Machine Learning Homework 9

He Li, 2014012684

2017/11/30

Problem 1. *For the Randomized Weighted Majority Vote Algorithm, define expected loss:*

$$L_T = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{w_{t,i}}{\sum_{j=1}^{N} w_{t,j}} |\tilde{y}_{t,i} - y_t|$$

Then $\forall \beta \in (\frac{1}{2}, 1)$ *we have*

$$L_T \leq (2-\beta) m_T^* + \frac{\ln N}{1-\beta} \tag{1}$$

Proof. Define $W_t = \sum_{j=1}^{N} w_{t,j}$. Therefore, $W_1 = N$ and

$$L_T = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{w_{t,i}}{W_t} |\tilde{y}_{t,i} - y_t|$$

Denote

$$l_t = \frac{\sum_{\tilde{y}_{t,i} \neq y_t} w_{t,i}}{W_t}$$

Therefore,

$$\begin{aligned} W_{t+1} &= (1 - l_t) W_t + \beta l_t W_t \\ &= W_t \left(1 - l_t + \beta l_t\right) \end{aligned}$$

And we have

$$\begin{aligned} W_{\text{final}} &= W_1 \prod_{t=1}^{t} (1 - (1-\beta) l_t) \\ &\leq N \prod_{t=1}^{T} \exp\left\{-(1-\beta) l_t\right\} \\ &= N \exp\left\{-(1-\beta) \sum_{t=1}^{T} l_t\right\} \end{aligned}$$

Note that $\forall i$,

$$W_{\text{final}} \geq w_{T,i} = \beta^{m_T^{(i)}}$$

Therefore, $W_{\text{final}} \geq \beta^{m_T^*}$

$$N \exp\left\{-(1-\beta) \sum_{t=1}^{T} l_t\right\} \geq \beta^{m_T^*}$$

Therefore,

$$\sum_{t=1}^{T} l_t \leq \frac{\ln \frac{1}{\beta}}{1-\beta} m_T^* + \frac{\ln N}{1-\beta}$$

Note that $L_T = \sum_{t=1}^{T} l_t$ and when $\beta \in (\frac{1}{2}, 1)$

$$\frac{\ln \frac{1}{\beta}}{1-\beta} \leq 2 - \beta$$

So we have the conclusion

$$L_T \le (2-\beta)m_T^* + \frac{\ln N}{1-\beta}$$

$\square$

Problem 2. *For every choice of $f_1, f_2, \ldots$ the Multiplicative Weight Updating Algorithm goes into these two situations for at most $\frac{2\log N}{\epsilon\delta}$ $(0 < \epsilon < \delta)$*

Proof. Use Kullback–Leibler divergence $D(x\|x_t) = \sum_{i=1}^N x(i)\log\frac{x(i)}{x_t(i)}$ as potential function. Therefore, when $\langle f_t, x\rangle - \langle f_t, x_t\rangle > \delta$

$$\begin{aligned}
D(x\|x_{t+1}) &= \sum_{i=1}^N x(i)\log\frac{x(i)}{x_{t+1}(i)} \\
&= D(x\|x_t) + \sum_{i=1}^N x(i)\log\frac{x_t(i)}{x_{t+1}(i)} \\
&= D(x\|x_t) + \sum_{f_t(i)=1} x(i)\log\frac{x_t(i)}{\frac{(1+\epsilon)x_t(i)}{1+\epsilon\langle f_t, x_t\rangle}} \\
&= D(x\|x_t) + \log(1+\epsilon\langle f_t, x_t\rangle) - \langle f_t, x\rangle\log(1+\epsilon) \\
&\le D(x\|x_t) + \log(1+\epsilon\langle f_t, x_t\rangle) - (\langle f_t, x_t\rangle + \delta)\log(1+\epsilon) \\
&\le D(x\|x_t) - \delta\log(1+\epsilon)
\end{aligned}$$

Given the ground truth that

$$(1+\epsilon)^k \ge 1+\epsilon k$$

Also, when $\langle f_t, x\rangle - \langle f_t, x_t\rangle < -\delta$,

$$\begin{aligned}
D(x\|x_{t+1}) &= \sum_{i=1}^N x(i)\log\frac{x(i)}{x_{t+1}(i)} \\
&= D(x\|x_t) + \log(1+\epsilon\langle f_t, x_t\rangle) - \langle f_t, x\rangle\log(1+\epsilon) \\
&\le D(x\|x_t) + \log\left[1+\epsilon(\langle f_t, x\rangle + \delta)\right] - \langle f_t, x\rangle\log(1+\epsilon) \\
&\le D(x\|x_t) - \delta\log(1+\epsilon)
\end{aligned}$$

Therefore, denote n_t as the update times till time t, which is the times that $|\langle f_t, x\rangle - \langle f_t, x_t\rangle| > \delta$ till time t. Then,

$$\begin{aligned}
0 &\le D(x\|x_T) \\
&\le D(x\|x_1) - n_T\delta\log(1+\epsilon) \\
&\le \log N - n_T\delta\log(1+\epsilon)
\end{aligned}$$

Therefore,

$$n_T \le \frac{\log N}{\delta\log(1+\epsilon)} \le \frac{2\log N}{\epsilon\delta}$$

Given the fact that $0 < \epsilon < \delta < 1$ and $\log(1+\epsilon) \ge \frac{\epsilon}{2}$ when $\epsilon \in (0,1)$. $\square$