Applications of VC-Dimension

1 Polytope Preprocessing

Given a polytope $P \subseteq \mathbb{R}^d$, we want to preprocess it to efficiently answer the following type of queries: given a hyperplane characterized by an inequality $\boldsymbol{c}^\intercal \boldsymbol{x} \leq \gamma$, decide if the inequality holds for all $\boldsymbol{x} \in P$ or is violated. This is important for linear programming: the constraints are given ahead of time, and we get different objective functions characterized by $\boldsymbol{c}$.

We'd like to either correctly state that $\boldsymbol{c}^\intercal \boldsymbol{x} \leq \gamma$ for all $\boldsymbol{x} \in P$ or return a witness $\boldsymbol{x}$ with $\boldsymbol{c}^\intercal \boldsymbol{x} > \gamma$. An approximately correct answer would guarantee that $\boldsymbol{c}^\intercal \boldsymbol{x} \leq \gamma$ holds for a $(1-\varepsilon)$ fraction of P, or return a witness.

Our preprocessing algorithm will compute an ε-net N for P by random sampling. When a query $(\boldsymbol{c}, \gamma)$ arrives, we check the inequality for all $\boldsymbol{x} \in \mathbb{N}$. If one such $\boldsymbol{x}$ fails (*i.e.,* $\boldsymbol{c}^\intercal \boldsymbol{x} > \gamma$) return it as a witness; otherwise, return that most of P is on the correct side of the hyperplane. This works because, by definition, for each $S \subseteq P$ that is separated from $P \setminus S$ by a hyperplane such that $vol(S) \geq \varepsilon \cdot vol(P)$, we have $N \cap S \neq \emptyset$. In particular, this would have to hold for S defined by $(\boldsymbol{c}, \gamma)$. So if $(\boldsymbol{c}, \gamma)$ does not produce a witness, then the volume of its subset of P is at most $\varepsilon \cdot vol(P)$.

We have that

$$|N| = \Theta\big(\frac{d}{\varepsilon} \log \frac{d}{\varepsilon} + \frac{1}{\varepsilon} \log \frac{1}{\delta}\big)$$

is enough by the ε-net theorem, and N can be found by i.i.d. sampling from P. This finds an ε-net with probability at least $1-\delta$. In particular, if d, ε, and δ are constant, then $\mathcal{O}(1)$ samples is enough regardless of how complex the polytope is.

The fact that P is a polytope matters only in that it allows us to efficiently sample (nearly) uniformly. Note that we draw N once and then reuse it, even for adversarially chosen $(\boldsymbol{c}, \gamma)$. It's trivial to get the same guarantee if we choose N for any particular $(\boldsymbol{c}, \gamma)$.

2 Approximate Centerpoints

A centerpoint is one natural generalization of the median to higher dimensions. Let P be a point set in $\mathbb{R}^d$. We want a point $\boldsymbol{x} \in \mathbb{R}^d$ such that each hyperplane through $\boldsymbol{x}$ splits P in half. However, such a point may not exist. So we relax the definition to require a $\frac{1}{d+1}$ fraction of points on each side. This is tight on the d-dimensional simplex.

We will prove existence using the following theorem:

Theorem 1. *[Helly]. Let $X_1, X_2, \ldots, X_m \subseteq \mathbb{R}^d$ be convex sets such that any $d+1$ of them have nonempty intersection. Then the intersection of all of them is also nonempty.*

Now consider each set $P' \subseteq P$ with $|P'| < \frac{1}{d+1}|P|$ such that P' can be separated from $P \setminus P'$ by a hyperplane. For any such P', let $X_{P'}$ be the set of all points $\boldsymbol{v}$ such that no hyperplane through $\boldsymbol{x}$ separates P' from $P \setminus P'$. Each $X_{P'}$ is a convex set.

For any point $\boldsymbol{x} \in P \setminus P'$, consider a hyperplane $(\boldsymbol{c}, \gamma)$ such that $\boldsymbol{c}^\intercal \boldsymbol{y} \geq \gamma$ for all $\boldsymbol{y} \in P'$. Then because the hyperplane goes through $\boldsymbol{x}$ we have $\boldsymbol{c}^\intercal \boldsymbol{x} \geq \gamma$, so $\boldsymbol{x}$ is on the same side as P'. Thus, there is no hyperplane through $\boldsymbol{x}$ separating P' from $P \setminus P'$. This implies $X_{P'} \supseteq P \setminus P'$. Therefore,

$$|X_{P'}| > (1 - \frac{1}{d+1})|P|.$$

By a union bound, the intersection of the $X_{P_i'}$ is nonempty for any subset S of sets P_i' when $|S| \leq d+1$. And, by Helly's Theorem, the intersection of *all* the $X_{P_i'}$ is nonempty, and any point in the intersection is a centerpoint.

Now we know that the centerpoint actually exists. So we define the ε-approximate centerpoint as a point $\boldsymbol{x} \in \mathbb{R}^d$ such that no hyperplane through $\boldsymbol{x}$ separates P' from $P \setminus P'$ such that

$$|P'| < (\frac{1}{d+1} - \varepsilon)|P|.$$

We care about approximate centerpoints because the best-known algorithm for computing a centerpoint in d dimensions takes time $\mathcal{O}(n^{d-1})$, which is exponential in d. A simple algorithm for finding an ε-approximate centerpoint is to sample "enough" points from P i.i.d. uniformly at random; call this sample S. Then, use "brute force" to compute a centerpoint $\boldsymbol{x}$ of S.

We will want

$$\Theta(\frac{d^2}{\varepsilon^2} \log \frac{d}{\varepsilon} + \frac{1}{\varepsilon^2} \log \frac{1}{\delta})$$

points. That is, we will need an ε-sample. For the "brute force" algorithm, we will write an LP for $\boldsymbol{x}$ with a constraint for each subset $S' \subseteq S$ with $|S'| < \frac{1}{d+1}|S|$. This will be $\Theta(|S|^d)$ constraints, but this is independent of n with constants d, ε, and δ.

Why is a centerpoint of S an ε-approximate centerpoint of P? Assume that S is an ε-sample of P *w.r.t.* hyperplanes. By the ε-sample theorem, this happens with probability at least $1-\delta$. Assume for a contradiction that there exists $P' \subseteq P$ with $|P'| < (\frac{1}{d+1} - \varepsilon)|P|$ and a hyperplane through $\boldsymbol{x}$ that separates P' from $P \setminus P'$. That is,

$$\frac{|P'|}{|P|} < (\frac{1}{d+1} - \varepsilon)$$

and P' is obtained from a hyperplane. Because S is an ε-sample,

$$|\frac{|P' \cap S|}{|S|} - \frac{|P'|}{|P|}| \leq \varepsilon,$$

which implies

$$\frac{|P' \cap S|}{|S|} \leq \frac{|P'|}{|P|} + \varepsilon < \frac{1}{d+1}.$$

Therefore, we have a set $P' \cap S$ of size less than $\frac{1}{d+1}|S|$ that is obtained by a hyperplane through $\boldsymbol{x}$. But then $\boldsymbol{x}$ is not a centerpoint for S.

Moral. Drawing a constant-sized sample and computing on it can result in approximately correct answers.

3 Network Failure Detection

We are given an undirected, connected network $G = (V, E)$. We'd like to place monitors on a set D of nodes so as to detect "bad" network failures from the fact that at least one par of monitors became disconnected.

A "bad" network failure means that there is a set S with $|S|, |\bar{S}| \geq \varepsilon n$ such that S and $\bar{S}$ have been disconnected. A network failure means up to k edges are removed.

The goal is to report all "bad" network failures (and optionally more). Subject to meeting this goal, we'd like to minimize the size of D.

We will make D a set of

$$\Theta(\frac{k}{\varepsilon}\log\frac{k}{\varepsilon}+\frac{1}{\varepsilon}\log\frac{1}{\delta})$$

nodes sampled i.i.d. (The size of D is independent of n). Then with probability at least $1-\delta$, we detect every "bad" network failure.

Proof. Define a set system. We call a set $S \subseteq V$ separable by k edges iff there exists a set $F \subseteq E$ with $|F| \leq k$ of edges such that S and $\bar{S}$ are not connected in $(V, E \setminus F)$. In other words, the adversary could disconnect S from $\bar{S}$ by deleting at most k edges. Let $\mathcal{S}_k$ be the set system of all $S \subseteq V$ that are separable by k edges.

Lemma 2. *$(V, \mathcal{S}_k)$ has VC-dimension at most $2k+1$.*

Proof. Next class. □

Using the lemma, D is an ε-net for $(V, \mathcal{S}_k)$ with probability at least $1-\delta$. By definition, $D \cap S$ is nonempty for all $S \in \mathcal{S}_k$ with $|S| \geq \varepsilon n$. Thus, all "bad" network failures are detected. □