Homework 3 STOR 767: Theory Part

Due October 13, 2020

Student Name: Leo Li (PID: 730031954)

Instructions

- Edit this LaTeX file with your solutions and generate a PDF file from it. Upload **both the tex and the pdf** file to Sakai.
- Use proper fonts for a clear presentation:
  x for an observed value; X for a random variable; $\boldsymbol{X}$ for a vector; $\mathbf{X}$ for a matrix.
- You are allowed to work with other students but homework should be in your own words. Identical solutions will receive a **0** in grade and will be investigated.

1. Show that in AdaBoost, the training error satisfies the bound

$$\frac{1}{n}\sum_{1}^{n} I(y_i \neq H(\boldsymbol{x}_i)) \leq \frac{1}{n}\sum_{i=1}^{n} \exp(-y_i f(\boldsymbol{x}_i)) = \prod_{t=1}^{T} Z_t,$$

where $f(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})$ and $H(\boldsymbol{x}) = sign(f(\boldsymbol{x}))$.

Proof.

First, we would like to prove the first inequality, which is,

$$\frac{1}{n}\sum_{1}^{n} I(y_i \neq H(\boldsymbol{x}_i)) \leq \frac{1}{n}\sum_{i=1}^{n} \exp(-y_i f(\boldsymbol{x}_i)).$$

According to the fact that the 0-1 loss is bounded by the exponential loss, or in other words, $I(y_i \neq H(\boldsymbol{x}_i)) \leq \exp(-y_i f(\boldsymbol{x}_i))$, which can be easily verified by using Figure 1 in the Appendix. Therefore, the above inequality holds.

Second, we would like to show the second equality, which is,

$$\frac{1}{n}\sum_{i=1}^{n} \exp(-y_i f(\boldsymbol{x}_i)) = \prod_{t=1}^{T} Z_t.$$

Consider the updated weights in each step, we have that,

$$D_i^{(1)} = \frac{1}{n};$$

$$D_i^{(2)} = \frac{\exp(-y_i \alpha_1 h_1(\boldsymbol{x}_i))}{n Z_1};$$

$$D_i^{(3)} = \frac{D_i^{(2)} \exp(-y_i \alpha_2 h_2(\boldsymbol{x}_i))}{Z_2} = \frac{\exp(-y_i \alpha_1 h_1(\boldsymbol{x}_i)) \exp(-y_i \alpha_2 h_2(\boldsymbol{x}_i))}{n Z_1 Z_2}.$$

Using proof by induction, we can show that,

$$D_i^{(T)} = \frac{D_i^{(T-1)} \exp(-y_i \alpha_T h_T(\boldsymbol{x}_i))}{Z_T} = \frac{\prod_{t=1}^T \exp(-y_i \alpha_t h_t(\boldsymbol{x}_i))}{n \prod_{t=1}^T Z_t} = \frac{\exp(-y_i f(\boldsymbol{x}_i))}{n \prod_{t=1}^T Z_t}.$$

Since the sum of all the weights should equal to one, i.e., $\sum_{i=1}^n D_i^{(T)} = 1$, we have that,

$$\sum_{i=1}^n D_i^{(T)} = \frac{\frac{1}{n} \sum_{i=1}^n \exp(-y_i f(\boldsymbol{x}_i))}{\prod_{t=1}^T Z_t} = 1,$$

so that we have,

$$\frac{1}{n} \sum_{i=1}^n \exp(-y_i f(\boldsymbol{x}_i)) = \prod_{t=1}^T Z_t.$$

Since we have proved both of the inequality part and equality part, we have shown that the desired statement about the training error of the AdaBoost holds.

$\square$

Appendix

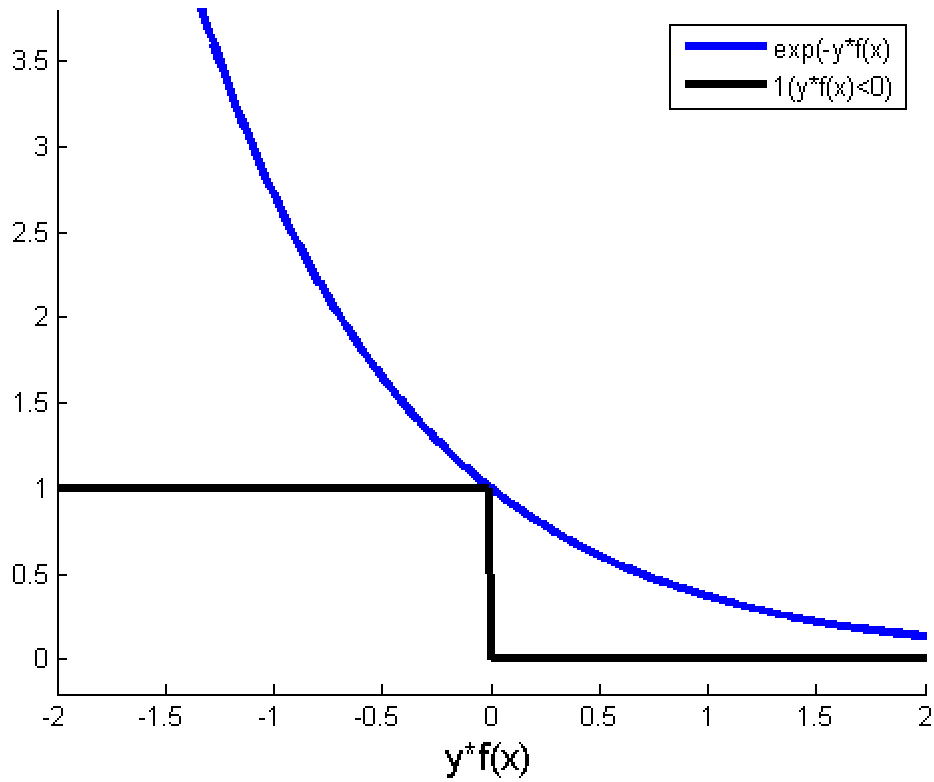

Figure 1: 0-1 loss function vs. exponential loss function