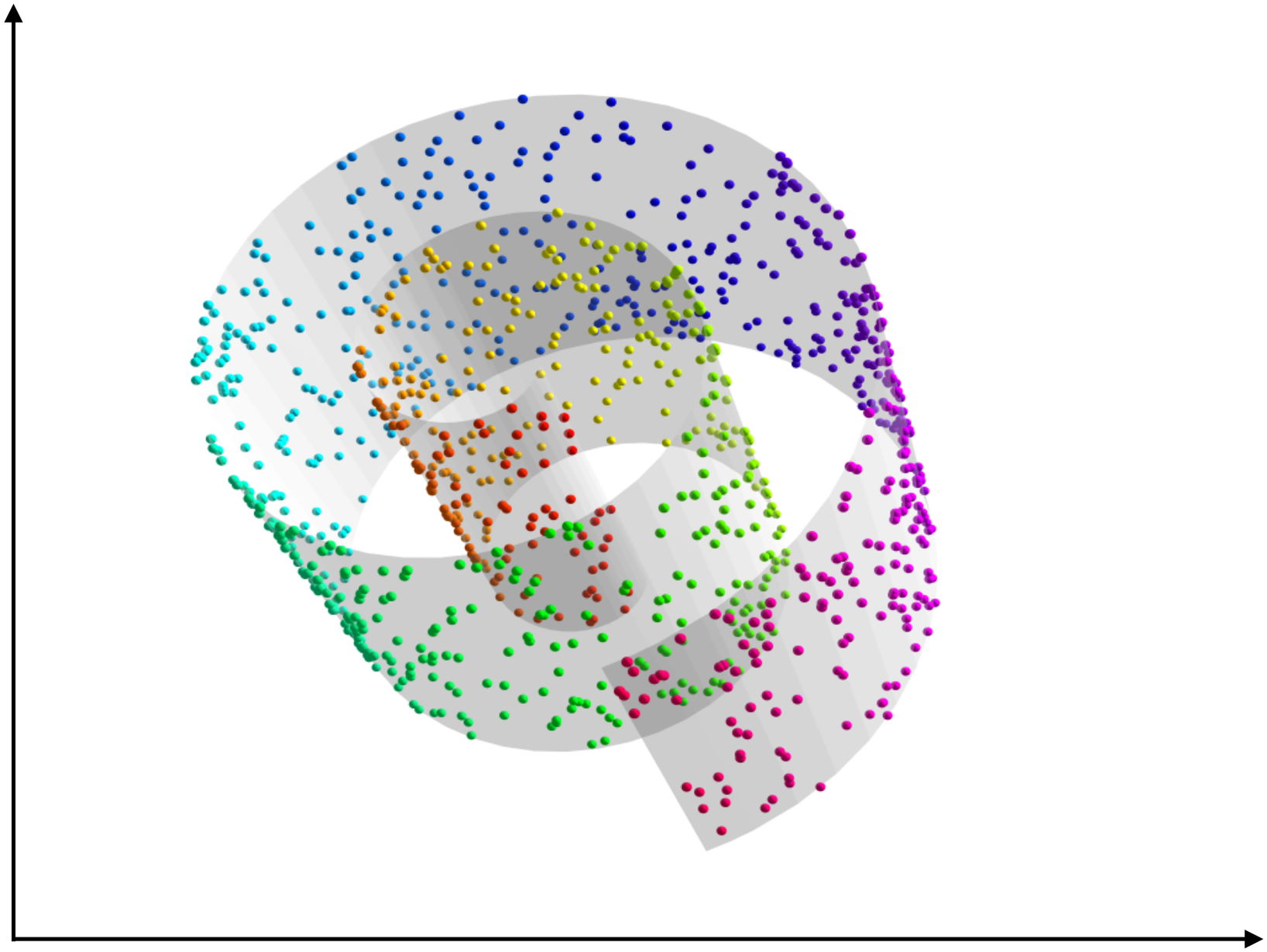

Plausible member of the set?

- Consider the space of images.


Plausible member of the set?

The Face Space

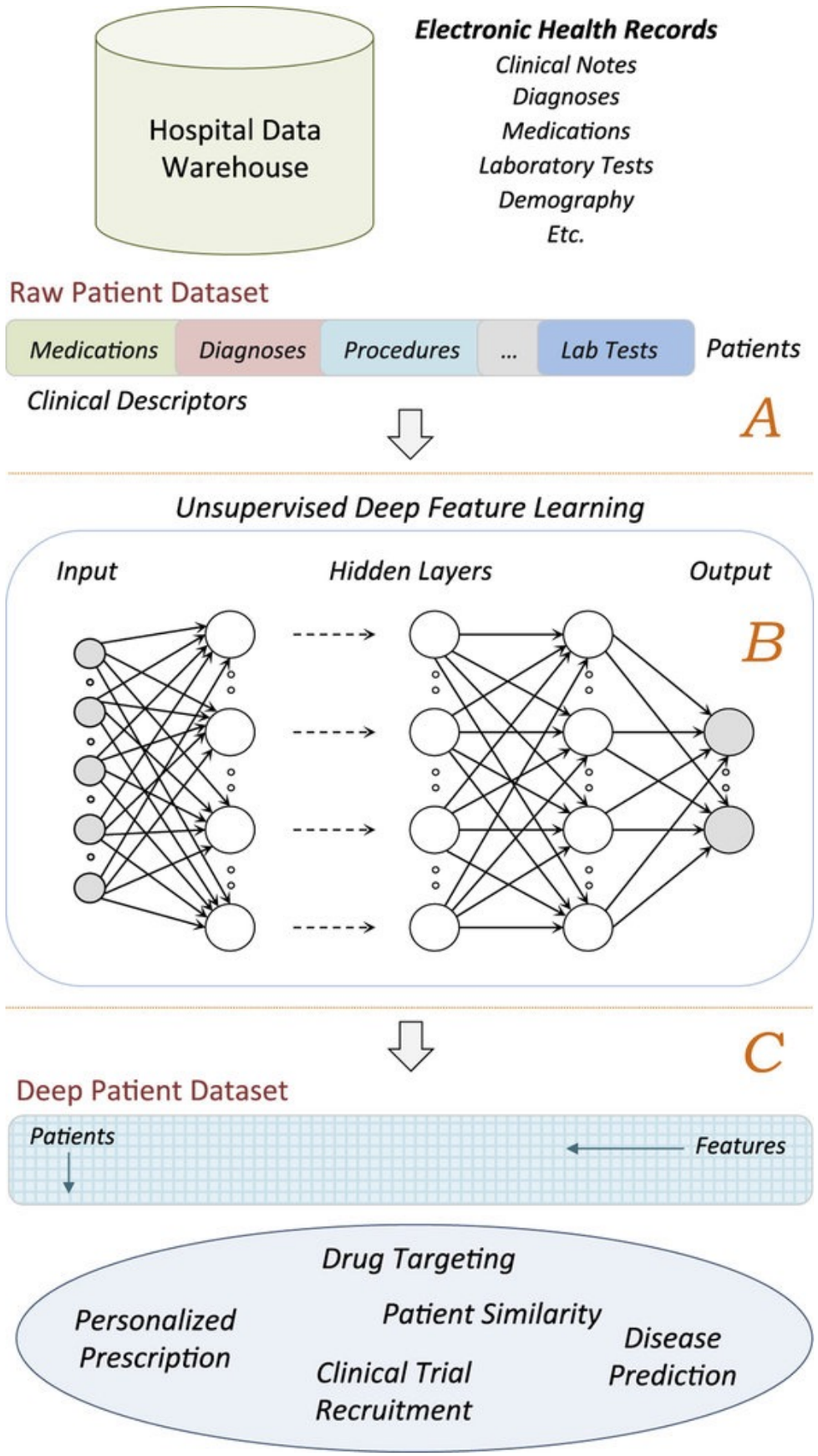




Hands on

Application

- Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records


Hospital Data Warehouse
Electronic Health Records
Clinical Notes
Diagnoses
Medications
Laboratory Tests
Demography
Etc.
Raw Patient Dataset
Medications
Diagnoses
Procedures
...
Lab Tests
Patients
Clinical Descriptors
A
Unsupervised Deep Feature Learning
Input
Hidden Layers
Output
B
C
Deep Patient Dataset
Patients
Features
Drug Targeting
Personalized Prescription
Patient Similarity
Disease Prediction
Clinical Trial Recruitment