Active Learning using Cutting-Plane Methods

William Chong, Jenna Lee

I. Introduction

Active learning with cutting plane methods has been proposed in [1] where the main task is to train a classifier on an unlabeled data set. In this scenario, the optimization is over the weights of the classifier in the unit norm ball, and the key sub-algorithm loops are the centering and querying loops. Active learning is prevalent in many scenarios where data acquisition is expensive or labeling process is unreliable, if not both.

II. Previous Works

(Pulled from [1]): Cutting-plane methods have been used as a training algorithm for general methods such as SVM [2]. With [1], the general active learning algorithm with cutting planes is proposed.

III. Methods

Active learning iteratively learns the label of new data points based on query to the information source. At each time step, algorithm chooses the unlabled samples to query about. By intuition, it is critical for the agent to query about meaningful samples for efficient learning. In this light, cutting plane method provides useful information on each sample (x, y) based on the amount of volume reduced with each learning.

We propose to first recreate the results of the paper on active learning with cutting planes from the general algorithm mentioned in the paper [1]. We will explore variations of the algorithm, namely the center and the query functions.

Center:

1) Chebyshev center
2) Center of gravity
3) Possible extension to center of pressure

Query:

1) Hit-and-run methods [3]
2) Dikin walk [4]
3) Possible extension to a simplex-based approach like Johnson's algorithm

Algorithm 4 Top: a generic cutting-plane active learning procedure; w^t is computed as the 'center' of $\mathcal{C}^t$ —center my refer to the center of gravity of the Chebyshev center. Bottom: a possible implementation of QUERY(): sampling strategies are given in, e.g., [2], [31], [32].

$\mathcal{C}^0 \leftarrow \mathcal{B}$
$t \leftarrow 0$
repeat
  $w^t \leftarrow$ center$(\mathcal{C}^t)$
  $x_{n_t}, y_{n_t} \leftarrow$ QUERY$(\mathcal{C}^t, D)$
  if $y_{n_t}\langle w^t, x_{n_t}\rangle < 0$ **then**
    $\mathcal{C}^{t+1} \leftarrow \mathcal{C}^t \cap \{z : y_{n_t}\langle z, x_{n_t}\rangle \geq 0\}$
    $t \leftarrow t + 1$
  end if
until $\mathcal{C}^t$ is small enough
return w^t

function QUERY$(\mathcal{C}, D)$
  Sample M points $s_1, \ldots s_M$ from $\mathcal{C}$
  $\mathbf{g} \leftarrow \sum_{k=1}^{M} s_k/M$
  $x \leftarrow \arg\min_{x_i \in D}\langle \mathbf{g}, x_i\rangle$
  $y \leftarrow$ get label from an expert
  return x, y
end function

Fig. 1. Active learning cutting plane algorithm

The baseline will explore the datasets used in [1], starting with the simple 2D case and possibly extending to the Reuters21578 —ModApte variation— and Newsgroups datasets.

As of now, the possible extensions are using the center of pressure, and using a simplex-based approach for the query. With the center of pressure approach, there are additional relevant weights that is correlates to certain partitions in the weight space, and the center of pressure is calculated with a weighted average. With the simplex-based approach, a simplex-based search algorithm maybe used to query informative points based on the search directions used in prior iterations.

REFERENCES

[1] L. Ralaivola and U. Louche, "From cutting planes algorithms to compression schemes and active learning," 2015.

[2] T. Joachims, T. Finley, and C.-N. J. Yu, "Cutting-plane training of structural svms," *Mach. Learn.*, vol. 77, p. 27–59, Oct. 2009.

[3] L. Lovász and S. Vempala, "Hit-and-run from a corner," Apr 2006.

[4] R. Kannan and H. Narayanan, "Random walks on polytopes and an affine interior point method for linear programming," 05 2009.