CSE 291 Scribe Notes Lecture 16

Zeyu Chen

May 15, 2017

Abstract

How to computational efficiently find an approximation of p using $q \in \mathcal{C}$ that minimizes the A_k distance: $|q - \hat{p}|_{A_k}$, where $\mathcal{C}$ is the family of distributions that are t-piecewise degree-d polynomial(t-piecewise degree-d polynomial: the $\mathbb{R}$ can be split into t pieces, and each piece the probability density function is a degree d polynomial)

1 Previous Lecture

We introduced the $|q - p|_{A_k}$ distance between p and q, which is $\frac{1}{2}\sum_I |q(I) - p(I)|$. If those two distributions have only a few crossings, then A_k distance can be a good standing of d_{TV}(total variation distance). A_k distance can be measured in a small VC-dimension which means that we can have a better bound in the number of samples to measure the distance. If we can split the $\mathbb{R}$ into $O(d(t+1))$ pieces and measure the A_k distance by catching the sign changes between p and q, then A_k is the same as d_{Tv}. If we cannot let p be exactly t-piecewise polynomial, we can $\delta - approximate$ p in error of $O(\delta + \epsilon)$ with time and sample complexity $O(t(d+1)/\epsilon^2)$.

2 Algorithm

1. Take Samples
2. Compute the empirical distribution $\hat{p}$
3. Find q t-piecewise degree-d that minimizes $|q - \hat{p}|_{A_k}$
4. If non-proper hypothesis q, return hypothesis not in $\mathcal{C}$

The algorithm above could end up with the case of non-proper hypothesis, but we can round the non-proper result to a proper result with a doubled error; however, there is not known algorithm to do that efficiently.

This still leads to the question of how to perform step 3 efficiently.

3 Examples

3.1 Special Case: t=1

First we will look at the special case where $t = 1$. We want to find degree-d polynomial q on $[0, 1]$ such that for any partition of $[0, 1]$ into intervals $I_1, I_2, ..., I_k$, $\sum_I |q(I_i) - \hat{p}(I_j)| < \epsilon$. We also assume that p is close to $\hat{p}$ in A_k distance and there exists some p that is close to $\hat{p}$ in A_k distance.

We can formulate this problem into a Linear Program. (Linear Program: A system of linear inequalities with some number of variables. Optimize the objective function with constraints like: $v_i \cdot x \geq b_i$, WiKi_LP)

The good side of linear program is that we know: (Theorem) there exists a polynomial time algorithm to find the solution; however, there might be infinite number of intervals (inequalities) that we need to consider for our LP, so we will end up with a horrible runtime.

It turns out that we do not actually need a list of equations for this algorithm to work. It is sufficient to have a separation oracle(a special version of LP *Lec_SO*), which can be done in the following:

- Given **X** return either
  - **X** is a solution
  - some constraint that **X** violates

In order to compute the A_k distance $|p-q|_{A_k}$ for some nice distribution p and q, we want to

1. partition intervals to break at where $p(x) = q(x)$[reduce to only finite number of end-points](There are d crossings in one interval, with t piece),
2. find the best intervals that minimize $\sum_{j=1}^{M} |p(I_j) - q(I_j)|$ using Dynamic Programming, where I_m ends at x.

Then, we can get some optimal partitions of $(I_1, I_2, ..., I_m)$ through Dynamic Programming by comparing on what is the best discrepancy between merging I_j and I_{j+1} and just I_j, then we can get the set of intervals that minimize the discrepancy. This dynamic programming also gives us a Separation Oracle (if the partition at X will give you some $A_k < \epsilon$) and we can apply Linear Program to see if there is such a q that $A_k distance < \epsilon$ and minimizes A_k

3.2 A more general case t>1

How to approximate (no exact solution to the minimum $A_k distance$)some distribution p where we assume $p =$ some t-piecewise degree-d polynomial There is no algorithm to minimize A_k distance but only to find good enough A_k distance approximation of p since we can always have a finer partition (larger t) to minimize the A_k distance. We also need $O(Nlog(\frac{k}{\epsilon}))$ samples as we need to take the union bound over $\frac{k}{\epsilon}$ terms.

As above we split $\mathbb{R}$ into intervals $J_1, J_2, ..., J_m$ where $P(J_i) \approx \frac{\epsilon}{k}$, where $k = 2t(d+1)$. We can do this if we have access to p or partition it approximate if we can access $\hat{p}$
But we need to know how good this approximation is Let $J_{a-b} = J_a \cup J_{a+1} \cup J_{a+2} ... \cup J_b$. Suppose we have N samples, and our empirical distribution $\hat{p}$ has an error between $\hat{p}(J_{a-b}) = p(J_{a-b}) \pm \sqrt{\frac{\frac{\epsilon}{k}(b-a+1)}{N}}$. Let $\sqrt{\frac{\frac{\epsilon}{k}(b-a+1)}{N}} = \sqrt{b-a+1}\delta$.

For $q \geq 0$ $|q(J_{a-b}) - \hat{p}(J_{a-b})| < \delta\sqrt{(b-a+1)}$ $\forall a \leq b$ where a-b does not cross a boundary for q. $|q(J_{a-b}) - p(J_{a-b})| < 2\delta\sqrt{(b-a+1)}$ (fix divider for the partition).

But how do we compute the $|q-p|_{A_k}$We partition the domain into k intervals and round the interval I_j's endpoints to some J_i's and this introduces $O(\epsilon)$ error.

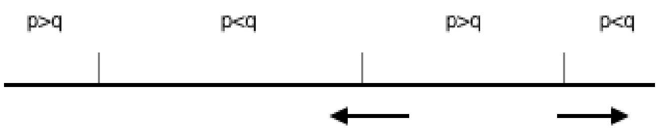

Figure 1: round the interval's endpoints where $p > q$ to intervals where $p < q$ so that we can end up with introducing $2(p(I_{err} - q(I_{err})))$, where $p(I_{err}) \leq \frac{\epsilon}{k}$, so for k intervals, we at introduce $O(\epsilon)$ error in total

After round all the endpoints, let $I_1 = J_{1-a_1}, I_2 = J_{a_1+1-a_2}, ..., |p-q|_{A_k} \leq O(\epsilon) + \sum 2\delta\sqrt{a_{j+1} - a_j} \leq O(\epsilon) + O(\delta)\sqrt{\sum a_{j+1} - a_j}\sqrt{k}$(Cauchy-Schwartz). Thus $|p-q|_{A_k} \leq O(\epsilon + \frac{\delta k}{\sqrt{\epsilon}})$.

Since $\delta = \sqrt{\frac{\epsilon}{kN}}$ and we need to get $\frac{\delta k}{\sqrt{\epsilon}} = \epsilon \rightarrow \delta = \frac{\epsilon^{\frac{3}{2}}}{k}$. By the equation $\frac{\epsilon^{\frac{3}{2}}}{k} = \sqrt{\frac{\epsilon}{kN}}$, we get $N = O(\frac{k}{\epsilon^2})$.

With the fixed dividers, we need to find where q needs to break at those dividers. Assume interval boundaries are at the boundaries at the J's.

1. We can find if there is a single degree-d polynomial q works on some J_{a-b} using Linear Program.
2. For each $m, 0 < m \leq t$, what is the largest interval J_{1-x}, such that can be done with m-piecewise polynomial q (how big could the piece be for some piece m).