Machine Learning Homework 5

He Li, 2014012684

2017/11/2

Problem 1. *AdaBoost is a greedy exponential loss with the following two properties, prove them.*

$$\alpha_t = \arg\min_{\alpha} \sum_{i=1}^{n} D_t(i) \exp\{-y_i \alpha_t h_t(x_i)\} \tag{1}$$

$$\prod_{t=1}^{T} z_t = \frac{1}{n} \sum_{i=1}^{n} \exp\left\{-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right\} = \frac{1}{n} \sum_{i} \exp\left\{-y_i f(x_i)\right\} \tag{2}$$

Proof. We first prove (1). Denote $F_t(\alpha) = \sum_{i=1}^{n} D_t(i) \exp\{-y_i \alpha_t h_t(x_i)\}$, when $F_t(\alpha)$ reaches its minimum,

$$\frac{\partial F_t(\alpha)}{\partial \alpha} = \sum_{i=1}^{n} -y_i h_t(x_i) D_t(i) \exp\{-y_i \alpha_t h_t(x_i)\} = 0$$

$$\Longrightarrow \sum_{i=1}^{n} I[y_i \neq h_t(x_i)] D_t(i) \exp\{I[y_i \neq h_t(x_i)]\alpha_t\} = \sum_{i=1}^{n} I[y_i = h_t(x_i)] D_t(i) \exp\{-I[y_i = h_t(x_i)]\alpha_t\}$$

Given the fact that

$$\sum_{i=1}^{n} D_t(i) I[y_i \neq h_t(x_i)] + \sum_{i=1}^{n} D_t(i) I[y_i = h_t(x_i)] = \sum_{i=1}^{n} D_t(i) = 1$$

We have

$$\epsilon_t \exp\{\alpha_t\} = (1 - \epsilon_t) \exp\{-\alpha_t\}$$

Therefore,

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} = \frac{1}{2} \ln \frac{1 + \gamma_t}{1 - \gamma_t}$$

We then prove (2). Given that

$$z_t := \sum_{i=1}^{n} D_t(i) \exp\left\{-y_i \alpha_t h_t(x_i)\right\}$$

$$D_t(i) := \frac{D_{t-1}(i) \exp\left\{-y_i \alpha_{t-1} h_{t-1}(x_i)\right\}}{z_{t-1}}$$

Obviously,

$$\begin{aligned}
z_T &= \sum_{i=1}^{n} D_T(i) \exp\left\{-y_i \alpha_T h_T(x_i)\right\} \\
&= \sum_{i=1}^{n} \frac{D_{T-1}(i) \exp\left\{-y_i \alpha_{T-1} h_{T-1}(x_i)\right\}}{z_{T-1}} \exp\left\{-y_i \alpha_T h_T(x_i)\right\} \\
&\ldots\ldots \\
&= \sum_{i=1}^{n} D_1(i) \frac{\exp\left\{-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right\}}{\prod_{t=1}^{T-1} z_t}
\end{aligned}$$

Therefore, we have our conclusion.

$$\prod_{t=1}^{T} z_t = \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right\} = \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\}$$

□

Problem 2. *Suppose $\gamma_t \geq \gamma \geq 0$ for $t \in [1, ..., T]$. Then*

$$\begin{aligned} P_s\left(yf(x) \leq 0\right) &= \frac{1}{n} I\left[y_i f(x_i) \leq 0\right] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\} \\ &\leq \left(1-\gamma^2\right)^{\frac{T}{2}} \end{aligned} \tag{3}$$

Proof. We first prove that

$$\frac{1}{n} I\left[y_i f(x_i) \leq 0\right] \leq \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\}$$

However, this is quite obvious given the fact that $I\left[y_i f(x_i) \leq 0\right] \leq \exp\left\{-y_i f(x_i)\right\}$ everywhere(surrogate loss of 0-1 loss).

We then prove that

$$\frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\} \leq \left(1-\gamma^2\right)^{\frac{T}{2}}$$

We already know that

$$\frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\} = \prod_{t=1}^{T} z_t$$

Therefore,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-y_i f(x_i)\right\} &= \prod_{t=1}^{T}\left\{\sum_{i=1}^{n} \exp\left\{\alpha_t\right\} D_t(i) I[y_i \neq h_t(x_i) + \exp\left\{-\alpha_t\right\} D_t(i) I[y_i = h_t(x_i)]\right\} \\ &= \prod_{t=1}^{T}\left\{\exp\{\alpha_t\}\epsilon_t + \exp\{-\alpha_t\}(1-\epsilon_t)\right\} \\ &= \prod_{t=1}^{T}\left\{\sqrt{\frac{1+\gamma_t}{1-\gamma_t}}\frac{1-\gamma_t}{2} + \sqrt{\frac{1-\gamma_t}{1+\gamma_t}}\frac{1+\gamma_t}{2}\right\} \\ &= \prod_{t=1}^{T}\sqrt{1-\gamma_t^2} \\ &\leq \left(1-\gamma^2\right)^{\frac{T}{2}} \end{aligned}$$

□

Problem 3. *Calculate the following function*

$$\sum_{i=1}^{n} D_{t+1}(i) I[y_i \neq h_t(x_i)] \tag{4}$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} D_{t+1}(i) I[y_i \neq h_t(x_i)] &= \sum_{i=1}^{n} \frac{D_t(i)\exp\{-y_i\alpha_t h_t(x_i)\}}{\sum_i D_t(i)\exp\{-y_i\alpha_t h_t(x_i)\}} I[y_i \neq h_t(x_i)] \\
&= \frac{\sum_{i=1}^{n} \exp\{\alpha_t\} D_t(i) I[y_i \neq h_t(x_i)]}{\sum_{i=1}^{n} \exp\{\alpha_t\} D_t(i) I[y_i \neq h_t(x_i) + \exp\{-\alpha_t\} D_t(i) I[y_i = h_t(x_i)]} \\
&= \frac{\exp\{\alpha_t\}\epsilon_t}{\exp\{\alpha_t\}\epsilon_t + \exp\{-\alpha_t\}(1-\epsilon_t)}
\end{aligned}$$

Given that

$$\alpha_t = \frac{1}{2}\ln\frac{1+\gamma_t}{1-\gamma_t} = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t}$$

We have

$$\begin{aligned}
\sum_{i=1}^{n} D_{t+1}(i) I[y_i \neq h_t(x_i)] &= \frac{\frac{1-\epsilon_t}{\epsilon_t}\epsilon_t}{\frac{1-\epsilon_t}{\epsilon_t}\epsilon_t + (1-\epsilon_t)} \\
&= \frac{1}{2}
\end{aligned}$$

□