Foundations of Machine Learning — Homework Assignment 3

Anirudhan J Rajagopalan
N18824115
ajr619

December 09, 2015

A. Boosting-type Algorithm

1. Bound of $1_{u \leq 0}$ Proof of convexity and differentiability

Given: $\phi_p(u) = \max((1+u)^p, 0)$

To prove: 1. Function $\phi_p(u)$ is convex and differentiable
And 2. $\forall u \in \mathbb{R}$ and $p > 1, 1_{u \leq 0} \leq \phi_p(-u)$

Proof of $\forall u \in \mathbb{R}$ and $p > 1, 1_{u \leq 0} \leq \phi_p(-u)$ There are three cases here:

When $u = 0$

$$\begin{aligned} 1_{u \leq 0} =& 1 \\ \phi_p(-u) =& \max((1-0)^p, 0) = 1 = 1_{u \leq 0} \end{aligned}$$

When $u < 0$

$$\begin{aligned} 1_{u \leq 0} =& 1 \\ \phi_p(-u) =& \max((1-u)^p, 0) \\ =& \max((1+u)^p, 0) \text{Since u is negative, -u is positive} \\ >& 1 \\ >& 1_{u \leq 0} \end{aligned}$$

When $u > 0$

$$\begin{aligned} 1_{u \leq 0} =& 0 \\ \phi_p(-u) =& \max((1-u)^p, 0) \\ =& \begin{cases} 0, & \text{if p is odd} \\ (1-u)^p, & \text{if p is positive} \end{cases} \\ \geq& 0 \quad \forall \quad \mathbb{R} \geq 1_{u \leq 0} \end{aligned}$$

Hence proved

Proof of Convexity and Differentiability: The function can be written as a piecewise function based on the value of p:

When p is even (2 and higher value even numbers)

$$\phi_p(u) = (1+u)^p$$

Since p > 1, the function is differentiable.

$$\begin{aligned} \phi_p^{'}(u) =& p(1+u)^{p-1} \\ \phi_p^{''}(u) =& (p-1)p(1+u)^{p-2} \\ \because p > 1, \phi_p^{''}(u) >& 0 \end{aligned}$$

Since the double derivative is greater than zero, the function is convex and differentiable.

When p is odd (1 and higher value odd numbers) the function can be defined using the piecewise function

$$\phi_p\left(x\right) = \begin{cases} 0, & \text{if } u \leq -1 \\ (1+u)^p, & otherwise \end{cases}$$

The function is piecewise continuous and differentiable. We have to check that the function is differentiable at u = -1 to show that the function is differentiable and continuous. We should also check the double derivative to show that the function is convex.

The first derivative can be found using limits.

$$\begin{aligned}
\lim_{h \to 0+} \frac{\phi_p(-1+h) - \phi_p(-1)}{h} &= \lim_{h \to 0+} \frac{(1+(-1+h))^p - (1-1)^p}{h} \\
&= \lim_{h \to 0+} \frac{h^p - 0}{h} \\
&= \lim_{h \to 0+} h^{p-1} \\
&=0
\end{aligned}$$

Also

$$\begin{aligned}
\lim_{h \to 0-} \frac{\phi_p(-1+h) - \phi_p(-1)}{h} &= \lim_{h \to 0+} \frac{(1+(-1+h))^p - (1-1)^p}{h} \\
&= \lim_{h \to 0-} \frac{h^p - 0}{h} \\
&= \lim_{h \to 0-} h^{p-1} \\
&=0
\end{aligned}$$

Since both the limit values are equal, the function is differentiable. Therefore:

$$\phi_p^{'}\left(x\right) = \begin{cases} 0, & \text{if } u \leq -1 \\ p(1+u)^{p-1}, & otherwise \end{cases}$$

For showing the function is convex, we find the double differeniable using limits for $\phi_p^{'}$ and show that it is non negative

$$\begin{aligned}
\lim_{h \to 0+} \frac{\phi_p^{'}(-1+h) - \phi_p^{'}(-1)}{h} &= \lim_{h \to 0+} \frac{p(1+(-1+h))^{p-1} - 0}{h} \\
&= \lim_{h \to 0+} \frac{p(h)^{p-1}}{h} \\
&= \lim_{h \to 0+} p(h)^{p-2} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\lim_{h \to 0-} \frac{\phi_p^{'}(-1+h) - \phi_p^{'}(-1)}{h} &= \lim_{h \to 0-} \frac{p(1+(-1+h))^{p-1} - 0}{h} \\
&= \lim_{h \to 0-} \frac{p(h)^{p-1}}{h} \\
&= \lim_{h \to 0-} p(h)^{p-2} \\
&= 0
\end{aligned}$$

Since the left and right derivatives are equal, the function is $\phi_p^{'}$ is differentiable. Which implies that the original function ϕ_p is double differentiable.

$$\phi_p^{''}(x) = \begin{cases} 0, & \text{if } u \le -1 \\ (p-1)p(1+u)^{p-2}, & otherwise \end{cases}$$

Since in both cases the double differentiable is non negative, the function is convex and differentiable.

C. Randomized Halving

1. Psuedo code

Algorithm 1 Randomized Halving

$H_1 \leftarrow H$
for t $\leftarrow$ 1 to T **do**
  $RECEIVE(x_t)$
  $r_t \leftarrow \frac{\sum_{i:y_{t,i}=1} 1}{|H_t|}$
  $p_t \leftarrow 1$
  if $r_t \leq \frac{3}{4}$ **then**
    $p_t \leftarrow [\frac{1}{2}\log_2 \frac{1}{1-r_t}]$
  $\hat{y_t} \leftarrow GetRandomNumberWithProbability([1,0],[p_t, 1-p_t])$
  $RECEIVE(y_t)$
  if $\hat{y_t} \neq y_t$ **then**
    $H_{t+1} \leftarrow \{c \in H_t : c(x_t) = y_t\}$
return H_{T+1}

2. Prove $\forall t \geq 1, E[\mu] \leq \frac{\phi_t - \phi_{t+1}}{2}$

Given: Potential function: $\phi_t = \log_2 \mid H_t \mid$ and $\mu_t = 1_{y_t \neq \hat{y_t}}$

Proof:

We are only considering the case when the predicted value $\hat{y_t}$ is not equal to the received value y_t. The value of expectation can be written as

$$\begin{aligned} E[\mu_t] =& p_t * 1 + (1-p_t) * 0 \\ =& p_t * 1 \end{aligned}$$

The probability of predicting 1 by the randomized algorithm is the probability of making a mistake since we are only considering the cases in which we make mistakes ($\mu_t = 1_{y_t \neq \hat{y_t}}$)

Therefore,

$$\begin{aligned} E[\mu_t] =& p_t \\ =& [\frac{1}{2}\log_2 \frac{1}{1-r_t}] 1_{r_t \leq \frac{3}{4}} + 1_{r_t > \frac{3}{4}} \\ \leq& [\frac{1}{2}\log_2 \frac{1}{1-r_t}] \end{aligned}$$

Since, the Expectation will be 1 when $r_t > \frac{3}{4}$ which corresponds to the maximum expectation here, we can upper bound the expectation by using $r_t \leq \frac{3}{4}$ as $[\frac{1}{2}\log_2 \frac{1}{1-r_t}]$ equals 1 when $r_t = \frac{3}{4}$

Let E_1, E_0, E_t denote the number of experts predicting 1, 0 and the total number of experts in a round t.

$$\begin{aligned}
E[\mu_t] \leq & [\frac{1}{2}\log_2 \frac{1}{1-r_t}] \\
= & [\frac{\log_2 \frac{1}{1-r_t}}{2}] \\
= & [\frac{\log_2 \frac{1}{1-r_t}}{2}] \\
= & [\frac{\log_2 \frac{1}{1-\frac{|E_1|}{|H_t|}}}{2}] \\
= & [\frac{\log_2 \frac{|H_t|}{|H_t|-|E_1|}}{2}] \\
= & [\frac{\log_2 \frac{|H_t|}{|E_0|}}{2}] \\
= & [\frac{\log_2 \mid H_t \mid - \log_2 \mid E_0 \mid}{2}] \\
= & [\frac{\log_2 \mid H_t \mid - \log_2 \mid H_{t+1} \mid}{2}] \\
= & [\frac{\phi_t - \phi_{t+1}}{2}] \\
\therefore E[\mu_t] \leq & \frac{\phi_t - \phi_{t+1}}{2}
\end{aligned}$$

3. Expected number of mistakes.

Given: N be the total number of experts at the beginnning of the iterations (denoted by H_1). Since we are considering a relizable scenario at the end of the algorithm the number of experts should be atleast one. So $H_T = 1$.

To Prove: The expected number of mistakes made by Randomized Halving is at most $\frac{1}{2}\log_2 N$

Proof:

Lets consider the total expectation of mistakes of the Randomized Halving algorithm over T iterations.

$$
\begin{aligned}
E[\mu_T] \leq& \sum_{t=1}^{T} \frac{\phi_t - \phi_{t+1}}{2} \\
\leq& \frac{(\phi_{H_1} - \phi_{H_2}) + (\phi_{H_2} - \phi_{H_3}) \cdots (\phi_{H_{T-2}} - \phi_{H_{T-1}}) + (\phi_{H_{T-1}} - \phi_{H_T})}{2} \\
\leq& \frac{\phi_{H_1} - \phi_{H_T}}{2} \\
\leq& \frac{\phi_N - \phi_T}{2} \\
\leq& \frac{\phi_N}{2} \\
\leq& \frac{1}{2} \log_2 N
\end{aligned}
$$

Hence proved.

Here $\phi_T = 0$ because the number of experts at line T is 1. Therefore, log of T will be zero.

4. [Bonus Question]

As we have seen in the previous answer, the mistakes made by the randomized algorithm is bounded by $\frac{1}{2}\log_2 N$. This upper bound is dependent only on the number of initial experts, N. Therefore any randomized algorithm that is dependent on the opinion of the experts to generate its predictions will have similar upper bound of $\lfloor \frac{1}{2}\log_2 N \rfloor$. The floor function is used as mistakes are natural numbers.