Sanghee Kim
CSCI 5622
Boosting
13 Mar 2015

Problem 6.3 Update guarantee. Assume that the main weak learner assumption of AdaBoost holds. Let h_t be the base learner selected at round t. Show that the base learner $h_t + 1$ must be different from h_t.

The meaning of the weak learner assumption of AdaBoost is that the algorithm can constantly search classifiers at least somewhat better than coin flipping probability which means usually $\epsilon_t < 1/2$.

To solve this problem, let's assume that $h_t + 1 = h_t$. Then,

$$\begin{aligned}
\epsilon_t + 1 &= \sum_{i=1}^{m} D_{t+1}(i) \mathbb{1}_{yi \neq h_t(x_i)} \\
&= \sum_{i=1}^{m} \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \mathbb{1}_{yi \neq h_t(x_i)}
\end{aligned}$$

For i's that satisfies $y_i \neq h_t(x_i)$, $y_i h_t(x_i)$ equals -1. Based on this,

$$\begin{aligned}
\epsilon_t + 1 &= \sum_{i=1}^{m} \frac{D_t(i) e^{\alpha_t}}{Z_t} \mathbb{1}_{yi \neq h_t(x_i)} \\
&= \frac{e^{\alpha t}}{Z_t} \sum_{i=1}^{m} D_t(i) \mathbb{1}_{yi \neq h_t(x_i)}
\end{aligned}$$

Also, we knew the following equation,

$$\begin{aligned}
Z_t &= 2\sqrt{\epsilon_t (1 - \epsilon_t)} \\
\alpha_t &= \frac{1}{2} \log \frac{1 - \epsilon_t}{\epsilon_t} \\
\sum_{i=1}^{m} D_t(i) \mathbb{1}_{yi \neq h_t(x_i)} &= \epsilon_t
\end{aligned}$$

Let's replace Z_t, α_t, and $\sum_{i=1}^{m} D_t(i) \mathbb{1}_{yi \neq h_t(x_i)}$

$$
\begin{aligned}
\epsilon_t + 1 &= \frac{e^{\frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}\epsilon_t \\
&= \frac{\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}\epsilon_t \\
&= \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \times \frac{\epsilon_t}{2\sqrt{\epsilon_t(1-\epsilon_t)}} \\
&= \frac{1}{2}
\end{aligned}
$$

Here, we can say that if $h_{t+1} = h_t$, $\epsilon_{t+1} = 1/2$. This can perfectly rebut the meaning of the weak learner assumption of AdaBoost. Therefore, the base learner $h_t + 1$ must be different from h_t.

■