COMPSCI 530: Design and Analysis of Algorithms	Oct 29, 2013
Lecture 17	
Lecturer: Pankaj K. Agarwal	*Scribe: Fangjian Guo*

1 Overview

In this lecture, we will study approximation algorithms for problems related to range spaces, especially the hitting set problem.

2 Range Spaces and the Hitting Set Problem

2.1 Range spaces

A range space is also called a set system or a hypergraph. Its definition is given as follows.

Definition 1. *A range space Σ is a pair $\Sigma = (X, R)$, where X is a finite set of objects and the ranges R is a set of subsets of X.*

For example, suppose we have $X = \{1,2,3\}$ and $R = \{\{1\},\{2,3\},\{1,2,3\}\}$, then $\Sigma = (X,R)$ is a range space. It shall be noted that R does not need to include all the subsets of X. Yet, when R is indeed the collection of all the subsets of X, the size of R is exponential, namely

$$|R| = 2^n,$$

where we denote $n = |X|$.

In this lecture, we will instead study the ranges spaces that are *well-behaved*, which means the size of R is polynomial, namely

$$|R| \leq n^{O(1)}.$$

The linear classification for points in the $\mathbb{R}^2$ is such an example.

Example 1. Let us suppose we have X as a finite set of points in $\mathbb{R}^2$, which can be imagined as two-dimensional data points drawn on a plane. For every line we draw in the plane, it will cut the space into two half-spaces. We can define the ranges R to be the set of points in a half-space for all possible lines, i.e.

$$R = \{\gamma \cap X \mid \gamma \text{ is a half-space}\}.$$

Then we will show $|R|$ is polynomial instead of exponential.

To count the size of $|R|$, as shown in Figure 1, one can consider moving and rotating the line l_1 while the classification does not change, until it is in the position of l_2, which passes through two points. That is to say, such a line will be the boundary that separates one classification from another, corresponding to a subset from another subset in R . By counting all these boundary lines, we know $|R| = 2 \times \binom{n}{2} = O(n^2)$. The same trick can be applied to other definitions of R. For example, if we let R to be $R = \{\gamma \cap X \mid \gamma \text{ is a disk (circle)}\}$, then we know $|R| = O(n^3)$ because three points can uniquely determine a circle.

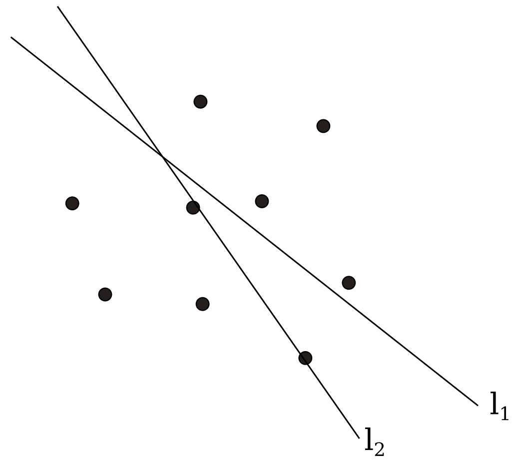

Figure 1: Moving and rotating l_1 to l_2 until it passes through two points.

2.2 The hitting set problem

Definition 2. *For a range space $\Sigma = (X, R)$, a hitting set is a subset $H \subseteq X$ such that $\forall \gamma \in R$, it holds that $H \cap \gamma \neq \varnothing$.*

That is, a hitting set H "hits" every range by having at least one of its element appearing in every range. A hitting set problem can be formulated as a 0-1 linear programming. If we define z_i for every element $x_i \in X$, where $z_i = 1$ denotes x_i is included by H and $z_i = 0$ otherwise. The LP can be written as below.

$$\begin{aligned} &\min \sum_{i=1}^{n} z_i \\ s.t. &\sum_{x_i \in \gamma} z_i \geq 1 \quad \forall \gamma \in R, \\ &z_i \in \{0, 1\}. \end{aligned}$$

Next we introduce ε-net as a relaxed version of the hitting set. Instead of requiring the set to "hit" every range, an ε-net is only required to hit the ranges that are "heavy" enough. To measure the heaviness, we can use a weight function $w : X \to \mathbb{R}^+$ that puts a weight for every object. And the weight of a set is simply defined as the total weights of the objects included, i.e.

$$w(\gamma) = \sum_{x \in \gamma} w(x).$$

Definition 3. *For a range space $\Sigma = (X, R)$, and for $0 < \varepsilon < 1$, a subset $A \subseteq X$ is an ε-net if $A \cap \gamma \neq \varnothing$ for all $\gamma \in R$ such that $w(\gamma) \geq \varepsilon w(X)$.*

When we choose an uniform weight $w(x) = 1$ for all $x \in X$, the condition is reduced to the condition on size $|\gamma| \geq \varepsilon |X|$.

2.3 Approximation algorithm for the hitting set

Generally speaking, it is difficult to approximate a hitting set problem. The best one can do for a general hitting set problem in polynomial time is an approximation with factor $O(\log n)$. Yet, an ε-net is easier to find an approximate or randomized solution if it is defined on a well-behaved range space ($|R|$ is polynomial). We state two facts without proof.

Fact 1. *For a range space $\Sigma = (X,R)$ with $|R| \leq |X|^{O(1)}$, an ε-net of size $O(1/\varepsilon)$ can be computed in $(|X|/\varepsilon)^{O(1)}$ time.*

Fact 2. *For a range space $\Sigma = (X,R)$ with $|R| \leq |X|^{O(1)}$, a random subset of size $O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ formed by including each $x \in X$ with probability $w(x)/w(X)$ is an ε-net with probability $p \geq 2/3$.*

Then, for a well-behaved range space, we can convert the hitting set problem to an ε-net problem and use either of the aforementioned algorithm to find an approximate or a probabilistically correct solution. As we know an ε-net only cares the ranges that are heavy, we can do the conversion by choosing ε and weights w carefully such that every range becomes heavy, that is $w(\gamma) \geq \varepsilon w(X)$ for all $\gamma \in R$.

Let us suppose we have an oracle that can tell us the size k of the optimal hitting set. Then we set $\varepsilon = \log\sqrt{2}/k$, where the choice of the constant will be clear later. The algorithm for choosing the weights w goes as follows.

Initially, we set $w(x) = 1$ for all $x \in X$. While $\exists\gamma \in R$ such that $w(\gamma) < \varepsilon w(X)$, we arbitrarily pick such a γ and double the weights for the objects included by γ, i.e.

$$w(x) \leftarrow 2w(x) \quad \forall x \in \gamma.$$

Repeat it until $w(\gamma) \geq \varepsilon w(X)$ for all $\gamma \in R$.

Theorem 3. *The algorithm will terminate in $O(k\log n)$ time.*

Proof. Let w_i be $w(X)$ after i iterations. Clearly, $w_0 = |X| = n$. In every iteration, we double the weight for the range that is lighter than εw_i. Therefore, we have

$$w_{i+1} \leq w_i + \varepsilon w_i = (1+\varepsilon)w_i.$$

By unrolling it, we have

$$w_i \leq (1+\varepsilon)^i n \leq \exp(\varepsilon)^i n = \exp(\varepsilon i + \log n).$$

Supposing the optimal hitting set is $H = \{x_1, x_2, \cdots, x_k\}$. By the definition of a hitting set, at least one of its element will appear in the range that we pick for weight-doubling in each iteration. Then, after i iterations, every $x_i \in H$ should be weight-doubled for at least i/k times, which gives us the lower bound for w_i as

$$w_i \geq w(H) \text{ after } i \text{ iterations } \geq k2^{i/k}.$$

By combining both bounds, we have

$$k2^{i/k} \leq \exp(\varepsilon i + \log n).$$

Now we can arrive at

$$i \leq \frac{k\log n}{\log(2)/2} = O(k\log n).$$

□

Clearly, when the algorithm terminates, all ranges will be heavy and thus an ε-net will also be a hitting set. However, in practice, we do not have an oracle and therefore do not know k. We can instead start with some guess of k and then run the reweighting algorithm for at most $\frac{k\log n}{\log(2)/2}$ iterations. If the algorithm terminates within the time limit, then we can continue to find an ε-net ; if not, we double our guess for k and repeat the process until the algorithm terminates within the limit.

3 Summary

In this lecture, we have introduced range spaces, the hitting set and the ε-net. We have seen that we can convert a hitting set problem on a well-behaved range space to an ε-net problem by a reweighting algorithm in polynomial time.

4 Further Reading Material

1. Har-Peled, Sariel. *Geometric Approximation Algorithms*. American Mathematical Soc., 2011. `http://sarielhp.org/book/`