Machine Learning

Answer Sheet for Homework 7

Da-Min HUANG

R04942045

Graduate Institute of Communication Engineering, National Taiwan University

January 6, 2016

Problem 1

Set $\mu_- = 1 - \mu_+$, we have

$$1 - \mu_+^2 - \mu_-^2 = 1 - \mu_+^2 - (1-\mu_+)^2 = (1-\mu_+)(1+\mu_+) - (1-\mu_+)^2 \tag{1}$$

$$= 2\mu_+(1-\mu_+) = -2\mu_+^2 + 2\mu_+ = -2\left(\mu_+ - \frac{1}{2}\right)^2 + \frac{1}{2} \tag{2}$$

$$\leq \frac{1}{2} \tag{3}$$

Hence, if $\mu_+ = 1/2 \in [0,1]$, then the maximum value of Gini index is $\boxed{1/2}$.

□

Problem 2

The normalized Gini index is

$$\frac{\left(1 - \mu_+^2 - \mu_-^2\right)}{\left(\frac{1}{2}\right)} = 2\left(1 - \mu_+^2 - \mu_-^2\right) \tag{4}$$

The squared error can be rewritten as

$$\mu_+\left(1 - (\mu_+ - \mu_-)\right)^2 + \mu_-\left(-1 - (\mu_+ - \mu_-)\right)^2 = 4\mu_+(1-\mu_+)^2 + 4\mu_+^2(1-\mu_+) \tag{5}$$

$$= 4\mu_+(1-\mu_+) \leq 4 \times \frac{1}{4} = 1 \tag{6}$$

Hence the normalized squared error is

$$
\begin{aligned}
4\mu_+ \left(1 - \mu_+\right) = 2\left(2\mu_+ \left(1 - \mu_+\right)\right) &= 2\left(\left(1 - \mu_+\right)\left(1 + \mu_+\right) - \left(1 - \mu_+\right)^2\right) && (7) \\
&= 2\left(1 - \mu_+^2 - \mu_-^2\right) && (8)
\end{aligned}
$$

which is equal to normalized Gini index.

□

Problem 3

The probability of one example not sampled is

$$\left(1 - \frac{1}{N}\right)^{pN} = \frac{1}{\left(\frac{N}{N-1}\right)^{pN}} = \frac{1}{\left(1 + \frac{1}{N-1}\right)^{pN}} = \left(\frac{1}{\left(1 + \frac{1}{N-1}\right)^{N}}\right)^{p} \tag{9}$$

As $N \to \infty$, we have

$$\lim_{N\to\infty}\left(\frac{1}{\left(1 + \frac{1}{N-1}\right)^{N}}\right)^{p} = \left(\lim_{N\to\infty}\frac{1}{\left(1 + \frac{1}{N-1}\right)^{N}}\right)^{p} = \left(\frac{1}{e}\right)^{p} = e^{-p} \tag{10}$$

So there approximately $e^{-p} \cdot N$ of the examples not sampled.

□

Problem 4

Since $G = \text{Uniform}\left(\{g_k\}_{k=1}^{3}\right)$, so if at least two terms of $\{g_k\}_{k=1}^{3}$ ouput wrong result, then G outputs wrong result. Let $\{E_k\}_{k=1}^{3}$ be the set of examples that $\{g_k\}_{k=1}^{K}$ got wrong results. Apparently $|E_3| > |E_2| > |E_1|$ and $|E_1| + |E_2| > |E_3|$. So

1. Maximum of $E_{\text{out}}(G)$ happens at $E_3 \subset (E_1 \cup E_2)$. Then G outputs wrong result in the region of E_3 with $E_{\text{out}}(G) = 0.35$.

2. Minimum of $E_{\text{out}}(G)$ happens at $E_i \cap E_j = \phi$, $i \neq j$ and $1 \leq i, j \leq 3$ with $i, j \in \mathbb{N}$. Then G always ouputs the correct result since $(E_1 \cup E_2 \cup E_3) \subset \{\text{all examples}\}$.

Hence, $0 \leq E_{\text{out}}(G) \leq 0.35$.

□

Problem 5

Since $G = \text{Uniform}\left(\{g_k\}_{k=1}^K\right)$, so if at least $(K+1)/2$ terms of $\{g_k\}_{k=1}^K$ ouput wrong result, then G outputs wrong result. Let $\{E_k\}_{k=1}^K$ be the set of examples that $\{g_k\}_{k=1}^K$ got wrong results.

If G outputs wrong result on some example $\mathbf{x}$, then we have

$$\mathbf{x} \in \bigcap_{i=1}^{((K+1)/2)+m} E_{\alpha_i} \tag{11}$$

where α_i is some index satisfies $1 \leq \alpha_i \leq K$ and $m \in (\mathbb{N} \cup \{0\})$ with $0 \leq m < (K+1)/2$. And

$$\left| \bigcap_{i=1}^{((K+1)/2)+m} E_{\alpha_i} \right| \leq \frac{2}{K+1+2m} \sum_{k=1}^{K} e_k \leq \frac{2}{K+1} \sum_{k=1}^{K} e_k \tag{12}$$

(12) holds due to

$$\bigcap_{i=1}^{((K+1)/2)+m} E_{\alpha_i} \subseteq E_\beta \tag{13}$$

where β is some index such that $|E_\beta| = \min_{\alpha_i} |E_{\alpha_i}|$. So

$$\left(\frac{K+1}{2} + m\right) \left| \bigcap_{i=1}^{((K+1)/2)+m} E_{\alpha_i} \right| \leq \left(\frac{K+1}{2} + m\right) |E_\beta| \leq \sum_{k=1}^{K} |E_k| \tag{14}$$

(14) holds since size of E_β is the samllest among $(((K+1)/2)+m)$ terms and $\sum_{k=1}^K |E_k|$ must contains the $(((K+1)/2)+m)$ terms.

Hence, we have

$$E_{\text{out}}(G) \leq \frac{2}{K+1} \sum_{k=1}^{K} e_k \tag{15}$$

$\square$

Problem 6

By the definition of U_t, we have

$$
\begin{aligned}
U_{t+1} &= \frac{1}{N} \sum_{n=1}^{N} \exp\left(-y_n \sum_{\tau=1}^{t} \alpha_\tau g_\tau\left(\mathbf{x}_n\right)\right) && (16)\\
&= \frac{1}{N} \sum_{n=1}^{N} \exp\left(-y_n \sum_{\tau=1}^{t-1} \alpha_\tau g_\tau\left(\mathbf{x}_n\right) - y_n \alpha_t g_t\left(\mathbf{x}_n\right)\right) && (17)\\
&= \frac{1}{N} \sum_{n=1}^{N} \exp\left(-y_n \sum_{\tau=1}^{t-1} \alpha_\tau g_\tau\left(\mathbf{x}_n\right)\right) \exp\left(-y_n \alpha_t g_t\left(\mathbf{x}_n\right)\right) && (18)\\
&= \sum_{n=1}^{N} u_n^{(t)} \exp\left(-y_n \alpha_t g_t\left(\mathbf{x}_n\right)\right) && (19)\\
&= \sum_{\substack{n \\ y_n \neq g_t(\mathbf{x}_n)}} u_n^{(t)} \exp\left(-y_n \alpha_t g_t\left(\mathbf{x}_n\right)\right) + \sum_{\substack{n \\ y_n = g_t(\mathbf{x}_n)}} u_n^{(t)} \exp\left(-y_n \alpha_t g_t\left(\mathbf{x}_n\right)\right) && (20)\\
&= \sum_{\substack{n \\ y_n \neq g_t(\mathbf{x}_n)}} u_n^{(t)} \exp\left(\alpha_t\right) + \sum_{\substack{n \\ y_n = g_t(\mathbf{x}_n)}} u_n^{(t)} \exp\left(-\alpha_t\right) && (21)\\
&= \exp\left(\alpha_t\right)\left(\epsilon_t\right) \sum_{n=1}^{N} u_n^{(t)} + \exp\left(-\alpha_t\right)\left(1-\epsilon_t\right) \sum_{n=1}^{N} u_n^{(t)} && (22)\\
&= U_t\left(\exp\left(\alpha_t\right)\left(\epsilon_t\right) + \exp\left(-\alpha_t\right)\left(1-\epsilon_t\right)\right) = U_t \cdot 2\sqrt{\epsilon_t\left(1-\epsilon_t\right)} && (23)
\end{aligned}
$$

Since

$$
U_1 = \sum_{n=1}^{N} u_n^{(1)} = \sum_{n=1}^{N} \frac{1}{N} = 1 \tag{24}
$$

we have

$$
\begin{aligned}
U_3 &= U_2 \cdot 2\sqrt{\epsilon_2\left(1-\epsilon_2\right)} = \left(U_1 \cdot 2\sqrt{\epsilon_1\left(1-\epsilon_1\right)}\right) \cdot 2\sqrt{\epsilon_2\left(1-\epsilon_2\right)} && (25)\\
&= 4\sqrt{\epsilon_1 \epsilon_2\left(1-\epsilon_1\right)\left(1-\epsilon_2\right)} && (26)
\end{aligned}
$$

which can be generalized as

$$
U_{T+1} = \prod_{t=1}^{T}\left(2\sqrt{\epsilon_t\left(1-\epsilon_t\right)}\right) \tag{27}
$$

□

Problem 7

To compute s_n, we need to find the optimal η of

$$\min_{\eta} \frac{1}{N} \sum_{n=1}^{N} \left((y_n - s_n) - \eta g_t \left(\mathbf{x}_n \right) \right)^2 := A \tag{28}$$

From $\partial A / \partial \eta = 0$, we have

$$\eta = \frac{\sum_{n=1}^{N} g_t \left(\mathbf{x}_n \right) \left(y_n - s_n \right)}{\sum_{n=1}^{N} g_t^2 \left(\mathbf{x}_n \right)} \tag{29}$$

Now $s_n = 0$ and $g_1 \left(\mathbf{x} \right) = 2$, so

$$\eta = \frac{2 \sum_{n=1}^{N} y_n}{4 \sum_{n=1}^{N}} = \frac{1}{2N} \sum_{n=1}^{N} y_n \tag{30}$$

Since $\eta = \alpha_1$, so

$$\alpha_1 g_1 \left(\mathbf{x}_n \right) = \frac{2}{2N} \sum_{n=1}^{N} y_n = \frac{1}{N} \sum_{n=1}^{N} y_n = s_n \tag{31}$$

□

Problem 8

From the equatio of optimal η, we have

$$\eta = \frac{\sum_{n=1}^{N} g_t \left(\mathbf{x}_n \right) \left(y_n - s'_n \right)}{\sum_{n=1}^{N} g_t^2 \left(\mathbf{x}_n \right)} = \frac{\sum_{n=1}^{N} y_n g_t \left(\mathbf{x}_n \right) - \sum_{n=1}^{N} s'_n g_t \left(\mathbf{x}_n \right)}{\sum_{n=1}^{N} g_t^2 \left(\mathbf{x}_n \right)} = \alpha_t \tag{32}$$

so

$$\sum_{n=1}^{N} y_n g_t \left(\mathbf{x}_n \right) - \sum_{n=1}^{N} s'_n g_t \left(\mathbf{x}_n \right) = \alpha_t \sum_{n=1}^{N} g_t^2 \left(\mathbf{x}_n \right) = \sum_{n=1}^{N} \alpha_t g_t^2 \left(\mathbf{x}_n \right) = \sum_{n=1}^{N} \left(s_n - s'_n \right) g_t \left(\mathbf{x}_n \right) \tag{33}$$

where s'_n is defined as the s_n in iteration $(t-1)$, so

$$\sum_{n=1}^{N} s_n g_t \left(\mathbf{x}_n \right) = \sum_{n=1}^{N} y_n g_t \left(\mathbf{x}_n \right) \tag{34}$$

□

Problem 9

□

Problem 10

□

Problem 11

□

Problem 12

□

Problem 13

□

Problem 14

□

Problem 15

□

Problem 16

□

Problem 17

□

Problem 18

□

Problem 19

□

Problem 20

□

Reference

[1] Lecture Notes by Hsuan-Tien LIN, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.