convex

by Harshith B V

Submission date: 23-Nov-2020 08:54PM (UTC+0530)
Submission ID: 1444726880
File name: BTP-1_Report_1.pdf (204.25K)
Word count: 1147
Character count: 6129

A NON CONVEX OPTIMIZATION BASED FRAMEWORK FOR FAIRNESS

Report submitted in partial fulfillment
of the requirements for the degree

of

Bachelor of Technology in
Computer Science and Engineering

by

Vivek Vardhan Adepu
(17CS10002)

Under the guidance of

Niloy Ganguly

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

November 2020

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, gures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: November 30, 2016 | Vivek Vardhan Adepu
Place: Kharagpur | 17CS10002

CERTIFICATE

This is to certify that the Dissertation Report entitled, **"A non convex optimization framework for fairness"** submitted by **Vivek Vardhan Adepu** to Indian Institute of Technology, Kharagpur, India, is a record of bona fide Project work carried out by him under my supervision and guidance and is worthy of consideration for the award of the degree of Bachelor of Technology in Computer Science and Engineering of the Institute. The Dissertation Report has fulfilled all the requirements as per the regulations of the institute and in my opinion reached the standard for submission.

Niloy Ganguly
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Date: 23 November 2020

Introduction

Bias is pervasive in our society. We tend to attribute discrimination to our own internalized prejudices and our inability to make decisions in truly objective ways. Because of this, machine learning algorithms seem like a compelling solution. However, these do fall prey to discriminatory bias. A classic example is what happened with Nikon's facial recognition algorithm, which automatically detected the presence of blinking in photos. However, this algorithm mistakenly flagged Asian people as blinking at a substantially higher rate than other demographics[5]. COMPAS, an algorithm used in several U.S. states to determine how likely a given defendant is to commit another crime in the future. This risk assessment is used to help determine high-impact consequences like probation and parole, but an analysis from ProPublica demonstrated that the algorithm's decisions can replicate racial discrimination[6]

In this project, we will be designing a non convex optimization framework to implement algorithmic fairness in machine learning applications

Background

- **Discrepancy**

We extend the notion of combinatorial discrepancy[8] to fairness for machine learning algorithms. Given a universal set X consisting of n elements and a collection $S = \{S_1, S_2,...,S_m\}$ of subsets of X, in combinatorial discrepancy tries to find a coloring $\chi : X \rightarrow \{-1, 1\}$ such that discrepancy of the system is minimized where discrepancy for a particular set S_j is defined as

$$D(S_j) = \left| \sum_{a \in S_j} \chi(a) \right|$$

We apply this in fairness by forming $S^j_i \in S$ which consist of all the elements of X that have a fixed value of j of the sensitive attribute i.

- **Convex Optimization:**

A mathematical optimization problem, or just optimization problem, has the form[2]

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \; f_i(x) \le b, \; i = 1, ..., m \end{aligned}$$

where the vector x = $(x_1, \ldots, x_n)$ is the optimization variable of the problem, the function f_0: $\mathbf{R}^n \to \mathbf{R}$ is the objective function, the functions f_i: $\mathbf{R}^n \to \mathbf{R}$, i = 1, . . . , m, are the constraint functions, and the constants $b_1, \ldots, b_m$ are the limits, or bounds, for the constraints.

A convex optimization problem is the one in which the objective and constraint functions are convex, i.e. they satisfy the inequality[2]

$$f_i(\alpha x + \beta y) \le \alpha f_i(x) \; + \; \beta f_i(y)$$

for all x, y $\in \mathbf{R}$ and all $\alpha, \beta \in \mathbf{R}$ with $\alpha + \beta = 1$, $\alpha \ge 0$, $\beta \ge 0$

Here, we are dealing with non-convex optimization problems which are not solvable by standard algorithms like least-squares, linear programming etc.

- ### Support Vector Machines(SVM):

Support vector machines are a set of supervised learning methods used for classification, regression and outliers detection[3].

A support vector machine algorithm finds a hyperplane in an N-dimensional space, N being the number of features, that distinctly classifies the data points. As there are many possible hyperplanes, it picks the one that has the maximum margin, i.e the maximum distance between data points of both classes. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we can maximize the margin of the classifier[4]

The Problem:

Concisely, we optimize the combined loss function as [7]:

Let w be a variable vector corresponding to an SVM classifier and f(w) be the objective function of the SVM and g(w) ∈ C be the feasible set of the corresponding optimization problem. Assuming that the points to be classified are mapped to a higher dimensional space via a mapping ϕ

$$\mathbf{w} = \sum_{i=1}^{n} c_i y_i \phi(\mathbf{x}_i)$$

where the c_i are solved by the following dual optimization problem

$$\min \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle c_j y_j - \sum_{i=1}^{n} c_i$$

$$\sum_{i=1}^{n} c_i y_i = 0 \ \forall i$$

$$0 \leq c_i \leq \frac{1}{2n\lambda} \ \forall i$$

The value of b can be solved as

$$\begin{aligned} b &= \langle \mathbf{w}, \phi(\mathbf{x}_j) \rangle - y_j \\ &= \left[\sum_{i=1}^{n} c_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \right] - y_j \end{aligned}$$

In our approach we will use this coloring directly in the discrepancy objective function and find a value of w that minimizes both the terms.

Thus our combined objective function becomes minimization of

$$
(1-\alpha)\Big\{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i K(\mathbf{x}_i,\mathbf{x}_j) c_j y_j - \sum_{i=1}^{n} c_i\Big\} \quad + \\
\alpha\{\max_{S_i \in S} \Big| \sum_{\mathbf{x}_j \in S_i} \mathrm{sgn}\Big(\Big[\sum_{i=1}^{n} c_i y_i K(\mathbf{x}_i,\mathbf{x}_j)\Big] - b\Big)\Big|\}
$$

$$
\begin{aligned}
\sum_{i=1}^{n} c_i y_i &= 0 \ \forall i \\
0 \le c_i &\le \frac{1}{2n\lambda} \ \forall i
\end{aligned}
$$

As sgn(x) function is not a continuous function, we approximate it with a non-convex smooth function like tanh(x) or $\frac{x}{\sqrt{x^2+1}}$ and use the similar techniques in the previous section to remove the modulus sign from the optimization which eventually gives us a non-convex optimization problem over the variables (c_1, c_2,...,c_n). In general SVM deals with a quadratic objective function but because of the joint optimization approach we need to solve a non-convex optimization problem

Future work

We are currently trying to leverage Zafar et al[1] by including the discrepancy objective in the objective function and building an optimizer for it. If not feasible, we will explore more approaches and try to solve the problem.

References

[1] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P Gummadi. Fairness constraints: A flexible approach for fair classification. J. Mach. Learn. Res., 20(75):1–42, 2019

[2] Boyd, S., Vandenberghe, L., Boyd, S. P. (2004). Convex Optimization. Spain: Cambridge University Press.

[3] Support Vector Machines — scikit-learn 0.23.2 documentation

[4] Support Vector Machine — Introduction to Machine Learning Algorithms | by Rohith Gandhi

[5] Face-Detection Cameras: Glitches Spur Charges of Racism - TIME

[6] How We Analyzed the COMPAS Recidivism Algorithm

[7] Bernhard Scholkopf and Alexander J Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. Adaptive Computation and Machine Learning series, 2018.

[8] Bernard Chazelle. The discrepancy method: randomness and complexity. Cambridge University Press, 2001.

convex

ORIGINALITY REPORT

56%	55%	14%	42%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Indian Institute of Technology, Kharagpure Student Paper	15%
2	sitn.hms.harvard.edu Internet Source	12%
3	www.docstoc.com Internet Source	6%
4	Submitted to Kingston University Student Paper	5%
5	d-nb.info Internet Source	4%
6	docplayer.net Internet Source	3%
7	Submitted to Chester College of Higher Education Student Paper	2%
8	ethesis.nitrkl.ac.in Internet Source	2%

9	Submitted to Panjab University Student Paper	2%
10	towardsdatascience.com Internet Source	2%
11	nitmanipur.ac.in Internet Source	1%
12	drum.lib.umd.edu Internet Source	1%
13	Perturbation Analysis of Optimization Problems, 2000. Publication	1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		