Imprecision in learning: introduction

Sebastien Destercke

Université de Technologie de Compiègne

Classical framework

1. A set **D** of (i.i.d.) **precise** data $\{x_i, y_i\}$ coming from $\mathscr{X} \times \mathscr{Y}$
2. Future data follow the **same** distribution D over $\mathscr{X} \times \mathscr{Y}$
3. A **precise** cost/reward $c_\omega(y)$ of predicting ω
4. Search for a model $M^* : \mathscr{X} \to \mathscr{Y}$

   $$M^* = \arg\min_{M \in \mathscr{M}} \sum_i c_{M(x_i)}(y_i)$$

   within a set $\mathscr{M}$
5. Producing **precise** predictions

Each assumption has been questioned in the past → in which case are IP approaches relevant ?

Imprecise prediction : what exists

Different approaches beyond IP :

- rejection or partial rejection using SVM, probabilistic thresholds
- conformal prediction (Vovk, Shafer, Gammerman)

Despite their possible efficiency, remain a minor field of activity

Imprecise prediction : perspectives/challenges

- make efficient imprecised predictions of complex structures
  - Graphs (block-clustering, social network analysis)
  - Preferences/recommendations (Angela Talk)
  - Multi-label data or multi-task problems
  - Sequences
- how to evaluate the different models ?
- what to do with the imprecise prediction once we have it ?

Cost of imprecision

Predict the rate someone would give a movie : **v**ery **b**ad, **b**ad, **g**ood, **v**ery **g**ood

Cost		Truth: vb	b	g	vg
Prediction	vb	0	1	2	3
	b	1	0	1	2
	g	2	1	0	1
	vg	3	2	1	0

Predictions "further away" from truth worse

Imprecise costs

Cost		Truth			
		vb	b	g	vg
Prediction	vb	0	1	2	3
	b	1	0	1	2
	g	2	1	0	1
	vg	3	2	1	0
	{vb,b}	?	?	?	?
	{vb,b,g}	?	?	?	?

How to fill up the matrix so that

- we can evaluate imprecise predictions
- we can learn efficiently a model that minimizes our cost

Non-identically distributed

- many problems where training $\{x_i, y_i\}$ is assumed to follow distribution D_1, but where new incoming data (of which you may or not have samples) may follow distribution D_2
  - Transfer learning (imprecise transport problem ?)
  - Concept drift
- can imprecise probability helps here ?
- some paper looking at ill-specified prior (Minimax Regret Classifier for Imprecise Class Distributions)

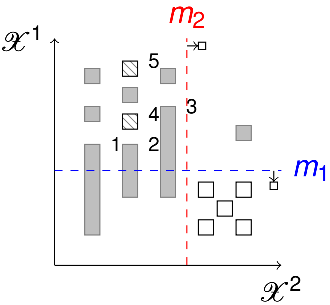
Imprecise data and models

- data $\{X_i, Y_i\}$ are now imprecise, i.e. $X_i \subseteq \mathscr{X}$, $Y_i \subseteq \mathscr{Y}$
- best model

$$M^* = \arg\min_{M \in \mathscr{M}} \sum_i c_{M(x_i)}(y_i)$$

no longer well-defined.

illustration



$$[\underline{R}(m_1),\overline{R}(m_1)] = [0,5]$$
$$[\underline{R}(m_2),\overline{R}(m_2)] = [1,3]$$
$$\inf R(m_1) - R(m_2) = -1$$
$$\inf R(m_2) - R(m_1) = -2$$

Imprecise data and models : some issues

1. Should we learn a set of models, or only one model ?
   - in the first case, how to learn it efficiently and in a compact way ? (taking every replacement not possible)
   - in the second case (most common in literature), what decision rule to pick ? Being optimistic (minimin) or pessimistic (maximin)
2. Under what assumptions about the imprecisiation process does the (optimal) model remain identifiable (Thomas talk ?)

Imprecise data and models : some issues

3. If model not identifiable (sets of possible model)
   - which features or labels among the data $\{X_i, Y_i\}$ should we query to improve the most our model (active learning)
   - in this case, can what we learn about the imprecisiation process help as well ?
4. Can the imprecisiation of the data provide more robust models ? → e.g., if we have few data