University of Waterloo
CS 480
r2knowle: 2023-11-13

Exercise # 1

Q1a) Our goal will be to prove that regardless of the updates, p_i will remain constant, as if this is the case it implies the updates are equivalent. Note that this means we need to prove for all w_i and $\widetilde{w}_i$ is must be the case that:

$$p_i^t = \frac{w_i^t}{\sum_j w_j^t} = \frac{\widetilde{w}_i^t}{\sum_j \widetilde{w_j}}$$

Further more this implies that we need to prove $w_i = a\widetilde{w}_i$ where $\exists a \in \mathbb{R}$ as if this is true it follows that:

$$\begin{aligned}\frac{w_i^t}{\sum_j w_j^t} &= \frac{a}{a} \times \frac{w_i^t}{\sum_j w_j^t} \\ &= \frac{aw_i^t}{\sum_j aw_j^t} \\ &= \frac{\widetilde{w}_i^t}{\sum_j \widetilde{w}_j^t}\end{aligned}$$

Therefore using induction we will prove that for any t that: $w_i^t = a\widetilde{w}_i^t$.

Basecase: At time t = 1 we initialize the weights to be $[\frac{1}{n}, ..., \frac{1}{n}] \in \mathbb{R}$ for both w_i and $\widetilde{w}_i$ therefore it follows that $w_i = a\widetilde{w}_i$ for the constant a = 1, satisfying our basecase.

Inductive Hypothesis: We are going to assume that for any $t \geq 0, w_i^t = a\widetilde{w}_i^t$, we are now going to prove given the definitions in the question that $w_i^{t+1} = a\widetilde{w}^{t_i+1}$.

Inductive Step: To prove that $w_i^{t+1} = a\widetilde{w}^{t_i+1}$, we will show that w_i^{t+1} is equivlant to $a\widetilde{w}^{t_i+1}$ given our hypothesis where $\exists a \in \mathbb{R}$. We will first simplify w_i^{t+1} to get:

$$\begin{aligned}w_i^{t+1} &= w_i^t \text{exp}(-y_i\beta_t h_t(\mathbf{x}_i)) \\ &= w_i^t \text{exp}(-y_i(\frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t})h_t(\mathbf{x}_i)) \\ &= w_i^t \text{exp}(-y_i h_t(\mathbf{x}_i)\frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}) \\ &= w_i^t \text{exp}(\log((\frac{1-\epsilon_t}{\epsilon_t})^{-y_i h_t(\mathbf{x}_i)\frac{1}{2}})) \\ &= w_i^t \frac{1-\epsilon_t}{\epsilon_t}^{-y_i h_t(\mathbf{x}_i)\frac{1}{2}} \\ &= w_i^t \frac{\epsilon_t}{1-\epsilon_t}^{y_i h_t(\mathbf{x}_i)\frac{1}{2}}\end{aligned}$$

Note that ϵ_t is constant for both types of update. Before we continue we also need to mention that y_i and h_i are only ever 1 or -1. This means that if they are the same value $y_i h_i = 1$ and -1 if otherwise.

It thus follows that $1 - |h_t(x_i) - y_i|$ will be equivalent, as if they are the same the equation will be 1, and if they are different its - 1. Therefore we can replace our equation to get:

$$
\begin{aligned}
&= w_i^t \widetilde{\beta}_t^{(1-|h_t(x_i)-y_i|)\frac{1}{2}} \\
&= w_i^t \widetilde{\beta}_t^{(1-|(2\widetilde{h}_t(x_i)-1)-(2\widetilde{y}_i-1)|)\frac{1}{2}} \\
&= w_i^t \widetilde{\beta}_t^{\frac{1}{2}-|\widetilde{y}_i-\widetilde{h}_t(x_i)|} \\
&= w_i^t \widetilde{\beta}_t^{1-|\widetilde{y}_i-\widetilde{h}_t(x_i)|} \widetilde{\beta}_t^{-\frac{1}{2}} \\
&= a\widetilde{w}_i^t \widetilde{\beta}_t^{1-|\widetilde{y}_i-\widetilde{h}_t(x_i)|} \widetilde{\beta}_t^{-\frac{1}{2}}
\end{aligned}
$$

Note that since a just needs to be a constant for each w_i but not for each iteration and β_t is a constant as ϵ_t is a constant, we will set a new $a_t = a_{t-1} \times \widetilde{\beta}_t^{-\frac{1}{2}}$ and thus we get:

$$
\begin{aligned}
&= a_{t-1}\widetilde{w}_i^t \widetilde{\beta}_t^{1-|\widetilde{y}_i-\widetilde{h}_t(x_i)|} \widetilde{\beta}_t^{-\frac{1}{2}} \\
w_i^{t+1} &= a_t\widetilde{w}_i^t \widetilde{\beta}_t^{1-|\widetilde{y}_i-\widetilde{h}_t(x_i)|} \\
w_i^{t+1} &= a\widetilde{w}_i^{t+1}
\end{aligned}
$$

Thus proving the induction, and showing how $\forall t \geq 0, w_i^t = a_t \widetilde{w}_i^t$. Which thus means p_i will be the same for both updates, and therefore both equations are equivalent.

Q1b) It then follows from our definition of expected values:

$$E[e^{-yH}|X = x] = e^{-H} \times Pr(y = 1|X = x) + e^{H} \times Pr(y = -1|X = x)$$

If we take the derivative w.r.t to H we thus get:

$$-e^{-H} \times Pr(y = 1|X = x) + e^{H} \times Pr(y = -1|X = x) = 0$$

We can then rearrange to get:

$$
\begin{aligned}
-e^{-H} \times Pr(y = 1|X = x) &= -e^{H} \times Pr(y = -1|X = x) \\
\frac{Pr[y = 1|X = x]}{Pr[y = -1|X = x]} &= \frac{-e^{H}}{-e^{-H}} \\
&= e^{2H}
\end{aligned}
$$

Now we will take the log of both sides to get:

$$
\begin{aligned}
\frac{Pr[y=1|X=x]}{Pr[y=-1|X=x]} &= e^{2H} \\
\log\frac{Pr[y=1|X=x]}{Pr[y=-1|X=x]} &= \log(e^{2H}) \\
&= 2H \\
&\propto H \\
&= \sum_{i}^{T} \beta_t h_t(x)
\end{aligned}
$$

Thus proving how that minimizer of the given exponential loss is proportional to the log odd loss as required.

Q1c) We are given the definition that:

$$\epsilon_t = \epsilon_t(h_t(x)) = \sum_{i=1}^{n} p_i^t \times [[h_t(x_i) \neq y_i]]$$

Note that since h_i and y_i is either -1 or 1, this would imply that $[[-h_t(x_i) \neq y_i]]$ is the same as $[[h_t(x_i) = y_i]]$. Therefore we can define $\widetilde{\epsilon_t}$ to be:

$$\widetilde{\epsilon_t} = \widetilde{\epsilon_t}(-h_t(x)) = \sum_{i=1}^{n} p_i^t \times [[h_t(x_i) = y_i]]$$

Notice that the sum of $\widetilde{\epsilon_t}$ and ϵ_t has the following property:

$$
\begin{aligned}
\epsilon_t + \widetilde{\epsilon_t} &= \sum_{i=1}^{n} p_i^t \times [[h_t(x_i) \neq y_i]] + \sum_{i=1}^{n} p_i^t \times [[h_t(x_i) = y_i]] \\
&= \sum_{i=1}^{n} p_i^t \\
&= \sum_{i=1}^{n} \frac{w_i^t}{\sum_{j=1}^{n} w_j^t} \\
&= \frac{\sum_{i=1}^{n} w_i^t}{\sum_{j=1}^{n} w_j^t} \\
&= 1
\end{aligned}
$$

Which thus implies that:

$$
\begin{aligned}
\epsilon_t &= 1 - \widetilde{\epsilon_t} \\
\widetilde{\epsilon_t} &= 1 - \epsilon_t
\end{aligned}
$$

Moving forward we can then define a new $\widetilde{\beta}_t$ based on the $\widetilde{\epsilon}_t$:

$$\widetilde{\beta}_t = \frac{1}{2} \log \frac{1-\widetilde{\epsilon}_t}{\widetilde{\epsilon}_t}$$

To prove the updates to w_t are equivalent we will prove that:

$$w_i^t \text{exp}(-y_i \beta_t h_t(x)) = w_i^t \text{exp}(-y_i \widetilde{\beta}_t \widetilde{h}_t(x))$$

To do so we will begin with:

$$\begin{aligned} w_i^t \text{exp}(-y_i \beta_t h_t(x)) &= w_i^t \text{exp}(y_i \beta_t \widetilde{h}_t(x)) \\ &= w_i^t \text{exp}(y_i (\frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t}) \\ &= w_i^t \text{exp}(y_i (\frac{1}{2} \log \frac{\widetilde{\epsilon}_t}{1-\widetilde{\epsilon}_t}) \widetilde{h}_t(x)) \\ &= w_i^t \text{exp}(y_i (\frac{1}{2} \log \frac{1-\widetilde{\epsilon}_t}{\widetilde{\epsilon}_t}^{-1}) \widetilde{h}_t(x)) \\ &= w_i^t \text{exp}(y_i (-\frac{1}{2} \log \frac{1-\widetilde{\epsilon}_t}{\widetilde{\epsilon}_t}) \widetilde{h}_t(x)) \\ &= w_i^t \text{exp}(-y_i \widetilde{\beta}_t \widetilde{h}_t(x)) \end{aligned}$$

Thus showing how the updates are equivalent to each other.

Q1d) To begin we are given the following minimizer:

$$\text{min}_\beta E[e^{-y\beta h_t(x)} | X = x]$$

From this we can then derive that:

$$= \text{min}_\beta \sum_i^n p_i \times e^{-y\beta h_t(x)}$$

Note that we have two cases, either the data is correctly classified in which case $-yh_t(x)$ is negative or the case where the data is classified incorrectly such that $-yh_t(x)$ is positive. Thus we can split this into:

$$= \text{min}_\beta \sum_{i,y=h_t(x)}^n p_i \times e^{-y\beta h_t(x)} + \sum_{i,y\neq h_t(x)}^n p_i \times e^{y\beta h_t(x)}$$

Another way we can write this is to use the $[[h_t \neq y_i]]$ notation as its equivalent:

$$= \text{min}_\beta \sum_i^n p_i \times e^{-\beta} \times [[h_t = y_i]] + \sum_i^n p_i \times e^{\beta} \times [[h_t \neq y_i]$$

$$= \text{min}_\beta e^{\beta} \sum_i^n p_i \times [[h_t = y_i]] + e^{\beta} \sum_i^n p_i \times [[h_t \neq y_i]$$

From the definition given we $\sum_i^n p_i \times [[h_t = y_i]] = 1 - \epsilon_t$ and $\sum_i^n p_i \times [[h_t \neq y_i]] = \epsilon_t$ and so we get:

$$
\begin{aligned}
&= \min_\beta e^{-\beta} \sum_i^n p_i \times [[h_t = y_i]] + e^\beta \sum_i^n p_i \times [[h_t \neq y_i] \\
&= \min_\beta e^{-\beta}(1 - \epsilon_t) + e^\beta \epsilon_t
\end{aligned}
$$

Taking the derivative w.r.t β gives us:

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

Showing that our β is optimal as required.

Q1e) We are given by equation 10), that: