CS170–Spring 2019 — Homework 12 Solutions

Ran Liao, SID 3034504227

April 21, 2019

Collaborators:Jingyi Xu, Renee Pu

1 Study Group

Name	SID
Ran Liao	3034504227
Jingyi Xu	3032003885
Renee Pu	3032083302

2 Experts Alternatives

(a) The adversary can set $\forall t\ c_1^t$ to be 1, and all other costs to be 0.

$$\begin{aligned}
\max_{C_i^t} R &= \max_{C_i^t} \frac{1}{T}(\sum_{t=1}^{T} c_{i(t)}^t - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t) \\
&= \frac{1}{T}(\sum_{t=1}^{T} c_1^t - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t) \\
&= \frac{1}{T}(\sum_{t=1}^{T} 1 - \sum_{t=1}^{T} 0) \\
&= 1
\end{aligned}$$

(b) For each step t, the adversary can set c_i^t to be 1 where p_i^t is 1, and all other costs to be 0.

In worst case, my strategy is choose experts as evenly as possible.

Therefore, $\min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t = \lfloor \frac{T}{n} \rfloor$.

$$\begin{aligned}
\max_{C_i^t} R &= \max_{C_i^t} \frac{1}{T}(\sum_{t=1}^{T} c_{i(t)}^t - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t) \\
&= \frac{1}{T}(\sum_{t=1}^{T} 1 - \lfloor \frac{T}{n} \rfloor) \\
&= \frac{1}{T}(T - \lfloor \frac{T}{n} \rfloor)
\end{aligned}$$

(c) Suppose $p_{i(min)}$ is the smallest number among p_i. The adversary can set $c_{i(min)}^t = c_{i(min)} = 0$, and all other costs to be 1.

$$\begin{aligned}
\max_{C_i^t} E[R] &= \max_{C_i^t} E[\frac{1}{T}(\sum_{t=1}^{T} c_{i(t)}^t - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t)] \\
&= \max_{C_i^t} \frac{1}{T}(E[\sum_{t=1}^{T} c_{i(t)}^t] - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t) \\
&= \max_{C_i^t} \frac{1}{T}(\sum_{t=1}^{T} \sum_{i=1}^{n} p_i c_i^t - \min_{1\leq i\leq n} \sum_{t=1}^{T} c_i^t) \\
&= \max_{C_i^t}(\sum_{i=1}^{n} p_i c_i - \min_{1\leq i\leq n} c_i) \\
&= (1 - p_{i(min)}) - 0 \\
&= 1 - p_{i(min)}
\end{aligned}$$

A evenly distributed p_i will minimize this regret. i.e. $p_i = \frac{1}{n}$.

3 Follow the regularized leader

(a) Set $A(t,1) = 0$ if t is odd, and $A(t,1) = 1$ otherwise.

Set $A(t,2) = 0$ if t is even, and $A(t,2) = 0.999$ otherwise.

My algorithm will keep alternating between strategy 1 and 2. It will alway choose the 0 payoff. But sticking to strategy 1 will give a payoff of 50. Sticking to strategy 2 will give a payoff of 49.95.

(b) Denote $c_t(i) = \sum_{1\le\tau\le t-1} A(\tau,i)$. Denote $p_t(i_{max})$ to be the largest number among $p_t(i)$.

To maximize $\sum_{i=1}^n (p_t(i)c_t(i))$, $p_t(i_{max})$ will be 1 and all others will be set with 0. Therefore, this algorithm is literally degenerate to the previous algorithm. Because of convexity, a distribution $D(t)$ can always be optimized until it reach a "single point".

(c)

$$
\begin{aligned}
\sum_{i=1}^{n}\left(p_t(i)\cdot\sum_{1\le\tau\le t-1}A[\tau,i]-\eta p_t(i)\ln p_t(i)\right) &= \sum_{i=1}^{n}\left(p_t(i)\cdot\left(\sum_{1\le\tau\le t-1}A[\tau,i]-\eta\ln p_t(i)\right)\right)\\
&= \eta\sum_{i=1}^{n}\left(p_t(i)\cdot\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}-\ln p_t(i)\right)\right)\\
&= \eta\sum_{i=1}^{n}\left(p_t(i)\cdot\ln e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}-\ln p_t(i)\right)}\right)\\
&\le \eta\ln\left(\sum_{i=1}^{n}p_t(i)\cdot e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}-\ln p_t(i)\right)}\right)\\
&= \eta\ln\left(\sum_{i=1}^{n}p_t(i)\cdot\frac{e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}\right)}}{e^{\ln p_t(i)}}\right)\\
&= \eta\ln\left(\sum_{i=1}^{n}p_t(i)\cdot\frac{e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}\right)}}{p_t(i)}\right)\\
&= \eta\ln\left(\sum_{i=1}^{n}e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}\right)}\right)
\end{aligned}
$$

(d) Let $l_i^{(t)} = -A[t,i]$, $(1-\epsilon) = e^{-\frac{1}{\eta}}$ and $w_i^{(0)} = 1$. Denote $A[i] = \sum_{\tau=1}^{t-1} A[\tau,i]$

$$
w_i^{(t)} = w_i^{(t-1)}(1-\epsilon)^{l_i^{(t-1)}} = w_i^{(0)}(1-\epsilon)^{\sum_{\tau=1}^{t-1} l_i^{(\tau)}} = (1-\epsilon)^{-\sum_{\tau=1}^{t-1} A[\tau,i]} = e^{\frac{\sum_{\tau=1}^{t-1} A[\tau,i]}{\eta}} = e^{\frac{A[i]}{\eta}}
$$

Therefore,

$$
p_t(i) = \frac{w_i^{(t)}}{\sum_{j=1}^n w_j^{(t)}} = \frac{e^{\frac{\sum_{\tau=1}^{t-1} A[\tau,i]}{\eta}}}{\sum_{j=1}^n e^{\frac{\sum_{\tau=1}^{t-1} A[\tau,j]}{\eta}}} = \frac{e^{\frac{A[i]}{\eta}}}{\sum_{j=1}^n e^{\frac{A[j]}{\eta}}}
$$

Then,

$$
\begin{aligned}
(1) &= \sum_{i=1}^{n}\left(p_t(i)\cdot\sum_{1\le\tau\le t-1}A[\tau,i]-\eta p_t(i)\ln p_t(i)\right)\\
&= \sum_{i=1}^{n}\left(p_t(i)\cdot\left(A[i]-\eta\ln p_t(i)\right)\right)\\
&= \sum_{i=1}^{n}\left(\frac{e^{\frac{A[i]}{\eta}}}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\cdot\left(A[i]-\eta\ln\frac{e^{\frac{A[i]}{\eta}}}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\right)\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\sum_{i=1}^{n}\left(e^{\frac{A[i]}{\eta}}\cdot\left(A[i]-\eta\ln\frac{e^{\frac{A[i]}{\eta}}}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\right)\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\sum_{i=1}^{n}\left(e^{\frac{A[i]}{\eta}}\cdot\left(A[i]-\eta\ln e^{\frac{A[i]}{\eta}}+\eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\right)\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\sum_{i=1}^{n}\left(e^{\frac{A[i]}{\eta}}\cdot\left(A[i]-\eta\frac{A[i]}{\eta}+\eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\right)\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\sum_{i=1}^{n}\left(e^{\frac{A[i]}{\eta}}\cdot\left(A[i]-A[i]+\eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\right)\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\sum_{i=1}^{n}\left(e^{\frac{A[i]}{\eta}}\cdot\eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\right)\\
&= \frac{1}{\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}}\cdot\sum_{i=1}^{n}e^{\frac{A[i]}{\eta}}\cdot\eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\\
&= \eta\ln\sum_{j=1}^{n}e^{\frac{A[j]}{\eta}}\\
&= \eta\ln\left(\sum_{i=1}^{n}e^{\left(\sum_{1\le\tau\le t-1}\frac{A[\tau,i]}{\eta}\right)}\right)\\
&= (2)
\end{aligned}
$$

Therefore, $\epsilon = 1-e^{-\frac{1}{\eta}}$.

4 Linear Classifiers using Multiplicative Weights

(a) Denote $X(i) = \sum_{t=1}^{T}(l_t[i] + \epsilon|l_t[i]|)$. And denote $i^* = \arg\min_{i=1}^{n} X(i)$.

$$\begin{aligned}
\sum_{t=1}^{T}\langle l_t, x_t\rangle &\leq \frac{\log(n)}{\epsilon} + \min_{i=1}^{n}\sum_{t=1}^{T}(l_t[i] + \epsilon|l_t[i]|) \\
&= \frac{\log(n)}{\epsilon} + X(i^*) \\
&= \frac{\log(n)}{\epsilon} + p^*[i^*]X(i^*) + (1 - p^*[i^*])X(i^*) \\
&\leq \frac{\log(n)}{\epsilon} + p^*[i^*]X(i^*) + \sum_{i\neq i^*} p^*[i]X(i) \\
&= \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} p^*[i]X(i) \\
&= \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} p^*[i](\sum_{t=1}^{T}(l_t[i] + \epsilon|l_t[i]|))
\end{aligned}$$

(b) The experts are the n features in a.

(c) $l(t) = -\sum_{i=1}^{m} a_i$

(d) The algorithm stops when $\langle a_i, x\rangle \geq 0$ for all $1 \leq i \leq m$.

(e) Denote $A = \sum_{i=1}^{m} a_i$.

$$w_i^{(t)} = w_i^{(t-1)}(1-\epsilon)^{l_i^{(t-1)}} = w_i^{(0)}(1-\epsilon)^{\sum_{\tau=1}^{t-1} l_i^{(\tau)}} = (1-\epsilon)^{-\sum_{\tau=1}^{t-1} A_i} = (1-\epsilon)^{(1-t)A_i}$$

Therefore,

$$\widetilde{x}_i = \frac{w_i^{(t)}}{\sum_{j=1}^{n} w_j^{(t)}} = \frac{(1-\epsilon)^{(1-t)A_i}}{\sum_{j=1}^{n}(1-\epsilon)^{(1-t)A_j}} = \frac{(1-\frac{1}{2}\eta)^{(1-t)A_i}}{\sum_{j=1}^{n}(1-\frac{1}{2}\eta)^{(1-t)A_j}}$$

See next page,

$$
\begin{aligned}
\sum_{t=1}^{T}\langle l_t, x_t\rangle &\le \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} p^*[i](\sum_{t=1}^{T}(l_t[i] + \epsilon|l_t[i]|)) \\
\sum_{t=1}^{T}\langle -A, x_t\rangle &\le \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i(\sum_{t=1}^{T}(-A_i + \epsilon|-A_i|)) \\
\sum_{t=1}^{T}\langle -A, x_t\rangle &\le \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i(\sum_{t=1}^{T}(-A_i + \epsilon)) \\
-\sum_{t=1}^{T}\langle A, x_t\rangle &\le \frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i(\sum_{t=1}^{T}(-A_i + \epsilon)) \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i(\sum_{t=1}^{T}(A_i - \epsilon)) \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i \sum_{t=1}^{T} A_i - \sum_{i=1}^{n} \widetilde{x}_i \sum_{t=1}^{T} \epsilon \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{\log(n)}{\epsilon} + \sum_{i=1}^{n} \widetilde{x}_i \sum_{t=1}^{T} A_i - T\epsilon \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{\log(n)}{\epsilon} + T\langle A, \widetilde{x}_i\rangle - T\epsilon \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{\log(n)}{\epsilon} + Tm\eta - T\epsilon \\
\sum_{t=1}^{T}\langle A, x_t\rangle &\ge -\frac{2\log(n)}{\eta} + Tm\eta - \frac{T\eta}{2}
\end{aligned}
$$

Let the right hand side be 0.

$$
\begin{aligned}
-\frac{2\log(n)}{\eta} + Tm\eta - \frac{T\eta}{2} &= 0 \\
Tm\eta - \frac{T\eta}{2} &= \frac{2\log(n)}{\eta} \\
T(m\eta - \frac{\eta}{2}) &= \frac{2\log(n)}{\eta} \\
T &= \frac{2\log(n)}{\eta^2(m - \frac{1}{2})}
\end{aligned}
$$