AI-3001: Advanced Topics in Machine Learning **Aug-Nov 2022**

Lecture 5: Exponential Weights Algorithm

Lecturer: Ganesh Ghalme *Scribes: Ganesh Ghalme*

5.1 Exponential Weights ExpWts algorithm

Setting

- Prediction with N experts (need not be constant experts)
- Prediction $p_t = \frac{\sum_{i=1}^{N} w_{i,t-1} f_{i,t}}{\sum_{i=1}^{N} w_{i,t-1}}$
- weight update rule $w_{i,t} = w_{i,t-1} e^{-\ell(f_{i,t}, y_t)}$ with $w_{i,0} = 1$
- Setting $\alpha = e^{-\eta}$, we have that the ExpWts is similar as weighted majority algorithm(there we had binary loss function whereas here the loss function is convex). In this setting (deterministic environment) we will consider a worst case regret guarantees.

1.	$\mathcal{D}$	decision space	convex set
2.	$\ell : \mathcal{D} \times \mathcal{Y} \to \mathbb{R}$	loss function	convex in the first argument
3.	$p_t \in \mathcal{D}$	Algorithms prediction at time t	can be a function of history of prediction and outcomes

Algorithm 1: ExpWts

Input: # experts: N, $\eta \in (0, \infty)$
Initialize: $w_{i,0} = 1$ for all $i \leq N$;
for $t = 1, 2, \cdots$ **do**
- **Input:** Expert recommendations $(f_{i,t})_{i=1}^{N}$;
- **Predict:** $p_t = \frac{\sum_{i=1}^{n} w_{i,t-1} \cdot f_{i,t}}{\sum_{i=1}^{n} w_{i,t-1}}$;
- **Observe:** y_t ;
- **Update**
  - $w_{i,t} \leftarrow w_{i,t-1} \cdot e^{-\eta \ell(f_{i,t}, y_t)}$;

end

Lemma 5.1. *Let $X \in [a, b]$ be a random variable, then for all $\eta \in \mathbb{R}$ we have* $\log(\mathbb{E}[e^{-\eta X}]) \leq -\eta \mathbb{E}[X] + \frac{\eta^2 (b-a)^2}{8}$.

Proof.

$$\log(\mathbb{E}[e^{-\eta X}]) \leq -\eta\mathbb{E}[X] + \frac{\eta^2(b-a)^2}{8} \Leftrightarrow \log(\mathbb{E}[e^{-\eta(X-\mathbb{E}[X])}) \leq \frac{\eta^2(b-a)^2}{8}$$
$$LHS = \log(\mathbb{E}[e^{-\eta(X-\mathbb{E}[X])}) \leq \log(\mathbb{E}[1-\eta(X-\mathbb{E}[X])] + \eta^2\frac{(X-\mathbb{E}(X))^2}{2})$$
$$= \log(1+\eta^2\frac{\mathtt{Var}(X)}{2}) \leq \eta^2\frac{\mathtt{Var}(X)}{2} \leq \eta^2\frac{(b-a)^2}{8} = RHS$$

The last inequality follows from the fact that any random variable $X \in [a,b]$ satisfies $0 \leq \mathtt{Var}(X) \leq \frac{(b-a)^2}{4}$. □

Claim 5.2. *Given a random variable $X \in [a,b]$ we have*

$$\mathit{Var}[X] \leq (b-a)^2/4$$

Proof. Let $\mu = \mathbb{E}[X]$. We have

$$\begin{aligned}
\mathtt{Var}[X] &= \mathbb{E}[X-\mu]^2 \\
&\leq \mathbb{E}[(X-\mu)^2 + (X-a)(b-X)] \\
&= \mathbb{E}[X^2 - 2\mu X + \mu^2 - X^2 + (a+b)X - ab] \\
&= -\mu^2 + (a+b)\mu - ab \\
&= -\mu(\mu-a) + b(\mu-a) \\
&= (\mu-a)(b-\mu)
\end{aligned}$$

The RHS is minimized for $\mu = (b+a)/2$ (check!). This gives the desired result.

Claim 5.3. *Let the loss function be convex and $\mathcal{D}, \mathcal{Y}$ be convex sets. Further, assume that the loss function is bounded in $[0,1]$ i.e. $\ell(x,y) \in [0,1]$ for all $x \in \mathcal{D}, y \in \mathcal{Y}$. Then the regret of* ExpWeights *algorithm is upper bounded by*

$$\mathcal{R}_T(\text{ExpWts}) \leq \frac{\log(N)}{\eta} + \frac{\eta \cdot T}{8}.$$

In particular, for $\eta = \sqrt{\frac{8\log(N)}{T}}$ we have $\mathcal{R}_T(\text{ExpWts}) \leq \sqrt{\frac{T\log(N)}{2}}$.

Proof. We will again use the potential function argument with $W_t := \sum_{i=1}^n w_{i,t-1}$ as in . However we will look at it slightly differently in this case.

$$\begin{aligned}
\log(W_{T+1}) &= \log\Big(\frac{W_{T+1}}{W_T} \cdot \frac{W_T}{W_{T-1}} \cdots \frac{W_2}{W_1} \cdot W_1\Big) \\
&= \sum_{t=1}^{T} \log\Big(\frac{W_{t+1}}{W_t}\Big) + \log(N) \\
&= \sum_{t=1}^{T} \log\Big(\sum_{i=1}^{N} (\frac{w_{i,t-1}}{W_t}) \cdot e^{-\eta\ell(f_{i,t},y_t)}\Big) + \log(N)
\end{aligned}$$

Consider a r.v. Y which takes value $\ell(f_{i,t}, y_t)$ with probability $\frac{w_{i,t-1}}{W_t}$

$$
\begin{aligned}
&= \sum_{t=1}^{T} \log\left(\mathbb{E}[\cdot e^{-\eta Y}]\right) + \log(N) \\
&\leq \sum_{t=1}^{T} \left[-\eta\mathbb{E}[Y] + \frac{\eta^2}{8}\right] + \log(N) && \text{(from Lemma 5.1)} \\
&= \sum_{t=1}^{T} \left[-\eta\mathbb{E}[\ell(f_{i,t}, y_t)] + \frac{\eta^2}{8}\right] + \log(N) \\
&\leq \sum_{t=1}^{T} -\eta\ell(\mathbb{E}[f_{i,t}], y_t) + \frac{\eta^2 T}{8} + \log(N) && \text{(Using Jensen's inequality)} \\
&= -\eta\sum_{t=1}^{T} \ell(p_t, y_t) + \frac{\eta^2 T}{8} + \log(N) && \text{(since } \mathbb{E}[f_{i,t}] = \textstyle\sum_{i=1}^{N}\left(\frac{w_{i,t-1}}{W_t}\right)f_{i,t} = p_t.\text{)}
\end{aligned} \tag{5.1}
$$

On the other hand, we have

$$\log(e^{-\eta\sum_{t=1}^{T}\ell(f_{i,t},y_t)}) \leq \log(W_{T+1}) \text{ for all } i = 1, 2, 3, \cdots, N \tag{5.2}$$

Using Equations 5.1 and 5.2 above, we have

$$\sum_{t=1}^{T} \ell(p_t, y_t) - \sum_{t=1}^{T} \ell(f_{i,t}, y_t) \leq \frac{\eta \cdot T}{8} + \frac{\log(N)}{\eta} \tag{5.3}$$

This holds for all experts i and hence also holds for the best expert. We optimize over the value of η to minimize the value of RHS to get an upper bound of $\sqrt{\frac{T\log(N)}{2}}$. □

Above theorem shows that one can achieve $O(1/\sqrt{T})$ convergence rate. This shows that convexity significantly increases the performance of the algorithm. Below is a brief summary of results we showed so far. In the next lecture we will look at the setting given in the fifth row of the table. We will motivate this setting by a portfolio optimization problem.

	Algorithm	Regret guarantee	Loss function	Comments
1.	MAJ WMAJ NAIVE	O(T)	0-1 loss	-
2.	RWMAJ	$O(\sqrt{T})$	0-1 loss	Expected loss
3.	WMAJ	$O(\sqrt{T})$	Convex	Deterministic alg arbitrary prediction from convex set
4.	FTBE	$O(\log(T))$	Squared loss (1/2-exp concave)	Constant experts(actions setting)
5.	WMAJ	$O(\log(T))$	exp-concave	Deterministic algorithm Constant experts (actions setting)

Table 5.1: Caption