Labelling as an unsupervised learning problem

Imanol Perez
(Joint work with T. Lyons)

Mathematical Institute
University of Oxford

9th Berlin-Oxford Meeting

University of Oxford
Mathematical Institute

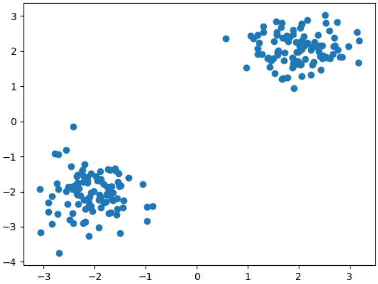
Oxford Mathematics

Unsupervised learning

- In machine learning, unsupervised learning is the task of unravelling hidden patterns in data without using labels assigned to each instance.

UNIVERSITY OF OXFORD
Mathematical Institute

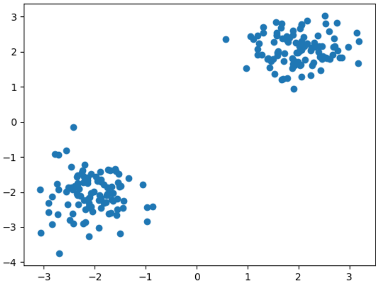
- In machine learning, unsupervised learning is the task of unravelling hidden patterns in data without using labels assigned to each instance.



Unsupervised learning

UNIVERSITY OF OXFORD
Mathematical Institute

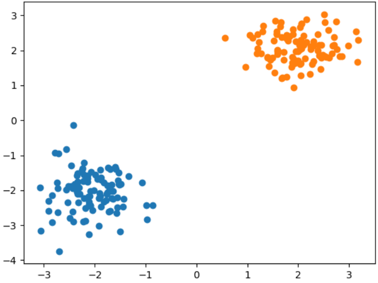
- An approach one could follow is to *cluster* the data, i.e. find a suitable partition of the dataset.



Unsupervised learning

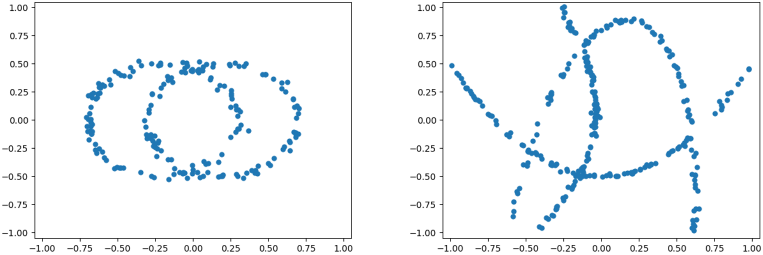
UNIVERSITY OF OXFORD
Mathematical Institute

- An approach one could follow is to *cluster* the data, i.e. find a suitable partition of the dataset.



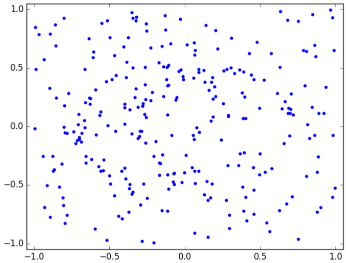
Unsupervised learning

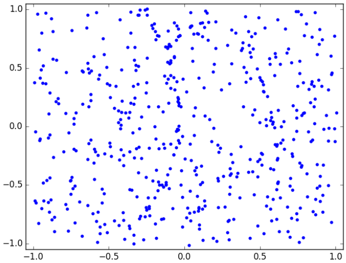
University of Oxford
Mathematical Institute



Unsupervised learning

UNIVERSITY OF OXFORD
Mathematical Institute





Labelling

- A cloud of points will share a label if unreasonably many of them share a common relationship.
- Points may have one, multiple or no labels.

- Let $\mathcal{X}$ be a topological space, called the *observation space*.

Labelling

Our framework

- Let $\mathcal{X}$ be a topological space, called the *observation space*.
- We will assume we have a *feature map* $\Phi : \mathcal{X} \to F$, with F a topological vector space (the *feature space*).

- Let $\mathcal{X}$ be a topological space, called the *observation space*.
- We will assume we have a *feature map* $\Phi : \mathcal{X} \to F$, with F a topological vector space (the *feature space*).
- $(\mathcal{X}, F, \Phi)$ will be called the *augmented observation space*.

- Let $\mathcal{X}$ be a topological space, called the *observation space*.
- We will assume we have a *feature map* $\Phi : \mathcal{X} \to F$, with F a topological vector space (the *feature space*).
- $(\mathcal{X}, F, \Phi)$ will be called the *augmented observation space*.
- Let μ be a probability measure on $\mathcal{X}$, called the *background noise*.

Mathematical Institute

- Let $\mathcal{X}$ be a topological space, called the *observation space*.
- We will assume we have a *feature map* $\Phi : \mathcal{X} \to F$, with F a topological vector space (the *feature space*).
- $(\mathcal{X}, F, \Phi)$ will be called the *augmented observation space*.
- Let μ be a probability measure on $\mathcal{X}$, called the *background noise*.
- The μ-augmented observation space is defined as the tuple $(\mathcal{X}, F, \Phi, \mu)$.

UNIVERSITY OF OXFORD
Mathematical Institute

Example (Polynomials in several variables)

In the case where the observation space is $\mathbb{R}^d$, we can define an augmented observation space by considering the feature map given by the mapping to a basis of polynomials of fixed degree.

Mathematical Institute

Example (Path space)

If $\mathcal{X} = \mathcal{V}^1([0, T]; \mathbb{R}^d)$, we can define an augmented observation space using the signature map as the feature map.

Labelling

UNIVERSITY OF OXFORD
Mathematical Institute

Definition

Given an augmented observation space $(\mathcal{X}, F, \Phi)$, we define the *space of potential labels* $\mathcal{F}_\Phi := \{\ell \circ \Phi : \ell \in F'\}$.

Definition

Given an augmented observation space $(\mathcal{X}, F, \Phi)$, we define the *space of potential labels* $\mathcal{F}_\Phi := \{\ell \circ \Phi : \ell \in F'\}$.

Definition (Label)

Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space, let $\mathcal{C} \subset \mathcal{X}$ and set $0 < \delta < 1$. A potential label $f \in \mathcal{F}_\Phi$ is called a (μ, δ)-label for $\mathcal{C}$ if there exists an interval $I \subset \mathbb{R}$ such that $0 \in I$, $f(\mathcal{C}) \subset I$ and

$$(f_*(\mu))(I) < \delta,$$

where $f_*(\mu)$ is the pushforward measure defined as $(f_*(\mu))(I) := \mu(f^{-1}(I))$.

Mathematical Institute

Definition

Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space, $\mathcal{C} \subset \mathcal{X}$ and $0 < \delta < 1$. The set of labels for $\mathcal{C}$ is defined as

$$\mathcal{L}_{\mu,\delta}(\mathcal{C}) := \{f \in \mathcal{F}_{\Phi} : f \text{ is a } (\mu, \delta)\text{-label for } \mathcal{C}\}.$$

Labelling

UNIVERSITY OF OXFORD
Mathematical Institute

- Typically, $|\mathcal{L}_{\mu,\delta}(\mathcal{C})| = \varnothing$ or ∞.

Labelling

Mathematical Institute

- Typically, $|\mathcal{L}_{\mu,\delta}(\mathcal{C})| = \varnothing$ or ∞.
- $\mathcal{L}_{\mu,\delta_1}(\mathcal{C}_1) \subset \mathcal{L}_{\mu,\delta_2}(\mathcal{C}_2)$ for all $0 < \delta_1 \leq \delta_2$, $\mathcal{C}_2 \subset \mathcal{C}_1 \subset \mathcal{X}$.

Labelling

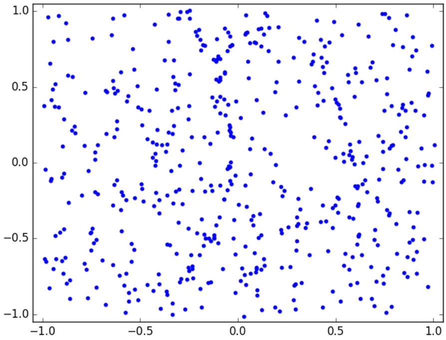
- Typically, $|\mathcal{L}_{\mu,\delta}(\mathcal{C})| = \varnothing$ or ∞.
- $\mathcal{L}_{\mu,\delta_1}(\mathcal{C}_1) \subset \mathcal{L}_{\mu,\delta_2}(\mathcal{C}_2)$ for all $0 < \delta_1 \leq \delta_2$, $\mathcal{C}_2 \subset \mathcal{C}_1 \subset \mathcal{X}$.
- If $f \in \mathcal{F}_\Phi$ is a (μ, δ)-label, then λf is a (μ, δ)-label for all $\lambda \in \mathbb{R} \setminus \{0\}$.

Labelling

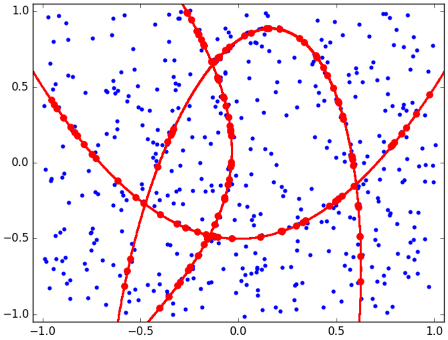
UNIVERSITY OF OXFORD
Mathematical Institute

- Typically, $|\mathcal{L}_{\mu,\delta}(\mathcal{C})| = \varnothing$ or ∞.
- $\mathcal{L}_{\mu,\delta_1}(\mathcal{C}_1) \subset \mathcal{L}_{\mu,\delta_2}(\mathcal{C}_2)$ for all $0 < \delta_1 \leq \delta_2$, $\mathcal{C}_2 \subset \mathcal{C}_1 \subset \mathcal{X}$.
- If $f \in \mathcal{F}_\Phi$ is a (μ, δ)-label, then λf is a (μ, δ)-label for all $\lambda \in \mathbb{R} \setminus \{0\}$.
- However, $\mathcal{L}_{\mu,\delta}(\mathcal{C})$ is not a vector space in general.

Examples

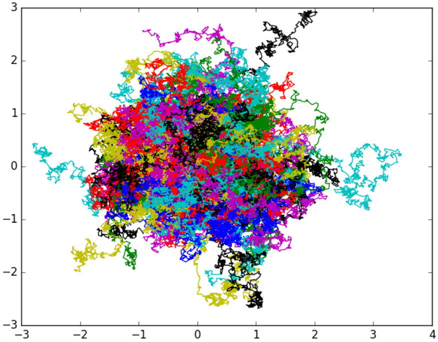


Examples



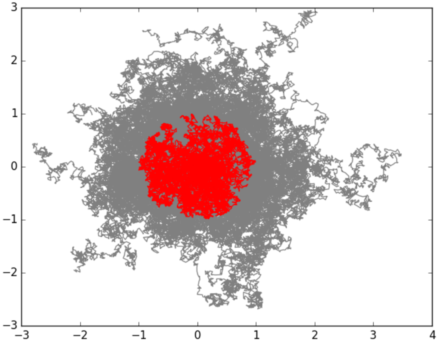
Examples

UNIVERSITY OF OXFORD
Mathematical Institute



Examples

UNIVERSITY OF OXFORD
Mathematical Institute



Label: $f(X) = \max_t |X_t| - 1.$

False discoveries

UNIVERSITY OF OXFORD
Mathematical Institute

- Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space.
- Let $\mathcal{C} \subset \mathcal{X}$ be a finite set of points independently distributed according to μ.

False discoveries

- Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space.
- Let $\mathcal{C} \subset \mathcal{X}$ be a finite set of points independently distributed according to μ.
- Given $\delta > 0$, what is the probability that there exists a (μ, δ)-label for $\mathcal{C}$?

False discoveries

UNIVERSITY OF OXFORD
Mathematical Institute

- Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space.
- Let $\mathcal{C} \subset \mathcal{X}$ be a finite set of points independently distributed according to μ.
- Given $\delta > 0$, what is the probability that there exists a (μ, δ)-label for $\mathcal{C}$?
- How does this depend on the sample size? And on δ, the feature map or the background noise?

False discoveries

UNIVERSITY OF OXFORD
Mathematical Institute

- Assume F is a D-dimensional Hilbert space with inner product $(\cdot, \cdot)_F$.

False discoveries

University of Oxford
Mathematical Institute

- Assume F is a D-dimensional Hilbert space with inner product $(\cdot,\cdot)_F$.
- Define the linear compact operator

$$
\begin{aligned}
\Gamma_\Phi : F' &\to \mathbb{R}^{|\mathcal{C}|} \\
\ell &\mapsto (\langle \ell, \Phi(X)\rangle)_{X\in\mathcal{C}}.
\end{aligned}
$$

False discoveries

UNIVERSITY OF OXFORD
Mathematical Institute

- Assume F is a D-dimensional Hilbert space with inner product $(\cdot,\cdot)_F$.
- Define the linear compact operator

$$
\begin{aligned}
\Gamma_\Phi : F' &\to \mathbb{R}^{|\mathcal{C}|} \\
\ell &\mapsto (\langle \ell, \Phi(X)\rangle)_{X\in\mathcal{C}}.
\end{aligned}
$$

- We want to study what is the probability that there exists $\ell \in F'$ such that $\{\langle \ell, \Phi(X)\rangle\}_{X\in\mathcal{C}}$ is contained in a *small* interval.
- Thus, we have to study the spectrum of the random matrix Γ_Φ.

UNIVERSITY OF OXFORD
Mathematical Institute

Theorem

Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space, with $\mathcal{X}$ a Banach space. Assume that

(i) $\mathbb{E}_\mu(\Phi(X)^*\Phi(X)) = I$.

(ii) $\|\Phi(x)\|_F \leq 1 + \|x\|_{\mathcal{X}}^k$ *for all* $x \in \mathcal{X}$, *where* $k > 0$.

(iii) $\mathbb{P}_\mu(\|\Phi(X)\|_F > \lambda) \lesssim \exp(-C\lambda^\alpha)$, *where* $\alpha \geq k$.

Then, $s_{min}(\Gamma_\Phi) \sim \sqrt{|\mathcal{C}|}$ *with probability higher than*
$1 - \exp(-p_{D,|\mathcal{C}|})$, *with* $p_{D,|\mathcal{C}|} \xrightarrow{|\mathcal{C}|\to\infty} \infty$.

University of Oxford
Mathematical Institute

Theorem (Probability of false discovery)

Let $(\mathcal{X}, F, \Phi, \mu)$ be a μ-augmented observation space and let $\mathcal{C} \subset \mathcal{X}$ be a finite set of points independently distributed according to μ. Assume that the feature map satisfies the assumptions of the previous theorem.
Then, give any $\epsilon > 0$, there exists $\delta_{\epsilon,|\mathcal{C}|} > 0$ such that the probability that there exists a $(\mu, \delta_{\epsilon,|\mathcal{C}|})$-label for $\mathcal{C}$ is less than $(C/\epsilon)^D \exp(-p_{D,|\mathcal{C}|})$, with $C > 0$ a constant.

UNIVERSITY OF OXFORD
Mathematical Institute

Proof (sketch)

We want to estimate:

$$\mathbb{P}_\mu(\exists\ (\mu,\delta) - \text{label for } \mathcal{C})$$

Proof (sketch)

We want to estimate:

$$\mathbb{P}_\mu(\exists\ (\mu,\delta)-\text{label for } \mathcal{C}) = \mathbb{P}_\mu\left(\bigvee_{f\in\mathcal{F}_\Phi} f \text{ is a } (\mu,\delta)-\text{label for } \mathcal{C}\right)$$

UNIVERSITY OF OXFORD
Mathematical Institute

Proof (sketch)

We want to estimate:

$$\mathbb{P}_\mu(\exists\ (\mu,\delta) - \text{label for } \mathcal{C}) = \mathbb{P}_\mu\left(\bigvee_{f\in\mathcal{F}_\Phi} f \text{ is a } (\mu,\delta) - \text{label for } \mathcal{C}\right)$$

- Obtain estimates for $\mathbb{P}_\mu(f \text{ is a } (\mu,\delta) - \text{label for } \mathcal{C})$, for a fixed $f \in \mathcal{C}$.
- Use an ϵ-net argument to restrict $\mathcal{F}_\Phi$ to some finite subset.

References

UNIVERSITY OF OXFORD
Mathematical Institute

Lyons, T. and Perez Arribas, I. "Labelling as an unsupervised learning problem." *ArXiv:1805.03911 (2018)*.