CS 395T Computational Complexity of Machine Learning

Lecture 11: February 24, 2005 — Lecturer: Adam Klivans

Scribe: Justin Brickell

11.1 Hardness Results for Learning Intersection of Halfspaces

Recall from the previous lecture the basic form of proving that a problem P is NP-hard by reduction. We take a problem Q that is known to be NP-hard, and *reduce* Q to P, *i.e.,* we show that given a solution to P, in polynomial time we can derive a solution to Q.

Given a hypergraph $H = (V, E)$ with $|V| = n$ we will provide a set of labelled examples F that is consistent with an intersection of l halfspaces, provided that H is l-colorable. First, some notation:

$$\text{For } v_i \in V, \qquad a(v_i) = \langle 0, ..., 0, 1, 0, ..., 0 \rangle, \text{ where the 1 occurs at position } i.$$
$$\text{For } e = (v_j, v_k, v_l) \in E, \qquad a(e) = a(v_j) + a(v_k) + a(v_l).$$

Now we provide the following set of labelled examples:

$$\begin{cases} (O^n, +) \\ (a(v_i), -) \text{ for all } v_i \in V \\ (a(e), +) \text{ for all } e \in E \end{cases}$$

Assume that H is l-colorable (*i.e.* V is colorable by l colors such that no edge is monochromatic). Then there is a function χ mapping vertices to the integers $1...l$. Consider the following halfspaces for $j \in 1...l$:

$$h_j(x) = \text{SIGN}\left(\sum_{i=1}^{n} w_{ji} x_i + \frac{1}{2}\right),$$

where w_{ji} is equal to -1 if $\chi(v_i) = j$ and is equal to n otherwise. For convenience, denote

$$g_j(x) = \sum_{i=1}^{n} w_{ji} x_i + \frac{1}{2}.$$

The intersection of these halfspaces is given by

$$h(x) = \bigwedge_{i=1}^{l} h_i(x).$$

Claim 1 *The intersection of halfspaces h is consistent with the examples.*

Proof: We will consider all three classes of examples in turn.

1. $(O^n, +)$
   In this case, $g_j(x) = \frac{1}{2}$ for all j, so $h_j(x)$ is positive (true) for all j.
2. $(a(v_i), -)$ for all $v_i \in V$
   Note that v_i must have some color b. Then by definition, $w_{bv_i} = -1$, so $h_b(a(v_i))$ is negative (false).
3. $(a(e), +)$ for all $e \in E$
   For all colors c, h_c is positive (true) because at least one of the w_{ci} must be equal to n (and not -1) due to the proper coloring.

We conclude that h is consistent with the examples. ■

Given an intersection of l halfspaces consistent with the examples, we can l-color H as follows:

For vertex v, consider $a(v)$, and color v equal to the index of the first halfspace that makes $a(v)$ false.

Claim 2 *This is a valid l-coloring of H.*

Proof: First, note that every halfspace has a positive threshold because of the $(O^n, +)$ example. Now, assume for the sake of a contradiction that some edge $e = (v_j, v_k, v_l)$ is *not* properly colored. Then $\chi(v_j) = \chi(v_k) = \chi(v_l) = c$. But then $h_c(a(e))$ must be negative, which is a contradiction. ■

11.2 Hardness results for learning 2-term DNFs

In this section, we will show that it is NP-hard to learn a 2-term DNF by an l-term DNF for any constant $l > 0$. There are two related problems that are currently open:

- Is it NP-hard to learn a 2-term DNF by an n-term DNF?
- Is it NP-hard to learn an intersection of 2 halfspaces by an intersection of n halfspaces?

Given a hypergraph $H = (V, E)$ with $|V| = n$ we will provide a set of labelled examples F that is consistent with an l-term DNF, provided that H is l-colorable. First, some notation:

$$\begin{aligned} \text{For } v_i \in V, \quad & b(v_i) = \langle 1, ..., 1, 0, 1, ..., 1 \rangle, \text{ where the 0 occurs at position } i. \\ \text{For } e = (v_j, v_k, v_l) \in E, \quad & b(e) = b(v_j) \wedge b(v_k) \wedge b(v_l). \end{aligned}$$

Next we provide the following set of labelled examples:

$$\begin{cases} (O^n, -) \\ (b(v_i), +) \text{ for all } v_i \in V \\ (b(e), -) \text{ for all } e \in E \end{cases}$$

Now we are ready to construct t, an l-term DNF that is consistent with the examples. For each color c, let

$$t_c = \bigwedge_{x_i \text{ not colored c}} x_i.$$

Now define t as

$$t = \bigvee_c t_c.$$

Claim 3 *The l-term DNF t is consistent with the examples.*

Proof: We will consider all three classes of examples in turn.

1. $(O^n, -)$
   t clearly evaluates to false.
2. $(b(v_i), +)$ for all $v_i \in V$
   Assume v_i is colored α. Then t_α will be true on $b(v_i)$.
3. $(b(e), -)$ for all $e \in E$
   For every color α, e must have at least one vertex that is not colored α. Therefore t_α is false on $b(e)$. Since this is true for all α, we conclude that t is false on $b(e)$.

We conclude that t is consistent with the examples. ■

Given an intersection of l-term DNF consistent with the examples, we can l-color H as follows:

For vertex v, consider $b(v)$, and color v equal to the index of the first term that satisfied by $b(v)$.

This is a valid coloring because if some edge e consisted of vertices i, j, k all colored α, then $b(e)$ would satisfy t_α, which is a contradiction.

11.3 Miscellaneous Hardness Results

Is it NP-hard to learn the intersection of 2 halfspaces by the intersection of n halfspaces?

This is an open problem, but the answer is probably "yes."

Here are some other hardness results whose proofs are not given in this course:

- For all $\epsilon > 0$, for all $c > 0$, it is NP-hard to learn an intersection of n^ϵ halfspaces by an intersection of n^c halfspaces.
- It is NP-hard to estimate $\chi(G)$ [The chromatic number of G] to within a $1 - \epsilon$ factor for all $\epsilon > 0$ *i.e.*, it is NP-hard to $N^{1-\epsilon}$ color an N^ϵ colorable graph.
- It is NP-hard to learn an n^ϵ-term DNF by an intersection of n^c halfspaces.

However, it is possible to learn an n^ϵ-term DNF by an $n^{\sqrt{n}}$-term DNF.

11.4 Uniform Distribution Learning

This is a learning model equivalent to PAC learning with regard to the uniform distribution. Assume the learner receives labelled examples chosen uniformly at random over $\{0,1\}^n$. With high probability (over the randomness of the learner and the received examples) the learner outputs h such that

$$\Pr_{x \in \{0,1\}^n}[h(x) = f(x)]$$

is large.

11.5 Basis for functions

In this section we will describe two different basis for all functions $f : \{0,1\}^n \to \mathbb{R}$.

1. The "term" basis.
   Has 2^n basis vectors. For a string $S \in \{0,1\}^n$, $t_s(x) = 1$ if $x = s$ and 0 otherwise.

2. The "parity" basis.
   For each $s \in [n]$,

$$\chi_s(x) = \begin{cases} 1 & \text{if } \sum_{i \in S} x_i \bmod 2 = 0 \\ -1 & \text{if } \sum_{i \in S} x_i \bmod 2 = 1 \end{cases}$$

   For f, g define

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) \cdot g(x) = \mathrm{E}[f(x) \cdot g(x)],$$

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

   Then any f can be uniquely expressed as

$$\sum_{s \subseteq [n]} \hat{f}(s) \chi_s(x),$$

   where $\hat{f}(s) = \langle f, \chi_s \rangle = \mathrm{E}[f \cdot \chi_s]$.