



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

Overview



Today:

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



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, ...).*

Supervised and unsupervised Learning (Ct'd)



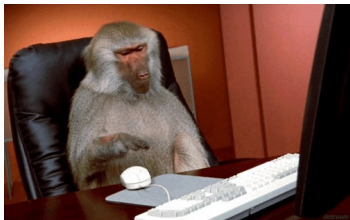
ML paradigm:

1. Encode $\mathbf{X}_t \in \mathbb{R}^d$.
2. Define Y_t .
3. Shape of f .
4. Define good: $\approx \rightarrow \ell$.
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.

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

→ Cannot do better than (encoding of) data!

Supervised and unsupervised Learning (Ct'd)

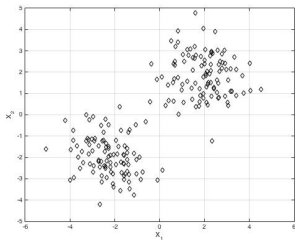
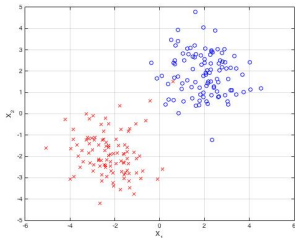


But:

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:*

$$\{(\mathbf{X}_t, Y_t)\}_{t=1}^n \subset \mathbb{R}^d \times \{-1, +1\}$$
$$\rightarrow f(\mathbf{X}_t) \approx Y_t$$

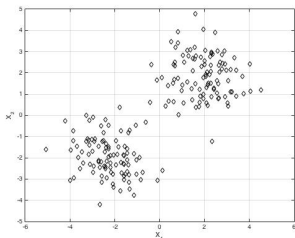
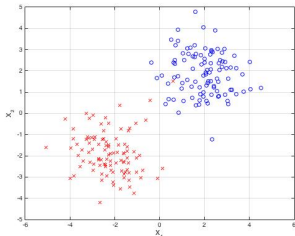
2. *Unsupervised learning:*

$$\{\mathbf{X}_t\}_{t=1}^n \subset \mathbb{R}^d$$
$$\rightarrow f(\mathbf{X}_t) \approx ?$$

? =

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

Supervised and unsupervised Learning (Ct'd)

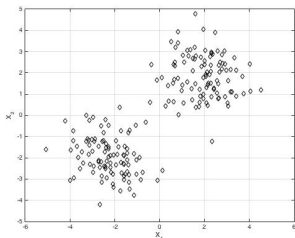
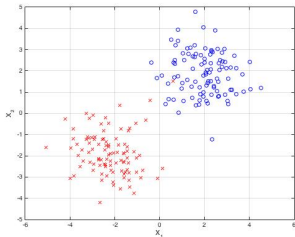


Principles for learning $\mathbf{X}_t \rightarrow Y_t$ by only having $\{\mathbf{X}_t\}$?

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 $p(\mathbf{X}_t)$.

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

Supervised and unsupervised Learning (Ct'd)



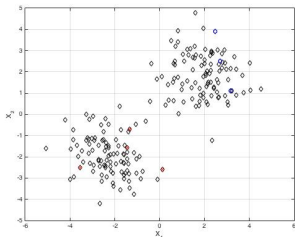
No natural notion of 'fit', but complexity control:

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

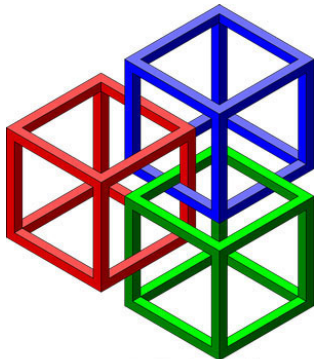


Risk for semi-supervised learning:

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

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

Semi-Supervised Learning (Ct'd)



ComputerHope.com

Framework:

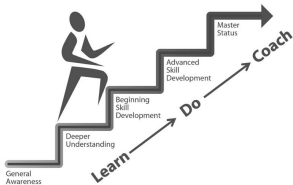
1. *Frequentist (there is one and only one truth).*
2. *Bayesian (the truth is a random sample).*
3. *Adversarial (data is arbitrary).*
4. Computational.

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.



THE VASCO NETWORK

ABOUT

"Any sufficiently advanced technology is indistinguishable from magic" - Arthur C. Clarke's Third Law.

The Vanishing & Appearing Sources during a Century of Observations (VASCO) project aims at finding astro-physically interesting mismatches between historical sky surveys: Which object flickered out from our celestial radar?, Which locations hint at astronomical spectacles to discover?, Which sky regions are displayed distinctly in different sky surveys?

Answers to these questions could lead to interesting scientific findings, like new astrophysical phenomena or to — who knows — interesting targets for follow-up SETI observations. The sheer scale of current existing sky surveys prompts us to develop new computational tools, with a glamorous role given to machine learning and artificial intelligence. The project gives a forum for interaction between professional scientists with various background, serious amateurs and curious citizen scientists.

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.