Name: Krishna Chaitanya Sripada

Ans 6.3

By weak learning assumption, there exists a hypothesis $h \in H$ where (D_{t+1} - error) < $\frac{1}{2}$. By examining the empirical error of h_t for the distribution D_{t+1}, we get,

$$Z_t = 2\sqrt{\varepsilon_t(1-\varepsilon_t)} \text{ and } \alpha_t = \tfrac{1}{2}\log\tfrac{1-\varepsilon_t}{\varepsilon_t}$$

Therefore,

$$\begin{aligned}
\hat{R}_{D_{t+1}} &= \sum_{i=1}^{m} \frac{D_t(i)e^{-\alpha_t y_i h_t(x_i)}}{Z_t} 1_{y_i h_t(x_i)<0} \\
&= \sum_{y_i h_t(x_i)<0}^{m} \frac{D_t(i)e^{\alpha_t}}{Z_t} \\
&= \frac{e^{\alpha_t}}{Z_t} \sum_{y_i h_t(x_i)<0}^{m} D_t(i) \\
&= \frac{e^{\alpha_t}}{Z_t}\varepsilon_t
\end{aligned}$$

Replacing the above equation with the values stated above, we get,

$$\begin{aligned}
&= \frac{\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}\varepsilon_t \\
&= \frac{1}{2}
\end{aligned}$$

This shows that h_t cannot be selected at round $t+1$.

Ans 6.6

We notice that the base hypotheses in this question can be defined to be threshold functions based on the first or second axis or constant functions. The hypotheses selected by AdaBoost are therefore chosen from the list of first or second axis or constant functions. It can be proved that the hypotheses selected in two consecutive rounds of AdaBoost are distinct. Also h_t and $-h_t$ can't be selected in consecutive rounds since misclassified and classified points by h_t are assigned to the same distribution. Thus, at each round, a distinct hypothesis is chosen. The points at coordinate $(-1,1)$ are misclassified by all the base hypothesis.

The algorithm stops when the best ε_t found is $\frac{1}{2}$. It can be shown that the error of the final classifier returned on the training set is $\frac{1}{4}(1-\varepsilon)$ since it misclassifies exactly the points at coordinate $(1,-1)$.