CS 395T Computational Complexity of Machine Learning

Lecture 8: Febrary 15, 2005 — Lecturer: Adam Klivans

Scribe: Dongmin

8.1 Learning Intersections of Halfspaces

8.1.1 Observation:

Consider the following assumptions,

- m points.
- R^n.

By Brute Force Search, we can find a intersection of t halfspaces consistent with m points with the running time bounds $O(m^{nt})$.

To do so, firstly, look at all subsets of m points of size $n+1$. (Set of all combinatorially distinct halfspaces) Then, look at all t-size subsets of halfspaces to find a consistent $h_1 \wedge \cdots \wedge h_t$.

8.2 Johnson-Lindenstrauss Theorem

Theorem 1 *Assume m points in R^n, $\{v^{(1)}, \cdots, v^{(m)}\} \subset R^n$. Then there exists a randomized mapping $\phi : R^n \to R^k$, with $k = O(\frac{\log m}{\epsilon^2})$ such that, $\forall\ v^{(i)}, v^{(j)}$ with high probability,*

$$(1-\epsilon) \leq \frac{\|\phi(v^{(i)}) - \phi(v^{(j)})\|_2^2}{\|v^{(i)} - v^{(j)}\|_2^2} \leq (1+\epsilon). \tag{8.1}$$

Proof: Let $v^{(1)}, \cdots, v^{(m)} \in R^n$. Choose $r^{(1)}, \cdots, r^{(k)} \in R^n$ with $r_j^{(i)} \sim N(0,1)$ and let $\phi(v) = (\langle v, r^{(1)}\rangle, \cdots, \langle v, r^{(k)}\rangle) \in R^k$.

Define $X = \langle v, r\rangle$. Then,

$$E[X] = E\left[\sum_{i=1}^{n} v_i r_i\right] = \sum_{i=1}^{n} v_i E[r_i] = 0, \tag{8.2}$$

$$Var[X] = Var\left[\sum_{i=1}^{n} v_i r_i\right] = \sum_{i=1}^{n} v_i^2 Var[r_i] = \sum_{i=1}^{n} v_i^2 = \|v\|_2^2. \tag{8.3}$$

And,

$$\begin{aligned}
E\left[\|\phi(v)\|_2^2\right] &= E\left[\sum_{i=1}^{k}\langle v, r^{(i)}\rangle^2\right] \\
&= \sum_{i=1}^{k} E\left[\langle v, r^{(i)}\rangle^2\right] \\
&= \sum_{i=1}^{k} E\left[\sum_{j=1}^{n}\left(v_j r_j^{(i)}\right)^2 + 2\sum_{p\neq q} v_p v_q r_p^{(i)} r_q^{(i)}\right] \\
&= \sum_{i=1}^{k}\left(\|v\|_2^2 + \sum_{p\neq q} v_p v_q E[r_p^{(i)}]E[r_q^{(i)}]\right) \\
&\quad \left(\because (\ 8.3) \text{ and the independence of } r_p^{(i)} \text{ and } r_q^{(i)}.\right) \\
&= k\|v\|_2^2. \qquad (8.4)
\end{aligned}$$

We will prove $E\left[\frac{\|\phi(u)-\phi(v)\|_2^2}{\|u-v\|_2^2}\right] = 1$ with high probability. In other words, $\|\phi(v)\|_2^2 = k\|v\|_2^2$ with probability $\geq 1 - \frac{1}{2m^2}$.

Rescale so that $\|v\|_2^2 = 1$, let $X_i = \langle v, r^{(i)}\rangle$, $X = \sum_{i=1}^{k} X_i^2$. Then, $X_i \sim N(0,1)$ and $E[X] = k$. Let $\alpha = (1+\epsilon)k$:

$$Pr[X \geq \alpha] \leq e^{\left(-\frac{k}{2}\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)\right)}. \qquad (8.5)$$

Therefore, if

$$k = \frac{4\log m}{\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)},$$

Then

$$Pr[X \geq (1+\epsilon)k] \leq \frac{1}{m^2}. \qquad (8.6)$$

Analogously,

$$Pr[X \leq (1-\epsilon)k] \leq \frac{1}{m^2}. \qquad (8.7)$$

By rescaling ϕ properly, the theorem holds.

* For the details of this proof please see Anupam Gupta's lecture notes #15, available from his web-page.

■

8.3 Learning Intersections of Halfspaces with Margin

Assume that the margin of two hyperplane is ρ. then, $\forall x \in S^{n-1}$, $\|x\|_2 = 1$, and $1 \geq |h_1(x)| \geq \rho$, $1 \geq |h_2(x)| \geq \rho$.

Consider applying random projection matrix to data $X \subset R^n$. Let $k = O\left(\frac{\log |X|}{\rho^2}\right)$ for random projection matrix $M \in R^{n \times k}$.

Then, the image of X w.r.t M: $M(X)$ classified by $h'_1 \wedge h'_2$ in R^k with margin $\frac{\rho}{2}$.

Now, we want to learn $h'_1 \wedge h'_2$ in R^k. Here, use Brute Force Search again.

- m examples.
- $(m)^{\frac{\log m}{\rho^2}}$ running time.

If we choose $m \approx \frac{\log\left(\frac{1}{\rho}\right)}{\rho^2 \gamma}$, Then, the total running time is approximately $O\left(\frac{1}{\rho^2 \gamma}\right)^{\frac{1}{\rho^2}} \cdot n^{O(1)}$.

For t halfspaces,

$$O\left(\frac{1}{\rho^2 \gamma}\right)^{\frac{t}{\rho^2}} \cdot n^{O(1)}. \tag{8.8}$$

References

[1] R. I. Arriaga and S. Vempala, An Algorithmic Theory of Learning: Robust Concepts and Random Projection, In *Proceedings of the 40th Annual Symposium on Foundation of Computer Science*, 1999.