Machine Learning Exercise 6

何舜成

2012011515

1. Show:

$$\frac{\phi(t+1)}{\phi(t)} = Z_{t+1}$$

Proof:

$$\begin{aligned}\frac{\phi(t+1)}{\phi(t)} &= \frac{\frac{1}{n}\sum_{i=1}^{n} e^{-y_i \sum_{\tau=1}^{t+1} \alpha_\tau h_\tau(x_i)}}{\frac{1}{n}\sum_{i=1}^{n} e^{-y_i \sum_{\tau=1}^{t} \alpha_\tau h_\tau(x_i)}} \\ &= \frac{\sum_{i=1}^{n} \prod_{\tau=1}^{t+1} e^{-y_i \alpha_\tau h_\tau(x_i)}}{\sum_{i=1}^{n} \prod_{\tau=1}^{t} e^{-y_i \alpha_\tau h_\tau(x_i)}}\end{aligned}$$

Note that

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

and then

$$e^{-\alpha_t y_i h_t(x_i)} = \frac{Z_t D_{t+1}(i)}{D_t(i)}$$

Therefore we get

$$\begin{aligned}\prod_{\tau=1}^{t} e^{-y_i \alpha_\tau h_\tau(x_i)} &= \prod_{\tau=1}^{t} \frac{Z_t D_{t+1}(i)}{D_t(i)} \\ &= \frac{D_{t+1}(i)}{D_1(i)} \prod_{\tau=1}^{t} Z_t \\ &= n D_{t+1}(i) \prod_{\tau=1}^{t} Z_t\end{aligned}$$

at the last step we use $D_1(i) = \frac{1}{n}$.

$$\begin{aligned}\frac{\phi(t+1)}{\phi(t)} &= \frac{\sum_{i=1}^{n} nD_{t+2}(i)\prod_{\tau=1}^{t+1} Z_\tau}{\sum_{i=1}^{n} nD_{t+1}(i)\prod_{\tau=1}^{t} Z_\tau} \\ &= Z_{t+1}\frac{\sum_{i=1}^{n} D_{t+2}(i)}{\sum_{i=1}^{n} D_{t+1}(i)} \\ &= Z_{t+1}\end{aligned}$$

2. Show

$$Z_t = \sqrt{1-\gamma_t^2}$$

Proof:

The original equation can be transformed in the following steps

$$Z_t = \sqrt{1-\gamma_t^2} \tag{1}$$

$$\sum_{i=1}^{n} D_t(i)e^{-\alpha_t y_i h_t(x_i)} = \sqrt{1-\gamma_t^2} \tag{2}$$

$$\sum_{i=1}^{n} D_t(i)\left(\sqrt{\frac{1+\gamma_t}{1-\gamma_t}}\right)^{-y_i h_t(x_i)} = \sqrt{1-\gamma_t^2} \tag{3}$$

$$\sum_{i=1}^{n} D_t(i)(1+\gamma_t)^{-\frac{y_i h_t(x_i)}{2}}(1-\gamma_t)^{\frac{y_i h_t(x_i)}{2}} = (1+\gamma_t)^{\frac{1}{2}}(1-\gamma_t)^{\frac{1}{2}} \tag{4}$$

$$\sum_{i=1}^{n} D_t(i)(1+\gamma_t)^{-\frac{1+y_i h_t(x_i)}{2}}(1-\gamma_t)^{-\frac{1-y_i h_t(x_i)}{2}} = 1 \tag{5}$$

let $I = \{i|y_i h_t(x_i) = +1\}$, and $A = \sum_{i\in I} D_t(i)$, $B = \sum_{i\notin I} D_t(i)$. Naturally $A + B = 1$. Then we get

$$\gamma_t = \sum_{i=1}^{n} D_t(i)y_i h_t(x_i) = \sum_{i\in I} D_t(i) - \sum_{i\notin I} D_t(i) = A - B$$

and

$$\sum_{i\in I} D_t(i)(1+\gamma_t)^{-1} + \sum_{i\notin I} D_t(i)(1-\gamma_t)^{-1} = 1 \tag{6}$$

therefore

$$
\begin{aligned}
A(1+\gamma_t)^{-1} + B(1-\gamma_t)^{-1} &= 1 && (7)\\
A(1+A-B)^{-1} + B(1-A+B)^{-1} &= 1 && (8)\\
A(1-A+B) + B(1+A-B) &= (1-A+B)(1+A-B) && (9)
\end{aligned}
$$

It is trivial to prove equation (9) using $A+B=1$.

3. Calculate the value of

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

What is the meaning of the result?

Proof:

$$
\begin{aligned}
\sum_{i=1}^{n} D_{t+1}(i) I(y_i \neq h_t(x_i)) &= \sum_{i=1}^{n} \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \frac{1 - y_i h_t(x_i)}{2}\\
&= \frac{1}{2} - \frac{1}{2Z_t} \sum_{i=1}^{n} D_t(i) e^{-\alpha_t y_i h_t(x_i)} y_i h_t(x_i)
\end{aligned}
$$

Next we will show this equation true using the same treatment in the previous problem.

$$
\begin{aligned}
\sum_{i=1}^{n} D_t(i) e^{-\alpha_t y_i h_t(x_i)} y_i h_t(x_i) &= 0\\
\sum_{i=1}^{n} D_t(i) \left(\sqrt{\frac{1-\gamma_t}{1+\gamma_t}}\right)^{y_i h_t(x_i)} y_i h_t(x_i) &= 0\\
\sum_{i \in I} D_t(i) \sqrt{\frac{1-\gamma_t}{1+\gamma_t}} - \sum_{i \notin I} D_t(i) \sqrt{\frac{1-\gamma_t}{1+\gamma_t}} &= 0\\
\sum_{i \in I} D_t(i)(1-\gamma_t) - \sum_{i \notin I} D_t(1+\gamma_t) &= 0\\
A(1-A+B) - B(1-A+B) &= 0
\end{aligned}
$$

It is trivial to prove the last equation using $A+B=1$. Then

$$\sum_{i=1}^{n} D_{t+1}(i) I(y_i \neq h_t(x_i)) = \frac{1}{2}$$

It means that if we use the training data weights in round $t+1$ to test the classifier in round t, the classifier will give the worst prediction, equivalent to random guess.