E1 245: Online Prediction & Learning	**Fall 2015**
Lecture 7 — August 25	
Lecturer: Aditya Gopalan	*Scribe: Arvind Kumar*

7.1 RECAP

Randomized-EXP-WTS(η)
INIT :

$$W_{i,1} = 1 \quad \forall i \in \mathscr{E}$$

for $t \geq 1$:
Sample an expert

$$\mathbb{I}_t \sim W_t$$

Predict

$$\hat{p}_t = f_{I\,i,t}$$

$$W_{i,t+1} \rightarrow W_{i,t} \exp\left[-\eta\, l\left(f_{i,t}, y_t\right)\right] \quad \forall i$$

7.2 Regret bound for R-EXPWTS

Theorem 7.1. *High probability regret bound for R-EXPWTS.*

$$For \quad l : \mathscr{D} \times \mathscr{Y} \rightarrow [0,1], \quad |\mathscr{E}| = N, \quad \forall \quad 0 < \delta < 1$$

$$\mathbb{P}\left[\hat{L_T} - \mathbb{E}\left[\hat{L_T}\right] > \sqrt{\frac{T}{2}\log\frac{1}{\delta}}\right] \quad \leq \quad \delta$$

Proof: With probability $1-\delta$

$$\hat{L_T} - \min_{i \in \mathscr{E}} L_{T\,i,T} \quad \leq \quad \sqrt{\frac{T}{2}\log\frac{1}{\delta}} + \sqrt{\frac{T}{2}\log N}$$

Recall Hoeffding's inequality $\Rightarrow$
If $X_1, X_2 .... X_N$ are independent random variables with values in [0,1] $\quad \forall \varepsilon > 0$,

$$\mathbb{P}\left[\sum_{i=1}^{n} x_i \geq \frac{1}{n}\sum_{i=1}^{n} EX_i + \varepsilon\right] \leq \exp\left(-2n\varepsilon^2\right)$$

Equivalently, for any $0 < \delta < 1$,

$$\sum_{i=1}^{n} X_i \leq \sum_{i=1}^{n} EX_i + \sqrt{\frac{n}{2}\log\frac{1}{\delta}}$$

with probability $\geq 1-\delta$
Applying Hoffeding to sequence of r.v.s $X_t \equiv l\left(f_{I_t}, y_t\right)$
distribution I_t depends on sequence of previous losses

$$\begin{aligned}
\mathbb{P}\left[I_t = i\right] &= \frac{\exp -\eta \sum_{s=1}^{t-1} l\left(f_{i,t}, y_t\right)}{\sum_{j\in[n]} \exp -\eta \sum_s l\left(f_{i,t}, y_t\right)} \\
\Rightarrow \sum_{t=1}^{T} l\left(f_{i,t}, y_t\right) &< \mathbb{E}\left[\sum_{t=1}^{T} l\left(f_{i,t}, y_t\right)\right] + \sqrt{\frac{T}{2}\log\frac{1}{\delta}}
\end{aligned}$$

□

7.3 MINIMAX REGRET AND LOWER BOUND FOR CONVEX OPTIMIZATION GAME

Regret of an algorithm $\mathscr{A}$, With losses function l ,Set of experts $\mathscr{E}$ and number of rounds T

$$R_T^{\mathscr{A}} = \sup_{y, f_t}\left[\sum_{t=1}^{T} l\left(\hat{p}_t, y_t\right) - \inf_{i\in\mathscr{E}} \sum_t l\left(f_{i,t}, y_t\right)\right]$$

$\mathscr{A}$ is defined as $\forall t \geq 1$
$\hat{p}_t: \left(\left(f_{i,1}\right)_i, y_1, \left(f_{i,2}\right)_i, y_2, \ldots\ldots, y_{t-1}, \left(f_{i,t}\right)_i\right) \rightarrow \mathscr{D}$
We showed the following
When D is convex and l is convex on $\mathscr{D}$,l $\in$ [0,1] , $|\mathscr{E}| = N$

$$R_T{}^{l,\mathscr{E}}\left(EXPWTS\right) \leq \sqrt{\frac{T}{2}\log N}$$

DEFINITION
MINIMAX REGRET $\rightarrow$ For a given set of $\mathscr{D}$,$\mathscr{Y}$,l and $|\mathscr{E}| = N$
We define minimax regret associated to (D, y, l, T, N) as

$$\begin{aligned}
V_T{}^{(N)} &= \inf_{\mathscr{A}} \sup_{|\mathscr{E}|=N} R_T{}^{(l,\mathscr{E})}\mathscr{A} \\
&= \inf_{\mathscr{A}} \sup_{|\mathscr{E}|=N}\left[\hat{L_T} - \min_{i\in\mathscr{E}} L_{i,T}\right]
\end{aligned}$$

$\forall\mathscr{D}$ convex ,l convex and l$\in$ [0,1]
Bound for exp-wts $\Rightarrow$

$$V_T{}^{(N)} \leq \sqrt{\frac{T}{2}\log N}$$

Theorem 7.2. *Consider y =* $\{0,1\}$ *,D = [0,1],* $l\,(p,y) = |p-y|$

$$\sup_{T\geq 1, N\geq 1} \frac{V_T{}^{(N),l,D,y}}{\sqrt{\frac{T}{2}\log N}} \geq 1$$

IMPLICATION :
Given any $0 < \mathscr{E} < 1$, $\quad \exists T \equiv T\,(\mathscr{E})\,, N = N\,(\mathscr{E})$ such that

$$V_T{}^{(N)} \geq (1-\mathscr{E})\sqrt{\frac{T}{2}\log N}$$

Let X be a random variable always in $\mathbb{S}$. Then

$$\sup_{x\in\mathbb{S}} f\,(x) \geq \mathbb{E}\,[f\,(x)]$$

Given T, $|\mathscr{E}| = N$,$\mathbb{A}$, Let's fix the expert advice $[f_{i,t}]_{i,t}$ to some arbitrary sequence.
For the algorithm $\mathbb{A} \rightarrow$

$$R_T{}^f\,(\mathbb{A}) = \sup_{(y_1,y_2,\ldots,y_t)} \left(\sum_{t=1}^{T}|\hat{p}_t - y_t| - \min_{i\in\mathscr{E}}\sum_{t=1}^{T}|f_{i,t}-y_t|\right) \geq \mathbb{E}\left[\left(\sum_{t=1}^{T}|\hat{p}_t - y_t| - \min_{i\in\mathscr{E}}\sum_{t=1}^{T}|f_{i,t}-y_t|\right)\right]$$

$Y_t \sim$ Bernoulli $\left(\frac{1}{2}\right)$ and iid

$$= \mathbb{E}_Y\left[\left(\sum_{t=1}^{T}|\hat{p}_t - y_t| - \mathbb{E}_Y\left[\min_i \sum_{t=1}^{T}|f_{i,t}-y_t|\right)\right]\right]$$

$$= \sum_{t=1}^{T}\mathbb{E}_Y\,[|\hat{p}_t - y_t|] - \mathbb{E}_Y\left[\min_i \sum_t |f_{i,t}-y_t|\right]$$

$$= \frac{T}{2} - \mathbb{E}_Y \left[\min_i \sum_t |f_{i,t} - y_t| \right]$$

$$= \mathbb{E}_Y \left[\sum_{t=1}^{T} \frac{1}{2} - \min_i \sum_t |f_{i,t} - y_t| \right]$$

$$= \mathbb{E}_Y \left[\max_i \sum_{t=1}^{T} \left(\frac{1}{2} - |f_{i,t} - y_t| \right) \right]$$

Note :

$$\frac{1}{2} - |f_{i,t} - y_t| = \frac{1}{2} - f_{i,t} \quad \text{when } y_t = 0$$

$$\frac{1}{2} - |f_{i,t} - y_t| = f_{i,t} - \frac{1}{2} \quad \text{when } y_t = 1$$

So

$$\frac{1}{2} - |f_{i,t} - y_t| = \left(f_{i,t} - \frac{1}{2} \right) \times (2y_t - 1)$$

Let's define $\sigma_t = (2y_t - 1) \in (-1, 1)$

$$R_T^f(\mathbb{A}) \geq \mathbb{E}_\sigma \left[\max_i \sum_{t=1}^{T} \left(f_{i,t} - \frac{1}{2} \right) \sigma_t \right]$$

where σ_t is RADEMACHER random variable.
so ,

$$\sup_{f_{i,t}} R_T^f(\mathbb{A}) \geq \sup_f \mathbb{E}_\sigma \left[\max_i \sum_{t=1}^{T} \left(f_{i,t} - \frac{1}{2} \right) \sigma_t \right]$$

$$\geq \mathbb{E}_f \left[\mathbb{E}_\sigma \left[\max_i \sum_{t=1}^{T} \left(|f_{i,t} - \frac{1}{2}| \right) \sigma_t \right] \right]$$

Since , $f_{i,t} - \frac{1}{2}$ can be either $-\frac{1}{2} or + \frac{1}{2}$

$$= \frac{1}{2}\mathbb{E}_f\left[\mathbb{E}_\sigma\left[\max_i \sum_{t=1}^{T} \mathbb{Z}_{i,t}\sigma_t\right]\right]$$

Where $\{\mathbb{Z}_{i,t}\}_{i,t}$ is iid RADEMACHER $\in (+-1)$ with p $= \frac{1}{2}$

$$= \frac{1}{2}\mathbb{E}_{\{Q_{I,T}\}}\left[\max_{i\in[N]} \sum_{t=1}^{N} \mathbb{Q}_{i,t}\right]$$

Note: $\forall i, \quad \sum_{t=1}^{T} \mathcal{Q}_{i,t}$ is just the position of independent standard random walk started at 0.

$$V_T^{(N)} \geq \frac{1}{2}\mathbb{E}_{\{Q_{I,T}\}}\left[\max_{i\in[N]} \sum_{t=1}^{T} \mathbb{Q}_{i,t}\right]$$

7.3.1 LEMMA

Take $N \times T$ iid RADEMACHER rvs ,$\{\mathbb{Q}\}_{i,t_{i\leq N,t\leq T}}$

$$\lim_{T\to\infty} \mathbb{E}\left[\frac{\max_{i\leq N}\sum_{t=1}^{T}\mathbb{Q}_{i,t}}{\sqrt{T}}\right] = \mathbb{E}\left[\max_{i\leq N} \mathbb{G}_i\right]$$

where $\mathbb{G} \sim \mathbb{N}(0,1)$ and further

$$\lim_{N\to\infty} \frac{\mathbb{E}\left[\max_i \mathbb{G}_i\right]}{\sqrt{2\log N}} = 1$$

So with this lemma , we get

$$\lim_{T\to\infty} \frac{V_T^{(N)}}{\sqrt{T}} \geq \frac{1}{2}\mathbb{E}\left[\max_{i\leq N} \mathbb{G}_i\right]$$

$$\lim_{N\to\infty}\lim_{T\to\infty} \frac{V_T^{(N)}}{\sqrt{\frac{T}{2}\log N}} \geq 1$$

7.4 Bibliography

[1] ”Prediction, Learning and Games” (PLG). Nicolo Cesa-Bianchi and Gabor Lugosi, Cambridge University Press, 2006 → Chapter 5, PLG Thm. 3.7