IE613: Online Machine Learning **Jan-Apr 2016**

Lecture 4: Weighted Majority Algorithm

Lecturer: M. K. Hanawal *Scribes: Manan Doshi*

4.1 Recap

For the unrealizable(agnostic case), we have computed the expected value of regret as follows:

$$\mathbb{E}(R(A,T)) = \sup_{\substack{\{(x_i,y_i), i\in[T]\} \\ \in (x,y)^T}} \left\{ \sum_{t=1}^{T} |p_t - y_t| - \inf_{h\in\mathcal{H}} \sum_{t=1}^{T} |h(x_t) - y_t| \right\} \tag{4.1}$$

where $p_t = Pr(\hat{y}_t = 1)$. Thus,

$$\mathbb{E}(|\hat{y}_t - y_t|) = |p_t - y_t|$$

4.2 Bounds of expected regret for agnostic case

Theorem 4.1 *For every hypothesis class $\mathcal{H}$, $\exists$ an Online Learning Algorithm, A, whose predictions come from $[0,1]$ and has a regret bound such that*

$$\mathbb{E}(R(A,T)) \leq O(\sqrt{T}) \qquad \forall h \in \mathcal{H}$$

In this chapter we will show an algorithm that achieves this upper bound. Specifically, we will show that

$$\sum_{t=1}^{T} |p_t - y_t| - \inf_{h\in\mathcal{H}} \sum_{t=1}^{T} |h(x_t) - y_t| \leq \sqrt{2\log(|\mathcal{H}|)T} = O(\sqrt{T}) \qquad \forall h \in \mathcal{H}$$

4.2.1 Weighted Majority Algorithm

The algorithm that achieves the above bound is the weighted majority algorithm. Consider the hypothesis class $\mathcal{H} = \{h_1, h_2, h_3 ... h_d\}$, where h_i's are various hypothesis and $|\mathcal{H}| = d$. The loss at a time instance is $l_{t,i} = |h_i(x_t) - y_t|$ where $h_i(x_t) = \hat{(y)}_t$ is the prediction by our online learning algorithm and y_t is the actual label.

Algorithm 4.1: The Weighted Majority Algorithm

Input : Hypothesis class $\mathcal{H}$
Parameter: $\eta \in [0, 1]$
Initialize : $\tilde{w}^{(1)} = [1, 1, 1, ..., 1]$ in $\mathbb{R}^d$
for $t \leftarrow 1$ **to** T **do**
  Set $w_i^{(t)} = \frac{\tilde{w}_i^{(t)}}{\sum_i \tilde{w}_i^{(t)}}$
  Play i according to the distribution $w^{(t)}$
  Receive loss vector $l_t = \{l_{t,i} : \forall i \in d\}$ where $l_{t,i}$ is the error in prediction of hypthesis h_i
  Update $\forall i, \tilde{w}_i^{(t+1)} = \tilde{w}_i^{(t)} e^{-\eta l_{t,i}}$

Theorem 4.2 *Let $d = |\mathcal{H}|$ and $T > 2\log(d)$, then,*

$$\mathbb{E}(R(A, T)) \leq \sqrt{2\log(dT)} = \sqrt{2\log(|\mathcal{H}|)T}$$

We need a couple of inequalities for this proof:

$$e^{-a} \leq 1 - a + \frac{a^2}{2} \qquad \forall a \in (0, 1) \tag{4.2}$$

$$e^{-a} \geq 1 - a \qquad \forall a \leq 1 \tag{4.3}$$

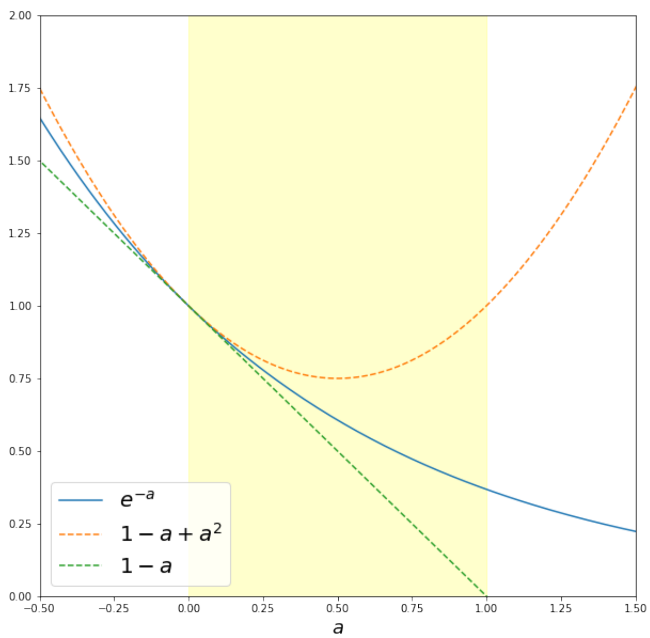

Figure 4.1: Plots demonstrating inequalities 4.2 and 4.3

Proof: Let $Z_t = \sum_{i=1}^{d} \tilde{w}_i^{(t)}$

$$
\begin{aligned}
\frac{Z_{t+1}}{Z_t} &= \frac{\sum_{i=1}^{d} \tilde{w}_i^{(t)} \cdot e^{-\eta l_{t,i}}}{Z_t} \\
&= \sum_{i=1}^{d} w_i^{(t)} e^{-\eta l_{t,i}} \\
&\leq \sum_{i=1}^{d} w_i^{(t)} \left\{ 1 - \eta l_{t,i} + \frac{\eta^2 l_{t,i}^2}{2} \right\} && \text{from equation 4.2} \\
&\leq 1 - \sum_{i=1}^{d} w_i^{(t)} \left\{ \eta l_{t,i} - \frac{\eta^2 l_{t,i}^2}{2} \right\} && \text{Since } \sum w_i^{(t)} = 1 \\
&\leq e^{-\sum_{i=1}^{d} w_i^{(t)} \left\{ \eta l_{t,i} - \frac{\eta^2 l_{t,i}^2}{2} \right\}} && \text{from equation 4.3} \\
\log \frac{Z_{t+1}}{Z_t} &\leq - \left[\eta \sum_{i=1}^{d} l_{t,i} w_i^{(t)} - \frac{\eta^2}{2} \sum_{i=1}^{d} w_i^{(t)} l_{t,i}^2 \right] && \text{Taking } log \text{ on both sides} \\
\sum_{t=1}^{T} \log \frac{Z_{t+1}}{Z_t} &\leq - \sum_{t=1}^{T} \left[\eta \sum_{i=1}^{d} l_{t,i} w_i^{(t)} - \frac{\eta^2}{2} \sum_{i=1}^{d} w_i^{(t)} l_{t,i}^2 \right] && \text{Summing over } t \\
\log \frac{Z_{T+1}}{Z_1} &\leq -\eta \sum_{t=1}^{T} \langle l_t, w^{(t)} \rangle + \frac{\eta^2}{2} T && \sum w_i^{(t)} l_{t,i}^2 \leq 1 \text{ since } \sum w_i^{(t)} = 1 \text{ and } l_{t,i}^2 \leq 1 \forall l_i \\
\log \frac{Z_{T+1}}{d} &\leq -\eta \sum_{t=1}^{T} \langle l_t, w^{(t)} \rangle + \frac{\eta^2}{2} T && \text{Since } Z_1 = d
\end{aligned}
$$

We have an upper bound for the intermediate quantity for $\log \frac{Z_{T+1}}{d}$. Now we try and get a lower bound.

$$
\begin{aligned}
\tilde{w}_i^{(t+1)} &= \tilde{w}_i^{(t)} e^{-\eta l_{t,i}} \\
&= e^{-\sum_{t=1}^{T} \eta l_{t,i}} \\
Z_{T+1} &= \sum_{i=1}^{d} e^{-\sum_{t=1}^{T} \eta l_{t,i}} \\
&\geq max \left(e^{-\sum_{t=1}^{T} \eta l_{t,i}} \right) \\
\log \frac{Z_{T+1}}{d} &\geq -\eta \min_i \sum_{t=1}^{T} l_{t,i} - \log d && \text{Taking log on both sides and subtracting } \log d
\end{aligned}
$$

Combining both the inequalities

$$
-\eta \min_i \sum_{t=1}^{T} l_{t,i} - \log d \leq \log \frac{Z_{T+1}}{d} \leq -\eta \sum_{t=1}^{T} \langle l_t, w^{(t)} \rangle + \frac{\eta^2}{2} T
$$

$$
\sum_{t=1}^{T} \langle l_t, w^{(t)} \rangle - \min_i \sum_{t=1}^{T} l_{t,i} \leq \frac{\log d}{\eta} + \frac{\eta T}{2}
$$

Note that $\langle l_t, w^{(t)} \rangle$ is the expected loss and is equal to $|p_t - y_t|$. Also, $l_{t,i}$ is the actual loss due to a prediction and is equal to $|h_i(x_t) - y_t|$. This leads us to the statement of theorem 4.2. It can be easily shown that the best bound is achieved by $\eta = \sqrt{\frac{2 \log d}{T}}$ ■