Machine Learning **Fall 2023**

Lecture 4: VC Theory

Lecturer: Liwei Wang *Scribe: San Zhang, Si Li*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

4.1 All Pairs Shortest Paths

The naive and obvious solution to All Pairs Shortest Path(APSP) problem is to run a Single Source Shortest Path algorithm from each starting vertex v. If the graph has arbitrary edge weights, it takes the Bellman-Ford algorithm $O(|E||V|^2)$ time to solve APSP. But there are better approaches.

4.1.1 Floyd-Warshall Algorithm: Dynamic Programming

Label the vertices $1, 2, \ldots, n$. Define $d^{(k)}(i, j)$ to be the length of a shortest path from i to j, using intermediate vertices from $\{1, 2, \ldots, k\}$ only. Obviously, $d^{(n)}(i, j)$ is the full problem.

...

4.2 Transitive Closure

Our goal is to achieve running time $O(M(n) \log n)$ for APSP where $M(n)$ is the time for $n \times n$ matrix multiplication. Let's see if we can achieve this for a simpler but related problem, namely *Transitive Closure*:

...

4.3 Infinite Hypothesis Space

Note that we still have:

$$\begin{aligned} &\Pr[\Pr_D(Y \neq \hat{f}(X) - \frac{1}{n}\sum_{i=1}^{n} I[Y_i \neq \hat{f}(X_i)] \geq \epsilon] \leq \\ &\Pr[\exists f \in \mathcal{F}, \Pr_D(Y \neq f(X)) - \frac{1}{n}\sum_{i=1}^{n} I[Y_i \neq f(X_i)] \geq \epsilon] \end{aligned} \tag{4.1}$$

4.3.1 Step I: Double Sample Trick

Lemma 4.1 *Consider $2n$ iid random variables $X_1, ..., X_n, X_{n+1}, ..., X_{2n}$ with $EX_i = p$. Let $\nu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \nu_2 = \frac{1}{n}\sum_{i=n+1}^{2n} X_i$. For $n \geq \frac{\ln 2}{\epsilon^2}$, we have:*

$$\frac{1}{2}\Pr(|\nu_1 - p| \geq 2\epsilon) \leq \Pr(|\nu_1 - \nu_2| \geq \epsilon) \leq 2\Pr(|\nu_1 - p| \geq \frac{1}{2}\epsilon)$$

Proof: For the second part, note that

$$\Pr(|\nu_1 - \nu_2| \geq \epsilon) \leq \Pr(|\nu_1 - p| \geq \frac{\epsilon}{2} \vee |\nu_2 - p| \geq \frac{\epsilon}{2})$$

For the first part, if $|\nu_1 - p| \geq 2\epsilon, |\nu_2 - p| \leq \epsilon$, we will always have $|\nu_1 - \nu_2| \geq \epsilon$. Therefore,

$$\Pr(|\nu_1 - \nu_2| \geq \epsilon) \geq \Pr(|\nu_1 - p| \geq 2\epsilon)\Pr(|\nu_2 - p| \leq \epsilon)$$

. ■

Therefore, according to this lemma, we have:

$$\begin{aligned}\Pr[\exists f \in \mathcal{F}, \Pr_D(Y \neq f(X)) - \frac{1}{n}\sum_{i=1}^{n} I[Y_i \neq f(X_i)] \geq \epsilon] \leq \\ 2\Pr[\exists f \in \mathcal{F}, \frac{1}{n}\sum_{i=1}^{n} I[Y_i \neq f(X_i)] - \frac{1}{n}\sum_{i=n+1}^{2n} I[Y_i \neq f(X_i)] \geq \frac{\epsilon}{2}]\end{aligned} \tag{4.2}$$

4.3.1.1 Step II: Sample and Permute

When drawing (x_i, y_i), we can follow thess two steps: first draw an unordered set $z_1, ..., z_{2n}(z_i = (x_i, y_i)$ and second generate a random permutation $\sigma \in S_{2n}$ as the order. With this method, we have:

$$\begin{aligned}2\Pr[\exists f \in \mathcal{F}, \frac{1}{n}\sum_{i=1}^{n} I[Y_i \neq f(X_i)] - \frac{1}{n}\sum_{i=n+1}^{2n} I[Y_i \neq f(X_i)] \geq \frac{\epsilon}{2}] = \\ 2\mathbb{E}_{(z_1,...,z_{2n})}\{\Pr_{\sigma \in S_{2n}}[\exists f \in \mathcal{F}, \frac{1}{n}\sum_{i=1}^{n} I[Y_{\sigma(i)} \neq f(X_{\sigma(i)})] - \frac{1}{n}\sum_{i=n+1}^{2n} I[Y_{\sigma(i)} \neq f(X_{\sigma(i)})] \geq \frac{\epsilon}{2}]\}\end{aligned} \tag{4.3}$$

References

[AGM97] N. Alon, Z. Galil and O. Margalit, On the Exponent of the All Pairs Shortest Path Problem, *Journal of Computer and System Sciences* **54** (1997), pp. 255–262.

[F76] M. L. Fredman, New Bounds on the Complexity of the Shortest Path Problem, *SIAM Journal on Computing* **5** (1976), pp. 83-89.