Machine Learning Theory (CS 6783)

Lecture 3 (Supplementary Material)

1 Stability of ϕ of Specific Form

Let $\mathcal{F} \subseteq \{\pm 1\}^n$ be a set of n dimensional ± 1 vectors. One form of ϕ we might be interested in for Cover's result is

$$\phi(y_1, \ldots, y_n) = \frac{1}{n} \min_{f \in \mathcal{F}} \sum_{t=1}^{n} \mathbf{1}\{y_t \neq f_t\} + C_n(\mathcal{F})$$

That is, we want our average error to be bounded by average classification error of the best $f \in \mathcal{F}$ (chosen in hindsight of course) + an extra slack of $C_n(\mathcal{F})$. To apply Cover's result from class, we need two things. First, that ϕ is stable and next, to be able to provide guarantee against any arbitrary adversary, we also need that $\mathbb{E}_\epsilon \left[\phi(\epsilon_1, \ldots, \epsilon_n)\right] \geq 1/2$.

Before we focus on these two properties, note that when $\mathcal{F} = \{(+1, \ldots, +1), (-1, \ldots, -1)\}$, then $\frac{1}{n} \min_{f \in \mathcal{F}} \sum_{t=1}^{n} \mathbf{1}\{y_t \neq f_t\}$ is exactly the average error of going with majority in hindsight and we already argued that for this case $C_n(\mathcal{F}) = \Theta(1/\sqrt{n})$ to ensure that $\mathbb{E}_\epsilon \left[\phi(\epsilon_1, \ldots, \epsilon_n)\right] \geq 1/2$.

Moving to the general case, first towards the question of stability we have the following claim.

Claim 1. *For any $\mathcal{F} \subseteq \{\pm 1\}^n$ if we consider any ϕ defined by*

$$\phi(y_1, \ldots, y_n) = \frac{1}{n} \min_{f \in \mathcal{F}} \sum_{t=1}^{n} \mathbf{1}\{y_t \neq f_t\} + C_n(\mathcal{F}),$$

such a ϕ is stable.

Proof. Consider any $i \in [n]$ and any $y_1, \ldots, y_{i-1}, y_{+1}, \ldots, y_n$. Without loss of generality, consider the case when $\phi(y_1, \ldots, y_{i-1}, +1, y_{i+1}, \ldots, y_n) \geq \phi(y_1, \ldots, y_{i-1}, -1, y_{i+1}, \ldots, y_n)$ because when it is not the case we simply consider the same argument below for the terms in the difference swapped.

$$\begin{aligned}
&\phi(y_1, \ldots, y_{i-1}, +1, y_{i+1}, \ldots, y_n) - \phi(y_1, \ldots, y_{i-1}, -1, y_{i+1}, \ldots, y_n) \\
&= \frac{1}{n}\left(\min_{f \in \mathcal{F}}\left\{\sum_{t=1}^{i-1} \mathbf{1}\{y_t \neq f_t\} + \mathbf{1}\{+1 \neq f_i\} + \sum_{i+1}^{n} \mathbf{1}\{y_t \neq f_t\}\right\} - \min_{f \in \mathcal{F}}\left\{\sum_{t=1}^{i-1} \mathbf{1}\{y_t \neq f_t\} + \mathbf{1}\{-1 \neq f_i\} + \sum_{i+1}^{n} \mathbf{1}\{y_t \neq f_t\}\right\}\right)
\end{aligned}$$

Noting that $-\min(\ldots)$ is same as $\max(-\ldots)$, and rearranging we get,

$$= \frac{1}{n} \max_{f' \in \mathcal{F}}\left(\min_{f \in \mathcal{F}}\left\{\sum_{t=1}^{i-1} \mathbf{1}\{y_t \neq f_t\} + \mathbf{1}\{+1 \neq f_i\} + \sum_{i+1}^{n} \mathbf{1}\{y_t \neq f_t\}\right\} - \sum_{t=1}^{i-1} \mathbf{1}\{y_t \neq f'_t\} - \mathbf{1}\{-1 \neq f'_i\} - \sum_{i+1}^{n} \mathbf{1}\{y_t \neq f'_t\}\right)$$

replacing the minimum with f' as the choice to use we get,

$$\leq \frac{1}{n}\max_{f'\in\mathcal{F}}\left(\sum_{t=1}^{i-1}\mathbf{1}\{y_t\neq f'_t\}+\mathbf{1}\{+1\neq f'_i\}+\sum_{i+1}^{n}\mathbf{1}\{y_t\neq f'_t\}-\sum_{t=1}^{i-1}\mathbf{1}\{y_t\neq f'_t\}-\mathbf{1}\{-1\neq f'_i\}-\sum_{i+1}^{n}\mathbf{1}\{y_t\neq f'_t\}\right)$$
$$= \frac{1}{n}\max_{f'\in\mathcal{F}}\{\mathbf{1}\{+1\neq f'_i\}-\mathbf{1}\{-1\neq f'_i\}\}$$
$$\leq \frac{1}{n}$$

Hence, ϕ is stable for any $\mathcal{F}$. □

Now that we have established that ϕ is stable, lets move to what the minimum value of $C_n(\mathcal{F})$ should be to ensure that $\mathbb{E}_\epsilon\left[\phi(\epsilon)\right]\geq 1/2$. To this end, clearly the smallest $C_n(\mathcal{F})$ is the one where we have the equality

$$\begin{aligned}\frac{1}{2} &= \mathbb{E}_\epsilon\left[\phi(\epsilon)\right]\\ &= \frac{1}{n}\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\sum_{t=1}^{n}\mathbf{1}\{\epsilon_t\neq f_t\}\right]+C_n(\mathcal{F})\end{aligned}$$

and so we have that

$$C_n(\mathcal{F}) = \frac{1}{2}-\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}\{\epsilon_t\neq f_t\}\right]$$

using that fact that for any $y, y'\in\{\pm1\}$, $\mathbf{1}\{y\neq y'\}=(1-y.y')/2$ we get,

$$\begin{aligned}&= \frac{1}{2}-\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{2}\left(1-\epsilon_t\cdot f_t\right)\right]\\ &= \frac{1}{2}-\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2}-\frac{1}{2}\epsilon_t\cdot f_t\right)\right]\\ &= \frac{1}{2}-\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\left(\frac{1}{2}-\frac{1}{2n}\sum_{t=1}^{n}\epsilon_t\cdot f_t\right)\right]\\ &= -\mathbb{E}_\epsilon\left[\min_{f\in\mathcal{F}}\left(-\frac{1}{2n}\sum_{t=1}^{n}\epsilon_t\cdot f_t\right)\right]\\ &= \frac{1}{2n}\mathbb{E}_\epsilon\left[\max_{f\in\mathcal{F}}\sum_{t=1}^{n}\epsilon_t\cdot f_t\right]\\ &= \frac{1}{2n}\mathbb{E}_\epsilon\left[\max_{f\in\mathcal{F}}f^\top\epsilon\right]\end{aligned}$$

where in the last line we can thing of f and ϵ as n dimensional vectors.