ISML-II: Machine Learning **Spring 2014**

Lecture 22- A Glimpse Beyond the Fence

Lecturer: Lorenzo Rosasco

We next try to give brief overview of 1) topics in machine learning that we have not touched upon, 2) some of the current and future challenges in machine learning.

22.1 Different Kinds of Data

Different machine learning approaches arise to deal with different kinds of input and output. Recall that, the input/output pairs are assumed to belong to an input space X and an output space Y, respectively. We call $Z = X \times Y$ the data space. We list a few examples of input and output spaces.

- Euclidean/Vector Spaces. Perhaps the simples example, covering many practical situations is $X = \mathbb{R}^d$, $d \in \mathbb{N}$.
- Probability distributions. We could set $X = \{x \in \mathbb{R}^d_+ \ : \ \sum_{j=1}^d x^j = 1$, $d \in \mathbb{N}$, and view elements of the space as probability distributions on a finite set Ω of dimension d. More generally given any probability space Ω we can view X as the space of probability distribution on Ω.
- Strings/Words. Given an alphabet Σ of symbols (letters) one could consider $X = \Sigma^p$, $p \in \mathbb{N}$, the (finite) space of strings (words) of p letters.
- Graphs. We can view X as collection of graphs, i.e. $X = \{\}$.

Clearly more exotic examples can be constructed considering compositions of the above examples, for example $X = \mathbb{R}^d \times \Sigma^p$, $d, p \in \mathbb{N}$ etc.

Next we discuss different choices of the output space and see how they often corresponds to problems with different names.

- Regression, $Y = \mathbb{R}$.
- Binary classification, $Y = \{-1, 1\}$. Where we note that here we could have taken $Y = \{0, 1\}$– as well as any other pair of distinct numeric values.
- Multivariate regression, $Y = \mathbb{R}^T$, $T \in \mathbb{N}$, each output is a vector.
- Functional regression, Y is a Hilbert space, for example each output is a function.
- Multi-category classification, $Y = \{1, 2, \ldots, T\}$, $T \in \mathbb{N}$, the output is one of T categories.
- Multilabel, $Y = 2^{\{1,2,\ldots,T\}}$, $T \in \mathbb{N}$, each output is any subset of T categories.

An interesting case is that of so called multitask learning. Here $Z = (X_1, Y_1) \times (X_2, Y_2) \times \cdots \times (X_T, Y_T))$ and the training set is is $S = (S_1, S_2, \ldots, S_T)$. Here we can view each data space/training set as corresponding to different yet related tasks. In full generality, input/output spaces and data cardinality can be different.

22.2 Data and Sampling Models

The standard data model we consider is a training set as an i.i.d. sample from a distribution p on the data space Z.

- Semisupervised, the more general situation where unlabelled data S_u are available together with the labelled data S.
- Transductive, related to the above setting, unlabelled data S_u are available together with the labelled data and the goal is to predict the label of the unlabeled data set S_u.
- Online/Dynamic Learning, the data are not i.i.d. The samples can be dependent, the samples can come from varying distribution or both.

22.3 Learning Approaches

- Online/Incremental
- Randomized
- distributed
- Online/Dynamic Learning, the data are not i.i.d. The samples can be dependent, the samples can come from varying distribution or both.
- Active,
- Reinforcement Learning.

22.4 Some Current and Future Challenges in Machine Learning

Challenges

$$1 \leftarrow \text{Data Size} \rightarrow \infty + \infty$$

22.4.1 Big Data?

Recent times have seen the development of technologies for gathering data-set of unprecedented size and complexity both in natural science and technology. On the one hand this has opened novel opportunities (e.g. online teaching), on the other had it has posed new challenges. In particular, the necessity has emerged to develop learning techniques capable to leverage predefined *budgets* and requisites in terms of

- Computations,
- Communications,
- Privacy.

22.4.2 Or Small Data?

One of the most evident difference between biological and artificial intelligence is the astounding ability of humans to generalize from limited supervised data. Indeed, while impressive, current artificial intelligent systems based on supervised learning required huge amounts of humanly annotated data.

- Unsupervised learning of data representation
- Learning under weak supervision.
- Learning and exploiting structure among learning tasks.