Indian Institute of Technology Kanpur
CS777 Topics in Learning Theory

QUESTION 1

Assignment Number: 2
Student Name: Gurpreet Singh
Roll Number: 150259
Date: March 13, 2018

From the discussion in class, we know about the following inequality, which I have used to prove some identities in this question.

If $a, b \geq 0$ and $a + b \geq c$, then for all $\eta \in [\,0, 1\,]$,

$$\eta \cdot a + (1 - \eta) \cdot b \quad \geq \quad c \cdot \min\,(\eta, 1 - \eta) \tag{1}$$

PART 1

$$\begin{aligned}
L^{0-1}(\eta) &= \min_{\hat{y}} \underset{Y \sim \eta}{\mathbb{E}} \left[\, l^{0-1}(\hat{y}, Y)\,\right] \\
&= \min_{\hat{y}} \underset{Y \sim \eta}{\mathbb{E}} \left[\, \mathbb{I}\,[\,\hat{y} \neq Y\,]\,\right] \\
&= \min_{\hat{y}} \underset{Y \sim \eta}{\mathbb{E}} \left[\, \mathbb{I}\,[\,\hat{y} = 1\,]\,\mathbb{I}\,[\,Y = -1\,] + \mathbb{I}\,[\,\hat{y} = -1\,]\,\mathbb{I}\,[\,Y = 1\,]\,\right] \\
&= \min_{\hat{y}}\ (1 - \eta)\ \mathbb{I}\,[\,\hat{y} = 1\,] + \eta\ \mathbb{I}\,[\,\hat{y} = -1\,] \\
&\geq \min\,(\eta, 1 - \eta)
\end{aligned}$$

The last inequality is from the discussion in class. However, for an algorithm that predicts $\hat{y}$ as follows,

$$\hat{y} \quad = \quad \begin{cases} +1 & \text{if } \eta > 0.5 \\ -1 & \text{else} \end{cases}$$

the equality is satisfied, that is $L^{0-1}(\eta) = \min\,(\eta, 1 - \eta)$.

Similarly for $L^{\sigma}(\eta)$,

$$\begin{aligned}
L^{\sigma}(\eta) &= \min_{\hat{y}} \underset{Y \sim \eta}{\mathbb{E}} \left[\, l^{\sigma}(\hat{y}, Y)\,\right] \\
&= \min_{\hat{y}} \underset{Y \sim \eta}{\mathbb{E}} \left[\frac{1}{1 + \exp\,(Y\hat{y})} \right] \\
&= \min_{\hat{y}}\ (1 - \eta)\frac{\exp\,(\hat{y})}{1 + \exp\,(\hat{y})} + \eta \frac{1}{1 + \exp\,(\hat{y})}
\end{aligned}$$

Again, from the discussion in class, we can say

$$L^{\sigma}(\eta) \quad \geq \quad \min\,(\eta, 1 - \eta)$$

We can find a prediction function such that this is an equality. The said prediction would be as follows

$$\hat{y} \quad = \quad \begin{cases} +\infty & \text{if } \eta > 0.5 \\ -\infty & \text{else} \end{cases}$$

Since for this predictor, the said inequality is indeed an equality, we can say that the minimum predictor gives $L^{\sigma}(\eta) = \min\,(\eta, 1 - \eta)$.

Hence, we have the result

$$L^{0-1}(\eta) = L^{\sigma}(\eta) = \min\,(\eta, 1 - \eta) \tag{2}$$

PART 2

From the previous part, we have

$$\begin{aligned} L^{0-1}(\hat{y},\eta) &= (1-\eta)\,\mathbb{I}[\,\hat{y}=1\,] + \eta\,\mathbb{I}[\,\hat{y}=-1\,] \\ L^{0-1}(\eta) &= \min(\eta, 1-\eta) \end{aligned}$$

Therefore, there are four cases for $L^{0-1}(\hat{y},\eta) - L^{0-1}(\eta)$, namely

$$L^{0-1}(\hat{y},\eta) - L^{0-1}(\eta) = \begin{cases} 1-2\eta & \text{if } \eta \leq 0.5, \hat{y} = +1 \\ 0 & \text{if } \eta \leq 0.5, \hat{y} = -1 \\ 0 & \text{if } \eta > 0.5, \hat{y} = +1 \\ 2\eta - 1 & \text{if } \eta > 0.5, \hat{y} = -1 \end{cases}$$

From this, we can directly represent this in the form given in the question, *i.e.*

$$L^{0-1}(\hat{y},\eta) - L^{0-1}(\eta) = |\,2\eta - 1\,| \cdot \mathbb{I}[\,\hat{y}\,(2\eta-1) < 0\,]$$

PART 3

From Part 1, we know

$$\begin{aligned} L^{0-1}(\text{sign}(\hat{y}),\eta) &= (1-\eta)\,\mathbb{I}[\,\text{sign}(\hat{y}) = +1\,] + \eta\,\mathbb{I}[\,\text{sign}(\hat{y}) = -1\,] \\ &= (1-\eta)\,\mathbb{I}[\,\hat{y} \geq 0\,] + \eta\,\mathbb{I}[\,\hat{y} < 0\,] \end{aligned}$$

Now, consider the term $2L^{\sigma}(\hat{y},\eta) - L^{0-1}(\text{sign}(\hat{y}),\eta)$.

$$2L^{\sigma}(\hat{y},\eta) - L^{0-1}(\text{sign}(\hat{y}),\eta) = (1-\eta)\left[\frac{2\exp(\hat{y})}{1+\exp(\hat{y})} - \mathbb{I}[\,\hat{y}\geq 0\,]\right] + \eta\left[\frac{2}{1+\exp(\hat{y})} - \mathbb{I}[\,\hat{y}<0\,]\right]$$

Since for $x \geq 0$, $\frac{2\exp(x)}{1+\exp(x)} \geq 1$, we have $\frac{2\exp(\hat{y})}{1+\exp(\hat{y})} - \mathbb{I}[\,\hat{y}\geq 0\,] \geq 0$. Similary, if $x < 0$, $\frac{2}{1+\exp(x)} > 1$. Therefore, we also have $\frac{2}{1+\exp(\hat{y})} - \mathbb{I}[\,\hat{y}<0\,] \geq 0$.

Hence, from Equation 1, we can say

$$\begin{aligned} 2L^{\sigma}(\hat{y},\eta) - L^{0-1}(\text{sign}(\hat{y}),\eta) &\geq \left[\frac{2\exp(\hat{y})+2}{1+\exp(\hat{y})} - \mathbb{I}[\,\hat{y}\geq 0\,] - \mathbb{I}[\,\hat{y}<0\,]\right] \cdot \min(\eta, 1-\eta) \\ &= \min(\eta, 1-\eta) \\ &= 2L^{\sigma}(\eta) - L^{0-1}(\eta) \end{aligned}$$

The last equality comes from Equation 2. Therefore, we have the desired result, that is

$$L^{0-1}(\text{sign}(\hat{y}),\eta) - L^{0-1}(\eta) \leq 2\,(L^{\sigma}(\hat{y},\eta) - L^{\sigma}(\eta)) \tag{3}$$

PART 4

From part 3, equation 3, we have the following result

$$L^{0-1}(\text{sign}(\hat{y}),\eta) - L^{0-1}(\eta) \leq 2\,(L^{\sigma}(\hat{y},\eta) - L^{\sigma}(\eta))$$

Suppose for some $f : \mathcal{X} \to \mathbb{R}$, we have $\hat{y} = f(\mathbf{x})$. Therefore, we can write

$$L^{0-1}(\text{sign}(f(\mathbf{x})),\eta) - L^{0-1}(\eta) \leq 2\,(L^{\sigma}(f(\mathbf{x}),\eta) - L^{\sigma}(\eta))$$

Since this inequality exists for all $\mathbf{x} \in \mathcal{X}$, we can take expectation on both sides over $\mathbf{X} \sim \mathcal{D}$ without disturbing the inequality. Therefore

$$\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{0-1}(\text{sign}(f(\mathbf{X})),\eta) - L^{0-1}(\eta)\right] \quad \leq \quad \underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[2\left(L^{\sigma}(f(\mathbf{X}),\eta) - L^{\sigma}(\eta)\right)\right]$$

Since both LHS and RHS are regret terms, we can replace them as follows,

$$\implies \mathcal{R}_{\mathcal{D}}^{0-1}[\text{sign} \circ f] \quad \leq \quad 2\mathcal{R}_{\mathcal{D}}^{\sigma}[f] \tag{4}$$

Hence, we have shown the inequality required.

PART 5

We have

$$l_{\alpha}^{\sigma} \quad = \quad \frac{1}{1 + \exp(\alpha \cdot \hat{y}y)} \qquad\qquad l^{\sigma} \quad = \quad \frac{1}{1 + \exp(\hat{y}y)}$$

Therefore, we can write $l_{\alpha}^{\sigma}(\hat{y}, y)$ as

$$l_{\alpha}^{\sigma}(\hat{y}, y) \quad = \quad l^{\sigma}(\alpha \cdot \hat{y}, y)$$

Therefore, we have

$$L_{\alpha}^{\sigma}(\hat{y}, \eta) \quad = \quad L^{\sigma}(\alpha \cdot \hat{y}, \eta)$$

Now, borrowing the results from part 1, we can write

$$L_{\alpha}^{\sigma}(\hat{y}, \eta) \quad = \quad L^{\sigma}(\alpha \cdot \hat{y}, \eta) \quad \geq \quad \min\{\eta, 1-\eta\}$$

Note. We can write this as the RHS is independent of the value of $\mathbf{y}$ and therefore always holds

We can use the same predictor as we used in part 1 to show that there is in fact an optimal predictor which gives the lowest pointwise error ($\min\{\eta, 1-\eta\}$). The optimal predictor being

$$\hat{y} \quad = \quad \begin{cases} +\infty & \text{if } \eta > 0.5 \\ -\infty & \text{else} \end{cases}$$

Hence, we have

$$L_{\alpha}^{\sigma}(\eta) \quad = \quad \min\{\eta, 1-\eta\} \quad = \quad L^{\sigma}(\eta)$$

From equation 4, we have

$$\begin{aligned} \underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{0-1}(\text{sign}(f(\mathbf{X})),\eta) - L^{0-1}(\eta)\right] \quad &\leq \quad 2\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{\sigma}(f(\mathbf{X}),\eta) - L^{\sigma}(\eta)\right] \\ \implies \underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{0-1}(\text{sign}(f(\mathbf{X})),\eta) - L^{0-1}(\eta)\right] \quad &\leq \quad 2\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{\sigma}(f(\mathbf{X}),\eta) - L_{\alpha}^{\sigma}(\eta)\right] \end{aligned}$$

Suppose we have some $g : \mathcal{X} \to \infty$ such that $f(\mathbf{x}) = \alpha \cdot g(\mathbf{x})$. Clearly, such a g will exist for all f. Therefore, we can replace f by $\alpha \cdot g$ as follows,

$$\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{0-1}(\text{sign}(\alpha \cdot g(\mathbf{X})),\eta) - L^{0-1}(\eta)\right] \quad \leq \quad 2\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{\sigma}(\alpha \cdot g(\mathbf{X}),\eta) - L_{\alpha}^{\sigma}(\eta)\right]$$

Also, since $\alpha > 0$, we have $\text{sign}\alpha \cdot g(\mathbf{x}) = \text{sign} g(\mathbf{x})$, and $L^{\sigma}(\alpha \cdot g(\mathbf{x}), \eta) = L_{\alpha}^{\sigma}(g(\mathbf{x}), \eta)$ for all $\mathbf{x} \in \mathcal{X}$. Hence, we can write this as

$$\begin{aligned} \underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L^{0-1}(\text{sign}(g(\mathbf{X})),\eta) - L^{0-1}(\eta)\right] \quad &\leq \quad 2\underset{\mathbf{X}\sim\mathcal{D}}{\mathbb{E}}\left[L_{\alpha}^{\sigma}(g(\mathbf{X}),\eta) - L_{\alpha}^{\sigma}(\eta)\right] \\ \implies \mathcal{R}_{\mathcal{D}}^{0-1}[\text{sign} \circ g] \quad &\leq \quad 2\mathcal{R}_{\mathcal{D}}^{\sigma,\alpha}[g] \end{aligned}$$

Therefore, we have the exact same regret transfer bound for l_{α}^{σ} as for l^{σ} with respect to l^{0-1}.