Robert Crimi
Machine Learning – Boosting
October 16, 2015

1)

a.
$Err(H) \leq \prod Z_t \qquad Z_t = (1 - \epsilon_t)e^{-\propto} + \epsilon_t \, e^{\propto} \qquad \gamma \leq \frac{1}{2} - \epsilon_t \rightarrow \epsilon_t \leq \frac{1}{2} - \gamma$

So,

$$Z_t \leq \left(\frac{1}{2} + \gamma\right) e^{-\propto} + \left(\frac{1}{2} - \gamma\right) e^{\propto}$$

$$\left(\frac{1}{2} + \gamma\right) e^{-\propto} = \left(\frac{1}{2} - \gamma\right) e^{\propto}$$

$$ln\left(\frac{1}{2} + \gamma\right) e^{-\propto} = ln\left(\frac{1}{2} - \gamma\right) e^{\propto}$$

$$ln\left(\frac{1}{2} + \gamma\right) + \ln\left(e^{-\propto}\right) = ln\left(\frac{1}{2} - \gamma\right) + ln(e^{\propto})$$

$$-\propto = ln\left(\frac{1}{2} - \gamma\right) + \propto - ln(\frac{1}{2} + \gamma)$$

$$2 \propto = ln\left(\frac{1}{2} + \gamma\right) - ln(\frac{1}{2} - \gamma)$$

$$\propto = \frac{1}{2} ln\left(\frac{\frac{1}{2} + \gamma}{\frac{1}{2} - \gamma}\right)$$

Optimal $\propto$: $0 < \propto \leq \frac{1}{2} ln\left(\frac{\frac{1}{2} + \gamma}{\frac{1}{2} - \gamma}\right)$

b.

$$P(right) = (1 - \epsilon_t)e^{-\propto} \quad P(wrong) = \epsilon_t\ e^{\propto}$$

So,

$$\frac{P_w}{P_r} = \frac{\epsilon_t\ e^{\propto}}{(1 - \epsilon_t)e^{-\propto}}$$

$$\frac{P_w}{P_r} = \frac{\epsilon_t\ e^{\frac{1}{2}ln\left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)}}{(1 - \epsilon_t)e^{-\frac{1}{2}ln\left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)}}$$

$$\frac{P_w}{P_r} = \frac{\epsilon_t}{(1 - \epsilon_t)} * \left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)^{.5} * \left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)^{.5}$$

$$\frac{P_w}{P_r} = \frac{\epsilon_t}{(1 - \epsilon_t)} * \frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}$$

$$\frac{P_w}{P_r} \leq \frac{\frac{1}{2}-\gamma}{\frac{1}{2}-\gamma} * \frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}$$

$$\frac{P_w}{P_r} \leq 1$$

Yes, the algorithm assigns the same probability mass function at each round, as the ratio of right to wrong is always 1.

c.

$$Z_t \le \left(\frac{1}{2}+\gamma\right)e^{-\propto} + \left(\frac{1}{2}-\gamma\right)e^{\propto}$$

$$Z_t \le \left(\frac{1}{2}+\gamma\right)\left(\frac{\frac{1}{2}-\gamma}{\frac{1}{2}+\gamma}\right)^{.5} + \left(\frac{1}{2}-\gamma\right)\left(\frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma}\right)^{.5}$$

$$Z_t \le 2\left(\left(\frac{1}{2}+\gamma\right)\left(\frac{1}{2}-\gamma\right)\right)^{.5}$$

So,

$$Err(H) \le 2\left(\left(\frac{1}{2}+\gamma\right)\left(\frac{1}{2}-\gamma\right)\right)^{\frac{T}{2}}$$

$$Err(H) \le 2\left(\frac{1}{4}-\gamma^2\right)^{\frac{T}{2}}$$

2)

$$D_{t+1}(i) = \frac{D_t e^{-\propto_t y_i h_t(x_i)}}{Z_t} \qquad \propto_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

So,

$$R_{D_{t+1}} = \frac{1}{Z_t}\sum_{i=1}^{m} D_{t,i}\, e^{-\propto_t y_i h_t(x_i)}$$

$$R_{D_{t+1}} = \frac{e^{\propto_t}}{Z_t}\sum_{i=1}^{m} D_{t,i}$$

$$R_{D_{t+1}} = \frac{e^{\propto_t}}{Z_t}\epsilon_t$$

$$R_{D_{t+1}} = \frac{e^{\frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)}}{2\sqrt{\epsilon_t\,(1-\epsilon_t)}}\epsilon_t$$

$$R_{D_{t+1}} = \frac{\frac{\sqrt{1-\epsilon_t}}{\sqrt{\epsilon_t}}}{2\sqrt{\epsilon_t}\,\sqrt{1-\epsilon_t}}\epsilon_t$$

$$R_{D_{t+1}} = \frac{\sqrt{(1-\epsilon_t)}}{\sqrt{\epsilon_t}} * \frac{1}{2\sqrt{\epsilon_t}\,\sqrt{(1-\epsilon_t)}}\epsilon_t$$

$$R_{D_{t+1}} = \frac{1}{2}$$

The weak learner assumption states that for a hypothesis h, there is a D_{t+1} error that is less than ½. Since our empirical error of the D_{t+1} distribution is equal to ½ , we would need to select a different h at round t+1.

3)

Assume $r = num\_points\_wrong$

$$\epsilon_1 \leq \frac{r}{m}$$

$$\propto_1 \leq \frac{1}{2} * \ln\left(\frac{\left(1 - \frac{r}{m}\right)}{\frac{r}{m}}\right))$$

$$D_{wrong} \leq \frac{\sqrt{\left(\frac{m}{r}\right) - 1}}{m}$$

$$D_{wrong} \leq \frac{1}{m * \sqrt{\left(\frac{m}{r}\right) - 1}}$$

Normalized, assuming m & r are positive:

$$D_{wrong} \leq \frac{1}{2 * \sqrt{1 - r} * \sqrt{m - r}}$$

$$D_{wrong} \leq \frac{1}{2 * r}$$

After another round, both distributions go back to 1/m.

So, at round T, the distributions are both 1/m if T is even and the above distributions if T is odd.