LEARNING PARTIAL BOOLEAN FUNCTIONS BY LOCAL SEARCH

BJOERN ANDRES

Abstract. We seek to separate disjoint sets $A \cup B = S \subseteq \{0,1\}^J$ by Boolean functions $f, g\colon \{0,1\}^J \to \{0,1\}$ such that $f(A) = 1$ and $g(B) = 1$ and $f \cdot g = 0$. Specifically, we consider f and g each defined by a disjunctive normal form, and we seek to minimize the sum of the lengths of these forms. This problem is known to be inapproximable to within $(1-\epsilon)\log|S|$ for any $\epsilon > 0$ unless P=NP. The task of this project is to design, implement and compare local search algorithms for the problem.

1. Preliminaries

Definition 1. A tuple (J, S, X, Y, x, y) is called *Boolean labeled data* iff J is finite and non-empty, called the set of *features*, $X = \{0,1\}^J$, called the *feature space*, S is finite and non-empty, called the set of *samples*, $Y = \{0,1\}$, called the *label space*, $x\colon S \to X$ and $y\colon S \to Y$. Boolean labeled data is called *ambiguous* iff there exist $s, s' \in S$ such that $x_s = x_{s'}$ and $y_s \neq y_{s'}$.

Definition 2. For any finite $X = \{0,1\}^J$, $\Gamma = \left\{\left(V, \bar{V}\right) \in 2^J \times 2^J | V \cap \bar{V} = \emptyset\right\}$ and $\Theta = 2^\Gamma$, the family $f : \Theta \to \{0,1\}^X$ such that, for any $\theta \in \Theta$ and any $x \in X$,

$$f_\theta(x) = \sum_{(J_0,J_1)\in\theta} \prod_{j\in J_0} x(j) \prod_{j\in J_1} (1 - x(j)) \tag{1}$$

is called the family of J-variate *disjunctive normal forms (DNFs)*.

For $R_l, R_d\colon \Theta \to \mathbb{N}_0$ such that for all $\theta \in \Theta$,

$$R_l(\theta) = \sum_{(J_0,J_1)\in\theta} (|J_0| + |J_1|) \quad , \tag{2}$$

$$R_d(\theta) = \max_{(J_0,J_1)\in\theta} (|J_0| + |J_1|) \quad , \tag{3}$$

$R_l(\theta)$ and $R_d(\theta)$ are called the *length* and *depth*, resp., of the DNF defined by θ.

Definition 3. For any non-ambiguous Boolean labeled data $D = (J, S, X, Y, x, y)$, for any non-empty set Θ and family $f\colon \Theta \to \{0,1\}^X$ of Boolean functions, and

TU Dresden, University of Tübingen
E-mail address: bjoern.andres@tu-dresden.de.

for any function $R\colon \Theta \to \mathbb{N}_0$, called a *regularizer*, the instance of the OPTIMAL-SEPARATION problem w.r.t. D, Θ, f and R has the form

$$\min_{(\theta,\theta')\in\Theta^2} \quad R(\theta) + R(\theta') \tag{4}$$

$$\text{subject to} \quad \forall s \in y^{-1}(1)\colon \quad f_\theta(x_s) = 1 \qquad \text{(exactness)} \tag{5}$$

$$\forall s \in y^{-1}(0)\colon \quad f_{\theta'}(x_s) = 1 \qquad \text{(exactness)} \tag{6}$$

$$f_\theta \cdot f_{\theta'} = 0 \ . \qquad \text{(non-contradictoriness)} \tag{7}$$

For any $m \in \mathbb{N}_0$, the instance of the SEPARABILITY problem w.r.t. D, Θ, f, R and m is to decide whether there exist $\theta, \theta' \in \Theta$ such that

$$\forall s \in y^{-1}(1)\colon \qquad f_\theta(x_s) = 1 \qquad \text{(exactness)} \tag{8}$$

$$\forall s \in y^{-1}(0)\colon \qquad f_{\theta'}(x_s) = 1 \qquad \text{(exactness)} \tag{9}$$

$$f_\theta \cdot f_{\theta'} = 0 \qquad \text{(non-contradictoriness)} \tag{10}$$

$$R(\theta) + R(\theta') \leq m \ . \qquad \text{(boundedness)} \tag{11}$$

Two remarks are in order. Firstly, any feasible solution $(\theta, \theta') \in \Theta^2$ to either of these problems defines a partial Boolean function $h\colon X' \to \{0,1\}$ with $X' = \{x' \in \{0,1\}^J \mid f_\theta(x') = 1 \vee f_{\theta'}(x') = 1\}$ and such that for all $x' \in X'$, we have $h(x') = 1$ iff $f_\theta(x') = 1$ (and equivalently, $h(x') = 0$ iff $f_{\theta'}(x') = 1$). Secondly, exactness means the labeled data is classified totally and without errors, i.e., for all $s \in S$, we have $x_s \in X'$ and $h(x_s) = y_s$.

2. Motivation

SEPARABILITY is known to be NP-hard, and OPTIMAL-SEPARABILITY is known to be inapproximable to within $(1-\epsilon)\log|S|$ for any $\epsilon > 0$ unless P=NP. It is unknown how close one can get to a $\log|S|$-approximation in polynomial time.

3. Related work

The standard textbook on Boolean functions is [2], and its two authors are leading researchers in this area. References to their publications are excellent pointers to specific recent work.

4. Task

The task of this project is to design, implement and empirically compare local search algorithms for the problem:

(1) A feasible solution to begin (initialize) with is given by

$$f_\theta(x) = \sum_{s\in y^{-1}(1)} \prod_{\{j\in J | x_s(j)=1\}} x(j) \prod_{\{j\in J|x_s(j)=0\}} (1 - x(j)) \tag{12}$$

$$f_{\theta'}(x) = \sum_{s\in y^{-1}(0)} \prod_{\{j\in J | x_s(j)=1\}} x(j) \prod_{\{j\in J|x_s(j)=0\}} (1 - x(j)) \tag{13}$$

Others can be proposed as part of this project.

(2) Two transformations that shall be searched over in order to reduce the objective function are described below. Others are to be proposed as part of this project.
   (a) removing a literal from a term

   (b) adding a literal to one term and simultaneously removing literals of the same variable from other terms.

(3) Algorithms are to be compared in terms of the objective value with respect to the running time, for any binary classification problem, e.g., distinguishing the digits 1 and 7 in the MNIST data set of hand-written digits [1].

References

[1] http://yann.lecun.com/exdb/mnist/.

[2] Yves Crama and Peter L. Hammer. *Boolean Functions: Theory, Algorithms, and Applications*. Cambridge University Press, 2011.