PROOF OF THE VC INEQUALITY

Hsuan-Tien Lin and Yaser S. Abu-Mostafa

Notations

$\underline{x} \in X^{2N}, \underline{x} = \underline{x}'\underline{x}'', |\underline{x}'| = |\underline{x}''| = N$

ν'_g : frequency of agreement of g on $\underline{x}'$

ν''_g : frequency of agreement of g on $\underline{x}''$

π_g : probability of agreement of g

G : a learning model

$L(\underline{x})$: the number of different patterns of G on $\underline{x}$

$L(N)$: the growth function of G

$G[\underline{x}]$: a subset of G that "represents" $L(\underline{x})$

$$\left\{\text{statement}\right\} = \begin{cases} 1 & \text{if statement is true} \\ 0 & \text{if statement is false} \end{cases}$$

$\Pr_p[\cdot]$ is w.r.t. $p(\underline{x})$. Thus,

$$\Pr_p\left[\text{statement}\right] = \sum_{\underline{x}} p(\underline{x}) \left\{\text{statement}\right\} = \sum_{\underline{x}'} \sum_{\underline{x}''} p(\underline{x}')p(\underline{x}'') \left\{\text{statement}\right\}$$

$\Pr[\cdot]$ is a short-hand of $\Pr_p[\cdot]$

Basic Equations

- Hoeffding's inequality: $\Pr\left[\left|\nu' - \pi\right| > \epsilon\right] \le 2\exp(-2\epsilon^2 N)$.

- Bin model: $\Pr\left[\nu' = \frac{K}{N}\right] = \binom{N}{K}\pi^K(1-\pi)^{N-K}$.
  Note that the term peaks around $K = \pi N$. For example, if πN is an integer, we could easily verify that for $\pi \in (0,1)$,

$$\frac{\Pr\left[\nu'=\frac{K+1}{N}\right]}{\Pr\left[\nu'=\frac{K}{N}\right]} < 1 \text{ and } \frac{\Pr\left[\nu'=\frac{K-1}{N}\right]}{\Pr\left[\nu'=\frac{K}{N}\right]} < 1.$$

Lemma 1 *(bounding the deviation with two half-vectors)*

$$\Pr\left[\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right] \ge \frac{1}{2}\Pr\left[\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right]$$

for N such that $\exp(-\frac{1}{2}\epsilon^2 N) \le \frac{1}{4}$.

Proof.

$$\begin{aligned}
& \Pr\left[\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right] \\
= & \sum_{\underline{x}'}\sum_{\underline{x}''} p(\underline{x}')p(\underline{x}'') \left\{\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
\ge & \sum_{\underline{x}'}\sum_{\underline{x}''} p(\underline{x}')p(\underline{x}'') \left\{\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\}\left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \\
= & \sum_{\underline{x}'} p(\underline{x}') \left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \sum_{\underline{x}''} p(\underline{x}'') \left\{\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
& \text{(for any } \mathbf{g} \text{ to be chosen later -- may depend on } \underline{x}'\text{)} \\
\ge & \sum_{\underline{x}'} p(\underline{x}') \left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \sum_{\underline{x}''} p(\underline{x}'') \left\{\left|\nu'_{\mathbf{g}} - \nu''_{\mathbf{g}}\right| > \tfrac{\epsilon}{2}\right\} \\
& \text{(restrict to some special cases that achieve the latter } \{\cdot\}\text{)} \\
\ge & \sum_{\underline{x}'} p(\underline{x}') \left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \sum_{\underline{x}''} p(\underline{x}'') \left\{\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \epsilon \text{ and } \left|\nu''_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \tfrac{\epsilon}{2}\right\} \\
= & \sum_{\underline{x}'} p(\underline{x}') \underbrace{\left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\}}_{(*)} \left\{\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \epsilon\right\} \sum_{\underline{x}''} p(\underline{x}'') \left\{\left|\nu''_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \tfrac{\epsilon}{2}\right\} \\
& \text{(if } (*) \text{ evaluates to 1, choose one } \mathbf{g} \text{ that achieves } (*)\text{; otherwise choose any } \mathbf{g}\text{)} \\
= & \sum_{\underline{x}'} p(\underline{x}') \left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \sum_{\underline{x}''} p(\underline{x}'') \left\{\left|\nu''_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \tfrac{\epsilon}{2}\right\} \\
& \text{(apply Hoeffding's inequality)} \\
\ge & \sum_{\underline{x}'} p(\underline{x}') \left\{\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right\} \left(1 - 2\exp(-\tfrac{1}{2}\epsilon^2 N)\right) \\
\ge & \frac{1}{2}\Pr\left[\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right]
\end{aligned}$$

■

Lemma 2 *(bounding the difference of half-vectors under specific conditions) Given any integer K from 0 to $2N$,*

$$\Pr\left[\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2} \middle| \nu'_g + \nu''_g = \tfrac{K}{N}\right] \le 2\exp(-\tfrac{1}{8}\epsilon^2 N).$$

Proof. The left-hand-side (LHS) is independent of π_g because

$$\Pr\left[\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2} \middle| \nu'_g + \nu''_g = \tfrac{K}{N}\right] = \frac{\binom{2N}{K}\sum_{i=0}^{2N}\binom{K}{i}\binom{2N-K}{N-i}\left\{\left|\frac{i}{N} - \frac{K-i}{N}\right| > \frac{\epsilon}{2}\right\}}{\binom{2N}{K}\sum_{i=0}^{2N}\binom{K}{i}\binom{2N-K}{N-i}}$$

From the equation above, we can see that under the condition of interest, if $\nu'_g = \frac{i}{N}$,

$$\left|\nu'_g - \nu''_g\right| = 2\left|\tfrac{i}{N} - \tfrac{K}{2N}\right|$$

Because LHS is independent of π_g, we can work on a pseudo-problem in which $\underline{x}$ is drawn from $q(\underline{x})$ and choose $\mathbf{g}$ such that $\pi_{\mathbf{g}} = \frac{K}{2N}$ without lose of generality. Then,

$$\left|\nu'_{\mathbf{g}} - \nu''_{\mathbf{g}}\right| = 2\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right|$$

Thus,

$$\begin{aligned}
\text{LHS} &= \Pr_q\left[\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \tfrac{\epsilon}{4} \middle| \nu'_{\mathbf{g}} + \nu''_{\mathbf{g}} = \tfrac{K}{N}\right] \\
&= \frac{\sum_{\underline{x}'}\sum_{\underline{x}''} q(\underline{x}')q(\underline{x}'')\left\{\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4}\right\}\left\{\nu'_{\mathbf{g}} + \nu''_{\mathbf{g}} = \frac{K}{N}\right\}}{\sum_{\underline{x}'}\sum_{\underline{x}''} q(\underline{x}')q(\underline{x}'')\left\{\nu'_{\mathbf{g}} + \nu''_{\mathbf{g}} = \frac{K}{N}\right\}} \\
&= \frac{\sum_{\underline{x}'} q(\underline{x}')\left\{\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4}\right\} f(\nu'_{\mathbf{g}})}{\sum_{\underline{x}'} q(\underline{x}') f(\nu'_{\mathbf{g}})},
\end{aligned}$$

where

$$f(\nu'_{\mathbf{g}}) = \sum_{\underline{x}''} q(\underline{x}'')\left\{\nu'_{\mathbf{g}} + \nu''_{\mathbf{g}} = \tfrac{K}{N}\right\} = \Pr_q\left[\nu''_{\mathbf{g}} = \tfrac{K}{N} - \nu'_{\mathbf{g}}\right].$$

Note that $f(\nu'_{\mathbf{g}})$ peaks around $\nu'_{\mathbf{g}} = \frac{K}{2N}$. Let

$$\sup_{\nu'_{\mathbf{g}}:\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \epsilon/4} f(\nu'_{\mathbf{g}}) = \alpha.$$

Then,

$$\left\{\begin{array}{lll} f(\nu'_{\mathbf{g}}) & \le & \alpha \quad \text{for } \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \\ f(\nu'_{\mathbf{g}}) & \ge & \alpha \quad \text{for } \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \frac{\epsilon}{4} \end{array}\right. .$$

Now,

$$
\begin{aligned}
\text{LHS} &= \frac{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} f(\nu'_{\mathbf{g}})}{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} f(\nu'_{\mathbf{g}}) + \sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \frac{\epsilon}{4} \right\} f(\nu'_{\mathbf{g}})} \\
&\quad \text{(apply } f(\nu'_{\mathbf{g}}) \ge \alpha \text{ under that condition)} \\
&\le \frac{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} f(\nu'_{\mathbf{g}})}{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} f(\nu'_{\mathbf{g}}) + \sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \frac{\epsilon}{4} \right\} \alpha} \\
&\quad \text{(use } \tfrac{A}{B} \le \tfrac{A+a}{B+a} \text{ when } A \le B \text{ and } A, B, a \text{ positive)} \\
&\le \frac{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} \alpha}{\sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right\} \alpha + \sum_{\underline{x}'} q(\underline{x}') \left\{ \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| \le \frac{\epsilon}{4} \right\} \alpha} \\
&= \Pr_q \left[\left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4} \right] \\
&\quad \text{(apply Hoeffding's inequality)} \\
&\le 2\exp(-\tfrac{1}{8}\epsilon^2 N)
\end{aligned}
$$

Note that we need $\alpha > 0$ above, which happens when the set $\left\{\nu'_{\mathbf{g}} : \left|\nu'_{\mathbf{g}} - \pi_{\mathbf{g}}\right| > \frac{\epsilon}{4}\right\}$ is nonempty. If the set is empty, LHS $= 0$ and the lemma is trivial.

■

Theorem 1 *(VC inequality) For N such that Lemma 1 is satisfied,*

$$\Pr\left[\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right] \le 4\cdot L(2N)\exp(-\tfrac{1}{8}\epsilon^2 N)$$

Proof.

$$\begin{aligned}
& \Pr\left[\sup_{g\in G}\left|\nu'_g - \pi_g\right| > \epsilon\right] \\
& \text{(apply Lemma 1)} \\
\le\ & 2\Pr\left[\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right] \\
=\ & 2\sum_{\underline{x}} p(\underline{x})\left\{\sup_{g\in G}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
=\ & 2\sum_{\underline{x}} p(\underline{x})\left\{\sup_{g\in G[\underline{x}]}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\}
\end{aligned}$$

We get the same probability if we first generate $\underline{x}_0$ according to $p(\underline{x}_0)$ and choose one of the $(2N)!$ permutations of $(1, 2, \cdots, 2N)$ to permute $\underline{x}_0$ into $\underline{x}$. Here $\underline{x}$ depends on $\underline{x}_0$ and the permutation $t = 1, 2, \cdots, (2N)!$, but $G[\underline{x}]$ and $G[\underline{x}_0]$ are equivalent.

$$\begin{aligned}
=\ & 2\sum_{\underline{x}_0} p(\underline{x}_0)\sum_{t=1}^{(2N)!}\frac{1}{(2N)!}\left\{\sup_{g\in G[\underline{x}_0]}\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
& \text{(a gross over-estimate)} \\
\le\ & 2\sum_{\underline{x}_0} p(\underline{x}_0)\sum_{t=1}^{(2N)!}\frac{1}{(2N)!}\sum_{g\in G[\underline{x}_0]}\left\{\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
=\ & 2\sum_{\underline{x}_0} p(\underline{x}_0)\sum_{g\in G[\underline{x}_0]}\sum_{t=1}^{(2N)!}\frac{1}{(2N)!}\left\{\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\right\} \\
& \text{(consider a distribution } q(\underline{x}) \text{ in which } q(\underline{x}) = \tfrac{1}{(2N)!} \text{ if and only if } \underline{x} \text{ is a permutation of } \underline{x}_0\text{)} \\
=\ & 2\sum_{\underline{x}_0} p(\underline{x}_0)\sum_{g\in G[\underline{x}_0]}\Pr_q\left[\left|\nu'_g - \nu''_g\right| > \tfrac{\epsilon}{2}\,\middle|\,\nu'_g + \nu''_g = \tfrac{1}{N}(\text{number of agreements of } g \text{ on } \underline{x}_0)\right] \\
& \text{(apply Lemma 2)} \\
\le\ & 2\sum_{\underline{x}_0} p(\underline{x}_0)\sum_{g\in G[\underline{x}_0]} 2\exp(-\tfrac{1}{8}\epsilon^2 N) \\
=\ & 4\sum_{\underline{x}_0} p(\underline{x}_0)L(\underline{x}_0)\exp(-\tfrac{1}{8}\epsilon^2 N) \\
\le\ & 4\sum_{\underline{x}_0} p(\underline{x}_0)L(2N)\exp(-\tfrac{1}{8}\epsilon^2 N) \\
=\ & 4\cdot L(2N)\exp(-\tfrac{1}{8}\epsilon^2 N)
\end{aligned}$$

■

