Semi-supervised learning: Between Supervised and Unsupervised Learning

Kristiaan Pelckmans

Division of Systems and Control (SysCon)
Department of Information Technology (IT)
Uppsala University (UU)
Sweden.

January 17, 2018

kp@it.uu.se

Overview

Today:

1. *Supervised and Unsupervised*
2. *Semi-Supervised Learning*
3. *Active Learning*
4. *Progressive Learning*
5. *Conclusions*
6. *VASCO*

kp@it.uu.se

Overview (Ct'd)

Caveats:

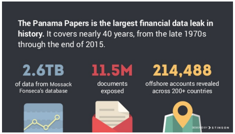
1. *Conceptual discussion.*
2. *No Deep nor Shallow models.*
3. *Scientific literature generally on the linear PERCEPTRON (or [SBD, 2009], [AB, 2017])*
4. *Density $\approx$ Generative models.*

Discussions:

1. *Prediction versus Recovering.*
2. *Progressive/Incremental learning.*
3. *(Anomaly) Detection.*
4. *Unmixing (latent variable).*
5. *Calibration (CV, Pokemons, ...).*

kp@it.uu.se

Supervised and unsupervised Learning (Ct'd)



$\downarrow$

$\downarrow$

ML paradigm:

1. *Encode* $\mathbf{X}_t \in \mathbb{R}^d$.
2. *Define* Y_t.
3. *Shape of* f.
4. *Define good:* $\approx \to \ell$.
5. *Define 'simple' (prior):* $\Omega(f)$.
6. *Dataset* $\{(\mathbf{X}_t, Y_t)\}_{t=1}^n$.
7. Optimize (fit) so that $f(\mathbf{X}_t) \approx Y_t$.
8. Calibrate/evaluate.

$$\min_{f \in \{f\}} \mathrm{Risk}_n(f) = \sum_{t=1}^{n} \ell\left(f(\mathbf{X}_t) \approx Y_t\right) + \gamma \Omega(f)$$

$\to$ Cannot do better than (encoding of) data!

kp@it.uu.se

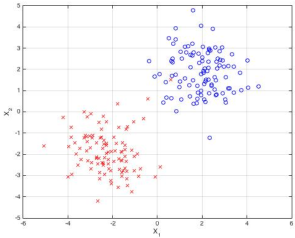
Supervised and unsupervised Learning (Ct'd)

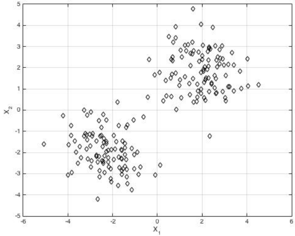
But:

1. *Getting labels is expensive*
2. *... and boring.*
3. *Are there principles for a shortcut?*

Yes ...

Supervised and unsupervised Learning (Ct'd)





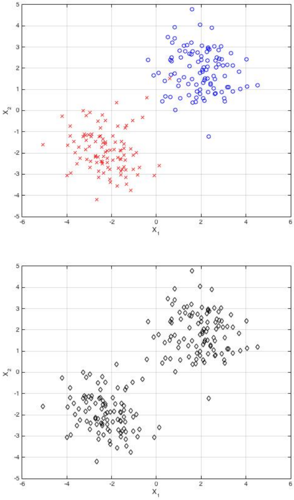
Labels Y_t sampled from $p(\cdot|\mathbf{X}_t)$:

1. *Supervised learning:*
   $\{(\mathbf{X}_t, Y_t)\}_{t=1}^n \subset \mathbb{R}^d \times \{-1, +1\}$
   $\to f(\mathbf{X}_t) \approx Y_t$
2. *Unsupervised learning:*
   $\{\mathbf{X}_t\}_{t=1}^n \subset \mathbb{R}^d$
   $\to f(\mathbf{X}_t) \approx ?$

$? =$

- Detection.
- Finding clusters.
- Nice 2/3D visualisation.
- Find latent (stable/causal) components.
- Concise explanation.
- Auto-encoder.
- ...

Supervised and unsupervised Learning (Ct'd)



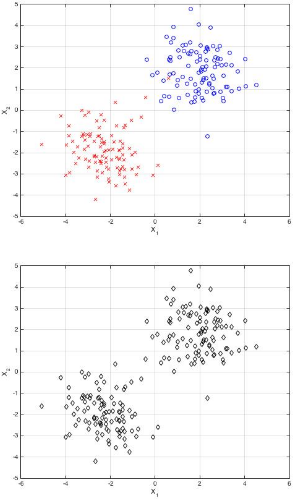
Principles for learning $\mathbf{X}_t \to Y_t$ by only having $\{\mathbf{X}_t\}$?

1. *Reduce range of possible* $\{f\}$.
2. *If* $\mathbf{X}_i \sim \mathbf{X}_j$, *then* $Y_i \sim Y_j$.
3. *Restrict space of interest of* $\mathbf{X}_t$.
4. *Estimate density* $p(\mathbf{X}_t)$.

$\to$ But no inherent implication on $\mathbf{X}_t \to Y_t$.

kp@it.uu.se

Supervised and unsupervised Learning (Ct'd)

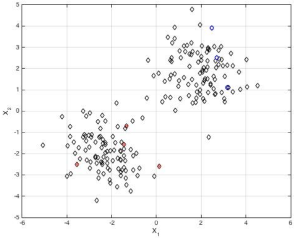


No natural notion of 'fit', but complexity control:

1. *Clustering.*
2. *Margin-based learning.*
3. *Graph-based approaches.*
4. *Subspace (manifold) learning.*
5. *Density estimation of* $\mathbf{X}_t$.

Semi-Supervised Learning

Semi-Supervised Learning

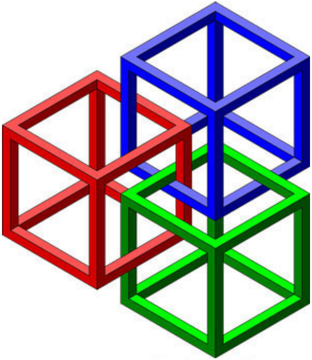


Risk for semi-supervised learning:

$$\text{Risk}_n(f) = \sum_{\sigma_t = 1} \ell\left(f(\mathbf{X}_t) \approx \sigma_t Y_t\right) + \gamma \Omega(f)$$

1. *Semi-supervised learning:* $\{(\mathbf{X}_t, \sigma_t Y_t)\}_{t=1}^n \subset \mathbb{R}^d \times \{-1, 0, +1\}$ $\to f(\mathbf{X}) \approx Y$
2. *Transductive learning:* $\{(\mathbf{X}_t, \sigma_t Y_t)\}_{t=1}^n \subset \mathbb{R}^d \times \{-1, 0, +1\}$ $\to \{y_t\}_{t=1}^n \subset \{-1, +1\}$
3. *Principles behind $\Omega(f)$?*

Semi-Supervised Learning (Ct'd)



Framework:

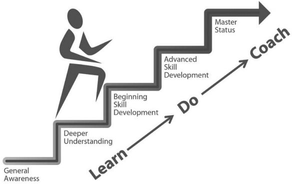
1. *Frequentist (there is one and only one truth).*
2. *Bayesian (the truth is a random sample).*
3. *Adversarial (data is arbitrary).*
4. Computational.

kp@it.uu.se

Alternative Learning Models.

1. *Active learning.*
2. *Selective sampling.*
3. *Learning with missing variables.*
4. *Matrix completion.*
5. *Boosting.*

Progressive Learning



1. *The more labels we get, the more complex bits can be learnt.*
2. *Learning-by-mistakes.*
3. *Discovery and detection.*

(video)

Conclusions

1. Labelling is expensive.
2. Inverse problem (understanding) or optimal prediction (engineering).
3. Extending theory behind observations.
4. Theory vs. empirical.
5. Eliminate possible explanations by counter-examples.

kp@it.uu.se

UPPSALA
UNIVERSITET

VASCO

1. Vanishing & Disappearing objects.
2. By cross-matching catalogues (USNO, 1950s, pannSTAR, 20xx, GAIA, 201x).
3. Identifying candidates for followup (BI, ESA, ...).
4. With ML to handle size.
5. Kick-off Feb. 2018.