Learning Mahalanobis metric spaces is an important problem that has found numerous applications. Several algorithms have been designed for this problem, including Information Theoretic Metric Learning (ITML) [Davis et al. 2007] and Large Margin Nearest Neighbor (LMNN) classification [Weinberger and Saul 2009]. We study the problem of learning a Mahalanobis metric space in the presence of adversarial label noise. To that end, we consider a formulation of Mahalanobis metric learning as an optimization problem, where the objective is to minimize the number of violated similarity/dissimilarity constraints. We show that for any fixed ambient dimension, there exists a fully polynomial-time approximation scheme (FPTAS) with nearly-linear running time. This result is obtained using tools from the theory of linear programming in low dimensions. As a consequence, we obtain a fully-parallelizable algorithm that recovers a nearly-optimal metric space, even when a small fraction of the labels is corrupted adversarially. We also discuss improvements of the algorithm in practice, and present experimental results on real-world, synthetic, and poisoned data sets.
In the case of Mahalanobis metric learning, we have \( X \subset \mathbb{R}^d \), with \( |X| = n \), for some \( d \in \mathbb{N} \), and the mapping \( f : \mathbb{R}^d \to \mathbb{R}^d \) is linear. Specifically, we seek to find a matrix \( G \in \mathbb{R}^{d \times d} \), such that for all \( \{p, q\} \in \mathcal{S} \), we have
\[
\|G_p - G_q\|_2 \leq u,
\]
and for all \( \{p, q\} \in \mathcal{D} \), we have
\[
\|G_p - G_q\|_2 \geq \ell.
\]

1.1 Our Contribution

In general, there might not exist any \( G \) that satisfies all constraints of type 1 and 2. We are thus interested in finding a solution that minimizes the fraction of violated constraints, which corresponds to maximizing the accuracy of the mapping. We develop a \((1 + \varepsilon)\)-approximation algorithm for optimization problem of computing a Mahalanobis metric space of maximum accuracy, that runs in near-linear time for any fixed ambient dimension \( d \in \mathbb{N} \). This algorithm is obtained using tools from geometric approximation algorithms and the theory of linear programming in small dimension. The following summarizes our result.

**Theorem 1.1.** For any \( d \in \mathbb{N}, \varepsilon > 0 \), there exists a randomized algorithm for learning \( d \)-dimensional Mahalanobis metric spaces, which given an instance that admits a mapping with accuracy \( r^* \), computes a mapping with accuracy at least \( r^* - \varepsilon \), in time \( d^{O(1)}n\log n/\varepsilon^{O(d)} \), with high probability.

The above algorithm can be extended to handle various forms of regularization. We also propose several modifications of our algorithm that lead to significant performance improvements in practice. The final algorithm is evaluated experimentally on both synthetic and real-world data sets, and in a data poisoning scenario, and is compared against the currently best-known algorithms for the problem.

1.2 Related Work

Several algorithms for learning Mahalanobis metric spaces have been proposed. Notable examples include the SDP based algorithm of Xing et al. [Xing et al.(2003)Xing, Jordan, Russell, and Ng], the algorithm of Globerson and Roweis for the fully supervised setting [Globerson and Roweis(2006)], Information Theoretic Metric Learning (ITML) by Davis et al. [Davis et al.(2007)Davis, Kulis, Jain, Sra, and Dhillon], which casts the problem as a particular optimization minimizing LogDet divergence, as well as Large Margin Nearest Neighbor (LMNN) by Weinberger et al. [Weinberger et al.(2006)Weinberger, Blitzer, and Saul], which attempts to learn a metric geared towards optimizing \( k \)-NN classification. We refer the reader to the surveys [Kulis et al.(2013), Li and Tian(2018)] for a detailed discussion of previous work. Our algorithm differs from previous approaches in that it seeks to directly minimize the number of violated pairwise distance constraints, which is a highly non-convex objective, without resorting to a convex relaxation of the corresponding optimization problem.

1.3 Organization

The rest of the paper is organized as follows. Section 2 describes the main algorithm and the proof of Theorem 1.1. Section 3 discusses practical improvements used in the implementation of the algorithm. Section 4 presents the experimental evaluation.

2 Mahalanobis Metric Learning as an LP-Type Problem

In this Section we present an approximation scheme for Mahalanobis metric learning in \( d \)-dimensional Euclidean space, with nearly-linear running time. We begin by recalling some prior results on the class of LP-type problems, which generalizes linear programming. We then show that linear metric learning can be cast as an LP-type problem.

2.1 LP-type Problems

Let us recall the definition of an LP-type problem. Let \( \mathcal{H} \) be a set of constraints, and let \( w : 2^\mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\} \), such that for any \( G \subset \mathcal{H} \), \( w(G) \) is the value of the optimal solution of the instance defined by \( G \). We say that \((\mathcal{H}, w)\) defines an LP-type problem if the following axioms hold:

(A1) **Monotonicity.** For any \( F \subseteq G \subseteq \mathcal{H} \), we have \( w(F) \leq w(G) \).

(A2) **Locality.** For any \( F \subseteq G \subseteq \mathcal{H} \), with \(-\infty < w(F) = w(G)\), and any \( h \in \mathcal{H} \), if \( w(G) < w(G \cup \{h\}) \), then \( w(F) < w(F \cup \{h\}) \).
More generally, we say that \((\mathcal{H}, w)\) defines an LP-type problem on some \(\mathcal{H}' \subseteq \mathcal{H}\), when conditions (A1) and (A2) hold for all \(F \subseteq G \subseteq \mathcal{H}'\).

A subset \(B \subseteq \mathcal{H}\) is called a basis if \(w(B) > -\infty\) and \(w(B') < w(B)\) for any proper subset \(B' \subseteq B\). A basic operation is defined to be one of the following:

- **(B0) Initial basis computation.** Given some \(G \subseteq \mathcal{H}\), compute any basis for \(G\).
- **(B1) Violation test.** For some \(h \in \mathcal{H}\) and some basis \(B \subseteq \mathcal{H}\), test whether \(w(B \cup \{h\}) > w(B)\) (in other words, whether \(B\) violates \(h\)).
- **(B2) Basis computation.** For some \(h \in \mathcal{H}\) and some basis \(B \subseteq \mathcal{H}\), compute a basis of \(B \cup \{h\}\).

### 2.2 An LP-type Formulation

We now show that learning Mahalanobis metric spaces can be expressed as an LP-type problem. We first note that we can rewrite (1) and (2) as

\[(p-q)^T A (p-q) \leq u^2, \quad (3)\]

and

\[(p-q)^T A (p-q) \geq \ell^2, \quad (4)\]

where \(A = G^T G\) is positive semidefinite.

We define \(\mathcal{H} = \{0,1\} \times \left(\mathbb{R}^d\right)^2\), where for each \((0, \{p,q\}) \in \mathcal{H}\), we have a constraint of type (3), and for every \((1, \{p,q\}) \in \mathcal{H}\), we have a constraint of type (4). Therefore, for any set of constraints \(F \subseteq \mathcal{H}\), we may associate the set of feasible solutions for \(F\) with the set \(\mathcal{A}_F\) of all positive semidefinite matrices \(A \in \mathbb{R}^{n \times n}\), satisfying (3) and (4) for all constraints in \(F\).

Let \(w : 2^\mathcal{H} \to \mathbb{R}\), such that for all \(F \in \mathcal{H}\), we have

\[w(F) = \begin{cases} \inf_{A \in \mathcal{A}_F} r^T A r & \text{if } \mathcal{A}_F \neq \emptyset \\ \infty & \text{if } \mathcal{A}_F = \emptyset \end{cases}, \]

where \(r \in \mathbb{R}^d\) is a vector chosen uniformly at random from the unit sphere from some rotationally-invariant probability measure. Such a vector can be chosen, for example, by first choosing some \(r' \in \mathbb{R}^d\), where each coordinate is sampled from the normal distribution \(\mathcal{N}(0, 1)\), and setting \(r = r'/\|r'\|_2\).

**Lemma 2.1.** When \(w\) is chosen as above, the pair \((\mathcal{H}, w)\) defines an LP-type problem of combinatorial dimension \(O(d^2)\), with probability 1. Moreover, for any \(n > 0\), if each \(r_i\) is chosen using \(\Omega(\log n)\) bits of precision, then for each \(F \subseteq \mathcal{H}\), with \(n = |F|\), the assertion holds with high probability.

**Proof.** Since adding constraints to a feasible instance can only make it infeasible, it follows that \(w\) satisfies the monotonicity axiom (A1).

We next argue that the locality axiom (A2) also holds, with high probability. Let \(F \subseteq G \subseteq \mathcal{H}\), with \(-\infty < w(F) = w(G)\), and let \(h \in \mathcal{H}\), with \(w(G) < w(G \cup \{h\})\). Let \(A_F \in \mathcal{A}_F\) and \(A_G \in \mathcal{A}_G\) be some (not necessarily unique)
With the interpretation of Mahalanobis metric learning as a LP-type problem given above, we thus obtain a linear time algorithm for Mahalanobis metric learning. This establishes that axiom (A2) holds with high probability. Each \( E \) is uniquely determined by specifying at most \( d \) points on its boundary (see [Welzl1991]). Therefore, each optimal solution can be uniquely specified as the intersection of at most \( O(d^2) \) constraints, and thus the combinatorial dimension is \( O(d^2) \).

Lemma 2.2. Any initial basis computation (B0), any violation test (B1), and any basis computation (B2) can be performed in time \( d^{O(1)} \).

Proof. The violation test (B1) can be performed by solving one SDP to compute \( w(B) \), and another to compute \( w(B \cup \{h\}) \). By Lemma 2.1 the combinatorial dimension is \( O(d^2) \), thus each SDP has \( O(d^2) \) constraints, and be solved in time \( d^{O(1)} \).

The basis computation step (B2) can be performed starting with the set of constraints \( B \cup \{h\} \), and iteratively remove every constraint whose removal does not decrease the optimum cost, until we arrive at a minimal set, which is a basis. In total, we need to solve at most \( d \) SDPs, each of size \( O(d^2) \), which can be done in total time \( d^{O(1)} \).

Finally, by the choice of \( w \), any set containing a single constraint in \( S \) is a valid initial basis.

2.3 Algorithmic implications

Using the above formulation of Mahalanobis metric learning as an LP-type problem, we can obtain our approximation scheme. Our algorithm uses as a subroutine an exact algorithm for the problem (that is, for the special case where we seek to find a mapping that satisfies all constraints). We first present the exact algorithm and then show how it can be used to derive the approximation scheme.

An exact algorithm. [Welzl1991] obtained a simple randomized linear-time algorithm for the minimum enclosing ball and minimum enclosing ellipsoid problems. This algorithm naturally extends to general LP-type problems (we refer the reader to [Har-Peled2011], Chazelle[2000]) for further details).

With the interpretation of Mahalanobis metric learning as an LP-type problem given above, we thus obtain a linear time algorithm for in \( \mathbb{R}^d \), for any constant \( d \in \mathbb{N} \). The resulting algorithm on a set of constraints \( F \subseteq H \) is implemented

```plaintext
Algorithm 2 An approximation algorithm for Mahalanobis metric learning.

procedure LPTML(F)
    for \( i = 0 \) to \( \log_{1+\epsilon} n \) do
        \( p \leftarrow (1+\epsilon)^{-i} \)
        for \( j = 1 \) to \( \log O(d^2) n \) do
            subsample \( F_j \subseteq F \), where each element is chosen independently with probability \( p \)
            \( A_{i,j} \leftarrow \text{Exact-LPTML}(F_j) \)
        end for
    end for
    return a solution out of \( \{A_{i,j}\}_{i,j} \), violating the minimum number of constraints in \( F \)
end procedure
```
by the procedure Exact-LPTML$(F; \emptyset)$, which is presented in Algorithm 1. The procedure LPTML$(F; B)$ takes as
input sets of constraints $F, B \subseteq \mathcal{H}$. It outputs a solution $A \in \mathbb{R}^{d \times d}$ to the problem induced by the set of constraints
$F \cup B$, such that all constraints in $B$ are tight (that is, they hold with equality); if no such solution solution exists, then
it returns nil. The procedure Basic-LPTML$(B)$ computes LPTML$(\emptyset; B)$. The analysis of [Welzl(1991)] implies that when Basic-LPTML$(B)$ is called, the cardinality of $B$ is at most the combinatorial dimension, which by Lemma 2.1
is $O(d^2)$. Thus the procedure Basic-LPTML can be implemented using one initial basis computation $(B0)$ and $O(d^2)$
basis computations $(B2)$, which by Lemma 2.2 takes total time $d^{O(1)}$.

An $(1 + \varepsilon)$-approximation algorithm. It is known that the above exact linear-time algorithm leads to an
nearly-linear-time approximation scheme for LP-type problems. This is summarized in the following. We refer the reader to
[Har-Peled(2011)] for a more detailed treatment.

**Lemma 2.3** ([Har-Peled(2011)], Ch. 15). Let $A$ be some LP-type problem of combinatorial dimension $\kappa > 0$, defined
by some pair $(\mathcal{H}, w)$, and let $\varepsilon > 0$. There exists a randomized algorithm which given some instance $F \subseteq \mathcal{H}$, with
$|F| = n$, outputs some basis $B \subseteq F$, that violates at most $(1 + \varepsilon)k$ constraints in $F$, such that $w(B) \leq w(B')$, for any
basis $B'$ violating at most $k$ constraints in $F$, in time $O \left( t_0 + \left( n + n \min \left\{ \frac{\log^{k+1} n}{\varepsilon^2 n}, \frac{\log^{k+2} n}{\kappa \varepsilon^2 n^2} \right\} \right) (t_1 + t_2) \right)$, where $t_0$
is the time needed to compute an arbitrary initial basis of $A$, and $t_1$, $t_2$, and $t_3$ are upper bounds on the time needed
to perform the basic operations $(B0)$, $(B1)$ and $(B2)$ respectively. The algorithm succeeds with high probability.

For the special case of Mahalanobis metric learning, the corresponding algorithm is given in Algorithm 2. The approxi-
mation guarantee for this algorithm is summarized in 1.1. We can now give the proof of our main result.

**Proof of Theorem 1.1.** Follows immediately by Lemmas 2.2 and 2.3. \qed

**Regularization.** We now argue that the LP-type algorithm described above can be extended to handle certain types
of regularization on the matrix $A$. In methods based on convex optimization, introducing regularizers that are convex
functions can often be done easily. In our case, we cannot directly introduce a regularizing term in the objective function
that is implicit in Algorithm 2. More specifically, let $\text{cost}(A)$ denote the total number of constraints of type (3)
and (4) that $A$ violates. Algorithm 2 approximately minimizes the objective function $\text{cost}(A)$. A natural regularized
version of Mahalanobis metric learning is to instead minimize the objective function $\text{cost}'(A) := \text{cost}(A) + \eta \cdot \text{reg}(A)$,
for some $\eta > 0$, and regularizer $\text{reg}(A)$. One typical choice is $\text{reg}(A) = \text{tr}(AC)$, for some matrix $C \in \mathbb{R}^{d \times d}$; the case $C = I$
corresponds to the trace norm (see [Kulis et al.(2013)]). We can extend the Algorithm 2 to handle any regularizer that can be expressed as a linear function on the entries of $A$, such as $\text{tr}(A)$. The following summarizes the
result.

**Theorem 2.4.** Let $\text{reg}(A)$ be a linear function on the entries of $A$, with polynomially bounded coefficients. For any
d \in \mathbb{N}, \varepsilon > 0, there exists a randomized algorithm for learning $d$-dimensional Mahalanobis metric spaces, which
given an instance that admits a solution $A_0$ with $\text{cost}'(A_0) = c^*$, computes a solution $A$ with $\text{cost}'(A) \leq (1 + \varepsilon)c^*$,
in time $d^{O(1)}n(\log n/\varepsilon)^{O(d)}$, with high probability.

**Proof.** If $\eta < t^i$, for sufficiently large constant $t > 0$, since the coefficients in $\text{reg}(A)$ are polynomially bounded, it follows
that the largest possible value of $\eta \cdot \text{reg}(A)$ is $O(\varepsilon)$, and can thus be omitted without affecting the result. Similarly,
if $\eta > (1/\varepsilon)n^t$, for sufficiently large constant $t' > 0$, since there are at most $(\kappa^2)$ constraints, it follows that the term
$\text{cost}(A)$ can be omitted form the objective. Therefore, we may assume w.l.o.g. that $\text{reg}(A_0) \in [e^{O(1)} / (1/\varepsilon)^{O(1)}]$. We
can guess some $i = O(\log n + \log(1/\varepsilon))$, such that $\text{reg}(A_0) \in [(1 + \varepsilon)^{-1}, (1 + \varepsilon)^i]$. We modify the SDP used
in the proof of Lemma 2.2 by introducing the constraint $\text{reg}(A) \leq (1 + \varepsilon)^i$. Guessing the correct value of $i$ requires
$O(\log n + \log(1/\varepsilon))$ executions of Algorithm 2, which implies the running time bound. \qed

3 Practical Improvements and Parallelization

We now discuss some modifications of the algorithm described in the previous section that significantly improve its
performance in practical scenarios, and have been implemented in our implementation.

**Move-to-front and pivoting heuristics.** We use heuristics that have been previously used in algorithms for linear
programming [Seidel(1990), Clarkson(1995)], minimum enclosing ball in $\mathbb{R}^3$ [Megiddo(1983)], minimum enclosing
ball and ellipsoid is $\mathbb{R}^d$, for any fixed $d \in \mathbb{N}$ [Welzl(1991)], as well as in fast implementations of minimum enclosing
ball algorithms [Gärtner(1999)]. The move-to-front heuristic keeps an ordered list of constraints which gets reorganized as the algorithm runs; when the algorithm finds a violation, it moves the violating constraint to the beginning of the list of the current sub-problem. The pivoting heuristic further improves performance by choosing to add to the basis the constraint that is “violated the most”. For instance, for similarity constraints, we pick the one that is mapped to the largest distance greater than $u$; for dissimilarity constraints, we pick the one that is mapped to the smallest distance less than $\ell$.

Approximate counting. The main loop of Algorithm 2 involves counting the number of violated constraints in each iteration. In problems involving a large number of constraints, we use approximate counting by only counting the number of violations within a sample of $O(\log 1/\varepsilon)$ constraints.

Early termination. A bottleneck of Algorithm 2 stems from the fact that the inner loop needs to be executed for $O(d^2 \log n)$ iterations. In practice, we have observed that a significantly smaller number of iterations is needed to achieve high accuracy. We denote by LPTML$_t$ for the version of the algorithm that performs a total of $t$ iterations of the inner loop.

Parallelization. Algorithm 2 consists of several executions of the algorithm Exact-LPTML on independently sampled sub-problems. Therefore, Algorithm 2 can trivially be parallelized by distributing a different set of sub-problems to each machine, and returning the best solution found overall.
### Table 1: Average accuracy and standard deviation over 50 executions of ITML, LMNN and LPTML.

| Data set  | ITML        | LMNN        | LPTML$_{t=2000}$ |
|-----------|-------------|-------------|------------------|
| Iris      | 0.96 ± 0.01 | 0.96 ± 0.02 | 0.94 ± 0.04      |
| Soybean   | 0.95 ± 0.04 | 0.96 ± 0.04 | 0.90 ± 0.05      |
| Synthetic | 0.97 ± 0.02 | 1.00 ± 0.00 | 1.00 ± 0.00      |

Figure 3: Accuracy in a data poisoning scenario. See Figure 2a.

## 4 Experimental Evaluation

We have implemented Algorithm 2, incorporating the practical improvements described in Section 3, and performed experiments on synthetic and real-world data sets. Our LPTML implementation and documentation can be found in our repository. We now describe the experimental setting and discuss the main findings.

### 4.1 Experimental Setting

**Classification task.** Each data set used in the experiments consists of a set of labeled points in $\mathbb{R}^d$. The label of each point indicates its class, and there is a constant number of classes. The set of similarity constraints $S$ (resp. dissimilarity constraints $D$) is formed by uniformly sampling pairs of points in the same class (resp. from different classes). We use various algorithms to learn a Mahalanobis metric for a labeled input point set in $\mathbb{R}^d$, given these constraints. The values $u$ and $\ell$ are chosen as the 90th and 10th percentiles of all pairwise distances. We used 2-fold cross-validation: At the training phase we learn a Mahalanobis metric, and in the testing phase we use $k$-NN classification, with $k = 4$, to evaluate the performance.

**Data sets.** We have tested our algorithm on the following synthetic and real-world data sets:

1. **Real-world:** We have tested the performance of our implementation on the Iris, Wine, Ionosphere and Soybean data sets from the UCI Machine Learning Repository.
2. **Synthetic:** Next, we consider a synthetic data set that is constructed by first sampling a set of 100 points from a mixture of two Gaussians in $\mathbb{R}^2$, with identity covariance matrices, and with means $(-3, 0)$ and $(3, 0)$ respectively; we then apply a linear transformation that stretches the $y$ axis by a factor of 40. This linear transformation reduces the accuracy of $k$-NN on the underlying Euclidean metric with $k = 4$ from 1 to 0.68.
3. **Data poisoning:** We modify the above synthetic data set by introducing a small fraction of points in an adversarial manner, before applying the linear transformation. Figure 2b depicts the noise added as five points labeled as one of the classes, and sampled from a Gaussian with identity covariance matrix and mean $(-100, 0)$ (Figure 2a).

**Algorithms.** We compare the performance of our algorithm against ITML and LMNN. We used the implementations provided by the authors of these works, with minor modifications.

---

2 [https://archive.ics.uci.edu/ml/datasets.php](https://archive.ics.uci.edu/ml/datasets.php)
4.2 Results

Accuracy. Algorithm 2 minimizes the number of violated pairwise distance constraints. It is interesting to examine the effect of this objective function on the accuracy of $k$-NN classification. Figure 1 depicts this relationship for the Wine data set. We observe that, in general, as the number of iterations of the main loop of LPTML increases, the number of violated pairwise distance constraints decreases, and the accuracy of $k$-NN increases. This phenomenon remains consistent when we first perform PCA to $d = 4, 8, 12$ dimensions.

Comparison to ITML and LMNN. We compared the accuracy obtained by LPTML, for $t = 2000$ iterations, against ITML and LMNN.

Table 1 summarizes the findings on the real-world and data sets and the synthetic data set without adversarial noise. We observe that LPTML achieves accuracy that is comparable to ITML and LMNN.

We observe that LPTML outperforms ITML and LMNN on the poisoned data set. This is due to the fact that the introduction of adversarial noise causes the relaxations used in ITML and LMNN to be biased towards contracting the $x$-axis. In contrast, the noise does not “fool” LPTML because it only changes the optimal accuracy by a small amount. The results are summarized in Figure 3.

The effect of dimension. The running time of LPTML grows with the dimension $d$. This is caused mostly by the fact that the combinatorial dimension of the underlying LP-type problem is $O(d^2)$, and thus performing each basic operation requires solving an SDP with $O(d^2)$ constraints. Figure 4 depicts the effect of dimensionality in the running time, for $t = 100, \ldots, 2000$ iterations of the main loop. The data set used is Wine after performing PCA to $d$ dimensions, for $d = 2, \ldots, 13$.

Parallel implementation. We implemented a massively parallel version of LPTML in the MapReduce model. The program maps different sub-problems of the main loop of LPTML to different machines. In the reduce
step, we keep the result with the minimum number of constraint violations. The implementation uses the mrjob [Yelp and Contributors(2019)] package. For these experiments, we used Amazon cloud computing instances of type m4.xlarge, AMI 5.20.0 and configured with Hadoop. As expected, the training time decreases as the number of available processors increases (Figure 5). All technical details about this implementation can be found in the parallel section of the documentation of our code.

5 Conclusions

We have shown that the problem of learning a Mahalanobis metric space can be cast as an LP-type problem. This formulation allows us to obtain an efficient approximation scheme using tools from the theory of linear programming in low dimensions. Specifically, we present a near-linear time \( (1 + \varepsilon) \)-approximation algorithm that minimizes the number of violated constraints. Experimental evaluation demonstrates that when compared to prior work, our method is significantly more robust against small adversarial modifications of the input labelling. Our approach also leads to a fully parallelizable algorithm.

It is an interesting research direction to extend our approximation algorithm to other classes of metric learning problems. One such case is when the input is specified as a set of ordered triples \((x, y, z)\), and the goal is to find a mapping \( f \) with \( \|f(x) - f(y)\|_2 \leq \|f(x) - f(z)\|_2 - m \), for some margin \( m > 0 \) (see [7]). Another important direction is to obtain geometric approximation algorithms for non-linear metric learning primitives, such as mappings computed by small depth neural networks.

References

[Chazelle(2000)] B. Chazelle. The discrepancy method: randomness and complexity. Cambridge University Press, 2000.
[Clarkson(1995)] K. L. Clarkson. Las vegas algorithms for linear and integer programming when the dimension is small. Journal of the ACM (JACM), 42(2):488–499, 1995.
[Davis et al.(2007)Davis, Kulis, Jain, Sra, and Dhillon] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In Proceedings of the 24th international conference on Machine learning, pages 209–216. ACM, 2007.
[Gärtner(1999)] B. Gärtner. Fast and robust smallest enclosing balls. In European Symposium on Algorithms, pages 325–338. Springer, 1999.
[Globerson and Roweis(2006)] A. Globerson and S. T. Roweis. Metric learning by collapsing classes. In Advances in neural information processing systems, pages 451–458, 2006.
[Har-Peled(2011)] S. Har-Peled. Geometric approximation algorithms. Number 173. American Mathematical Soc., 2011.
[Kulis et al.(2013)] B. Kulis et al. Metric learning: A survey. Foundations and Trends® in Machine Learning, 5(4):287–364, 2013.
[Li and Tian(2018)] D. Li and Y. Tian. Survey and experimental study on metric learning methods. Neural Networks, 105:447–462, 2018.
[Megiddo(1983)] N. Megiddo. Linear-time algorithms for linear programming in \( r^3 \) and related problems. SIAM journal on computing, 12(4):759–776, 1983.
[Seidel(1990)] R. Seidel. Linear programming and convex hulls made easy. In Proceedings of the sixth annual symposium on Computational geometry, pages 211–215. ACM, 1990.
[Weinberger et al.(2006)Weinberger, Blitzer, and Saul] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In Advances in neural information processing systems, pages 1473–1480, 2006.
[Welzl(1991)] E. Welzl. Smallest enclosing disks (balls and ellipsoids). In New results and new trends in computer science, pages 359–370. Springer, 1991.
[Xing et al.(2003)Xing, Jordan, Russell, and Ng] E. P. Xing, M. I. Jordan, S. J. Russell, and A. Y. Ng. Distance metric learning with application to clustering with side-information. In Advances in neural information processing systems, pages 521–528, 2003.
[Yelp and Contributors(2019)] Yelp and Contributors. Mrjob library, 2019. URL http://pythonhosted.org/mrjob/v0.6.8.dev0.