On a Combinatorial Problem in Machine Teaching

The formal model of machine learning (PAC)

- We have a concept class C of possible hypotheses
- C is given by a binary matrix M where the rows are concepts and the columns is the domain of the examples. $M(c, x) = 1$ if c is consistent with the data point $(x, 1)$.
- Then in PAC learning the question if how many data point do we need to estimate the correct concept when the data points are sampled at random.
- The worst case is related to the VC dimension of M. Which is the maximum number k of columns such that when we restrict the matrix to these columns there are 2^k (The maximum possible) different rows.

Machine teaching

- Now we have a teacher T which given a concept $c^* \in C$ chooses at set of set of examples $T(c^*) = w$ of minimal size so that the teacher can reconstruct c^*.
- We then try to minimize the teaching dimension which is $\max_{c \in C} T(c)$.

References

- R. L. Graham. On primitive graphs and optimal vertex assignments. *Annals of the New York academy of sciences*, 175(1):170–186, 1970.

  S. Hart. A note on the edges of the n-cube. *Discrete Mathematics*, 14(2):157–163, 1976.

  L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984. doi: 10.1145/1968.1972. URL

Example 1

- $$\mathbf{M} = \begin{array}{c} \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$
- $T(c_1) = \{x_1, x_2\}$, $T(c_2) = \{x_1, x_3\}$, $T(c_3) = \{x_1, x_2\}$, $T(c_4) = \{x_1, x_2\}$, $T(c_5) = \{x_4\}$.
- Teaching dimension is 2.

The consistency graph

- Let L be the set of labeled examples. For M we have $L = \{(x_1, 0), (x_1, 1), (x_2, 0), ..., (x_4, 1)\}$. Then let W be the power set of L.
- The consistency graph is the bipartite graph with partitions C and W (The power set of W). $c \in C$ has an edge to $w \in W$ if c is consistent with all the examples in w.

Example 2

- The consistency graph of M the node c_1 will be connected to the nodes $\{(x_1, 1)\}, \{(x_2, 0)\}$ and $\{(x3, 1), (x_4, 1)\}$ for example.
- c_1 will not be connect to the node $\{(x_1, 1), (x_2, 0), (x_3, 0)\}$ because c_1 is not consistent with the data point $(x_3, 0)$.

A matrix sum

- For a subset of the columns Q of a matrix N let $M(Q)$ be matrix restricted to the columns in Q and let $dif(N)$ be the number of unique rows in a matrix.
- $\boldsymbol{m_q(N) = \sum_{Q \in \binom{[1,n]}{q}} dif(N(Q))}$.
- So m_q takes all the subsets of size q of columns of M and counts how many unique rows there are for each such subset.
- This counts the number of $W-$ vertices on q examples which has at least on neighbor among the concepts.

Example 3

For the matrix M we have that $m_2(M) = dif(M(\{1, 2\})) + dif(M(\{1, 3\})) + dif(M(\{1, 4\})) + dif(M(\{2, 3\})) + dif(M(\{2, 4\})) + dif(M(\{3, 4\})) = 4 + 3 + 3 + 4 + 3 + 3 = 20$

Our results

- **We find the matrix M with the minimal value of $m_q(M)$ for all q**
- This matrix minimizes the number of $W-$vertices having a neighbor in the consistency graph.
- This matrix is the matrix with binary numbers from 0 to $|C|$. For example:

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Joakim Sunde, Brigt Håavardstun, Jan Kratochvíl, Jan Arne Telle

University of Bergen, Charles University