UNSUPERVISED LEARNING

LECTURE Ø 23/09/2024

ALEX RODRIGUEZ alejandro.rodriguezgarcia@units.it
via economo 12/3 3rd floor

ILARIA VASCOTTO

Practical things

Theoretical lectures Monday 9-12 4A H_2 bis

Lab Thursday 9-11 a.m. T-C D

Teams code: ps85zjm

Prerequisites:

+ Python
+ Basic algèbra

CONTENTS:

→ Basic notions about UL

→ Dimensionality reduction methods:
- General theory
- Classical Methods
- Advanced Methods

→ Intrinsic Dimension estimation

→ Density estimation:
- Histograms
- Kernel Density estimation
- k-NN " "

→ Clustering
- General Theory
- Classification of methods
- Classical algorithms
- New methods
- Validation

Lectures Nov 28th

Lab: Code methods
Understanding (efficacy) >> Fast programs (Efficiency)

Scikit-learn

Python Notebooks

Evaluation:

1. Project. 10 minutes presentation 12 points
   - (a) Quality of the presentation (time) 3 points
   - (b) Understanding of the method (implementation) 4 points
   - (c) Parameters of the method, assumptions, limitations 2 points
   - (d) Comparison with other methods from the lectures (3 points)

2. Questions

- Easy (8)
  - Medium (6)
    - Difficult (4)
    - Medium (4)
  - Easy (6)
    - Medium (4)
    - Easy (4)

Materials:

- Lecture Notes
- Recording
- Papers / Reviews

Artificial Intelligence

↳ Machine Learning $<$ Shallow / Deep learning

if yellow then
    if eats meat
    else tiger
        giraffe
else

Supervised Learning $<$ Classification / Regression $\}$ $P(y \mid \underline{x})$

y: class / function value; $\underline{x}$ → features

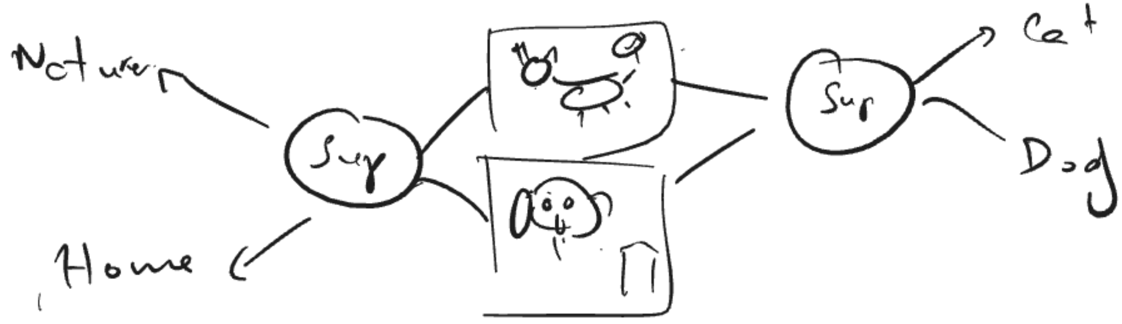
Unsupervised Learning $P(X)$

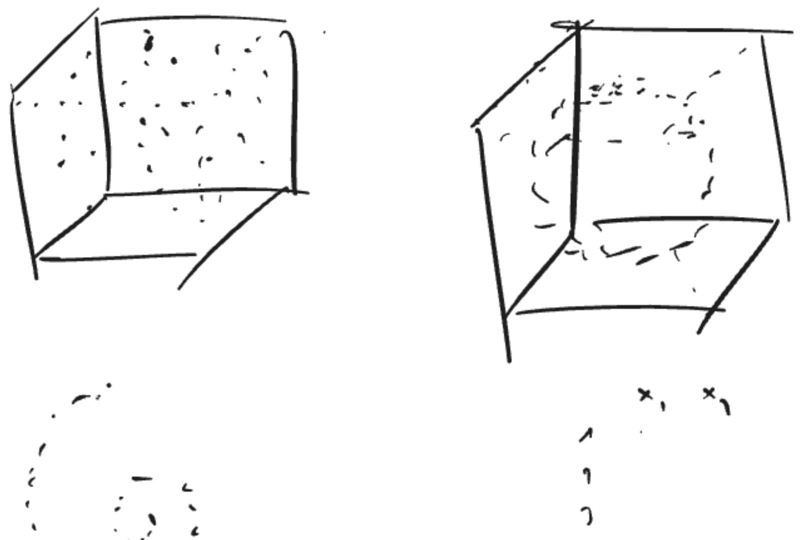
Reinforcement Learning $p(y \mid x)$ & $p(r)$

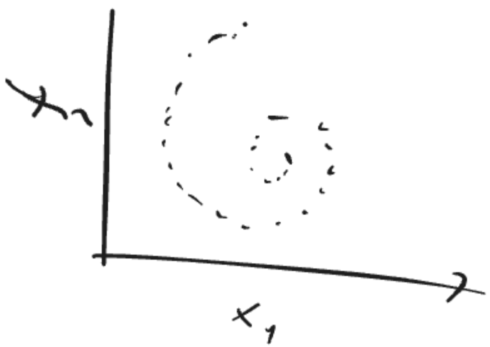
What do we learn on UL?

Patterns of $p(\underline{x})$

$D = 1280 \times 960 \rightarrow 1228800$ dimensions







Manifold learning

Manifold dimension (?) $\rightarrow$ Intrinsic dimension estimation

$D \longrightarrow d$ $\qquad$ d ?

$\mathbb{R}^D \rightarrow \mathbb{R}^d \rightarrow f[\mathbb{R}^D \rightarrow \mathbb{R}^d] \rightarrow$ Dimensionality Reduction

Notation

D Embedding dimension

$\{X\} \to x^i$ vector of features defining the i-th point

$x^i_j \quad j = 1 \ldots D$

$\mathbb{X}$ = matrix of dimension $N \times D$

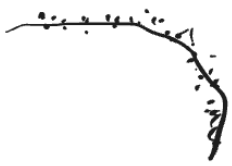
d : projection dimension

$Y \quad y^i$

$y^i_j \quad j = 1 \ldots d \qquad \mathbb{Y} \qquad N \times d$

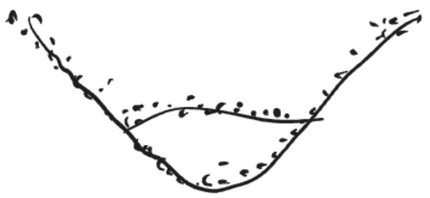
Some considerations

→ Manifold hypothesis is an approx.



→ Intrinsic dimension is scale dependent

→ Is not alway possible to find a closed form for the projection



Direct estimation of $p(x)$

Density estimation

$p(x)$

$p(x) \Longleftrightarrow p(y|x)$

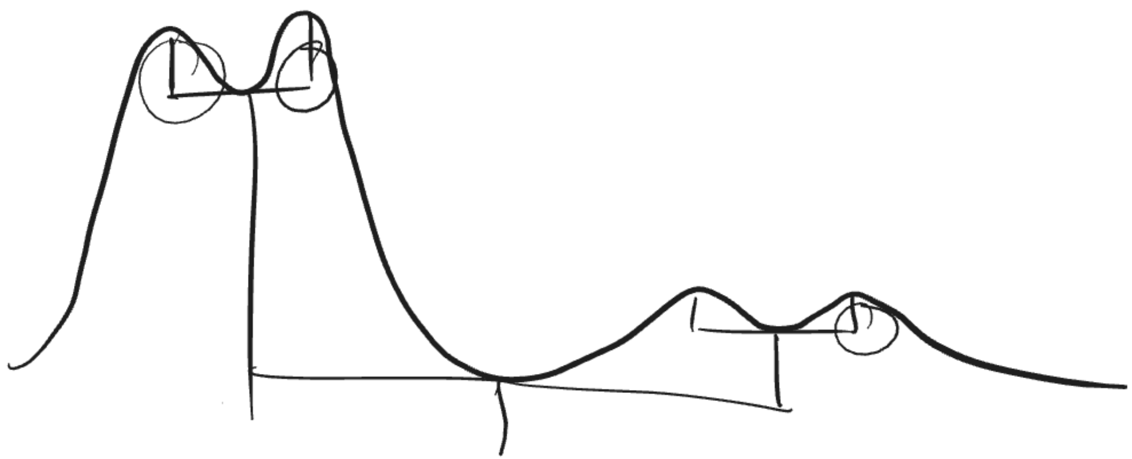
$p(x) \not\Longleftrightarrow p(y|x)$

Clustering

Can I divide my data into groups?

There exist groups in my data?

Hierarchies



- Anomaly detection $x^i \in \underline{p(x)}$?

- Generative models $p(x) \to \tilde{x}^i$

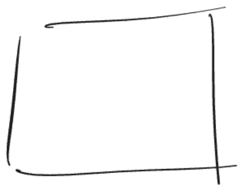
Course of dimensionality

Combinat. 2^2 $2^{10} \to 1024$

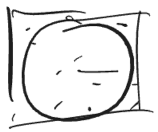
$2^{30} \to 1024 \times 1024 \times 1024$

Sampling

$\varepsilon_{p_i} \propto \frac{1}{\sqrt{N_i}}$



Distances



$p \quad \frac{n_0}{N} \sim \pi$

$$p = \frac{\pi^{D/2} r^D / (D/2)!}{(2r)^D}$$

$D \to \infty \to p \to \emptyset$