Assignment-1

Online Learning(IE-613)

Vinay Chourasiya (173190011)

Question-1

- **Weighted Majority Algorithm**
- Program file name - *Qus1.py*
- Plot of Pseudo Regret vs Learning Rate for $T = 10^5$ over 20 Sample paths. and value of $c[0.1, 2.1]$

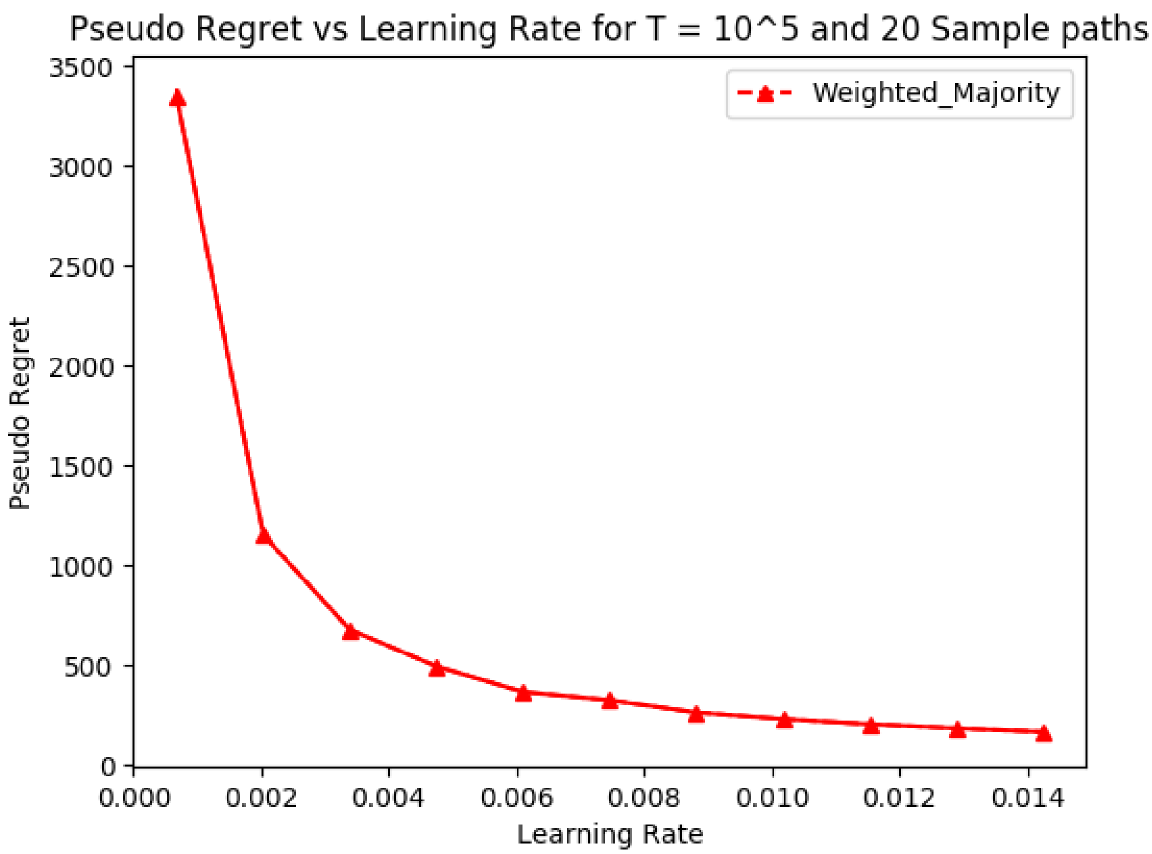

Figure 1: Pseudo Regret vs Learning Rate

Question-2

Exp3

Parameter for Exp3 algorithm is

$$\eta = c\sqrt{\frac{2\log K}{KT}}$$

- Program file name - *Q2exp3.py*
- Plot of Pseudo Regret vs Learning Rate for $T = 10^5$ over 50 Sample paths and $c[0.1, 2.1]$

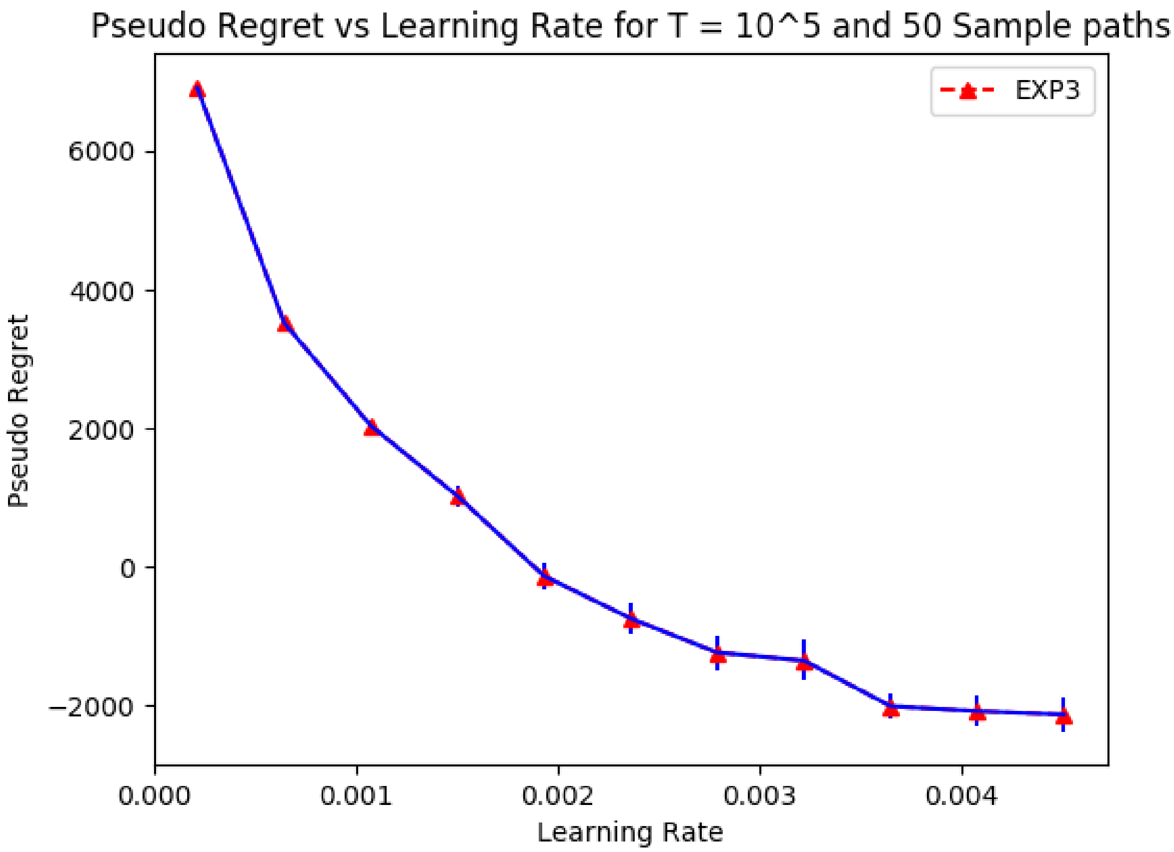

Figure 2: Pseudo Regret vs Learning Rate

Exp3.p

Parameter for Exp3 algorithm is

$$\eta = c * \sqrt{\frac{2 * \log K}{KT}}, \beta = \eta, \gamma = K\eta$$

- Program file name - *Q2exp3p.py*
- Plot of Pseudo Regret vs Learning Rate for $T = 10^5$ over 50 Sample paths and $c[0.1, 2.1]$

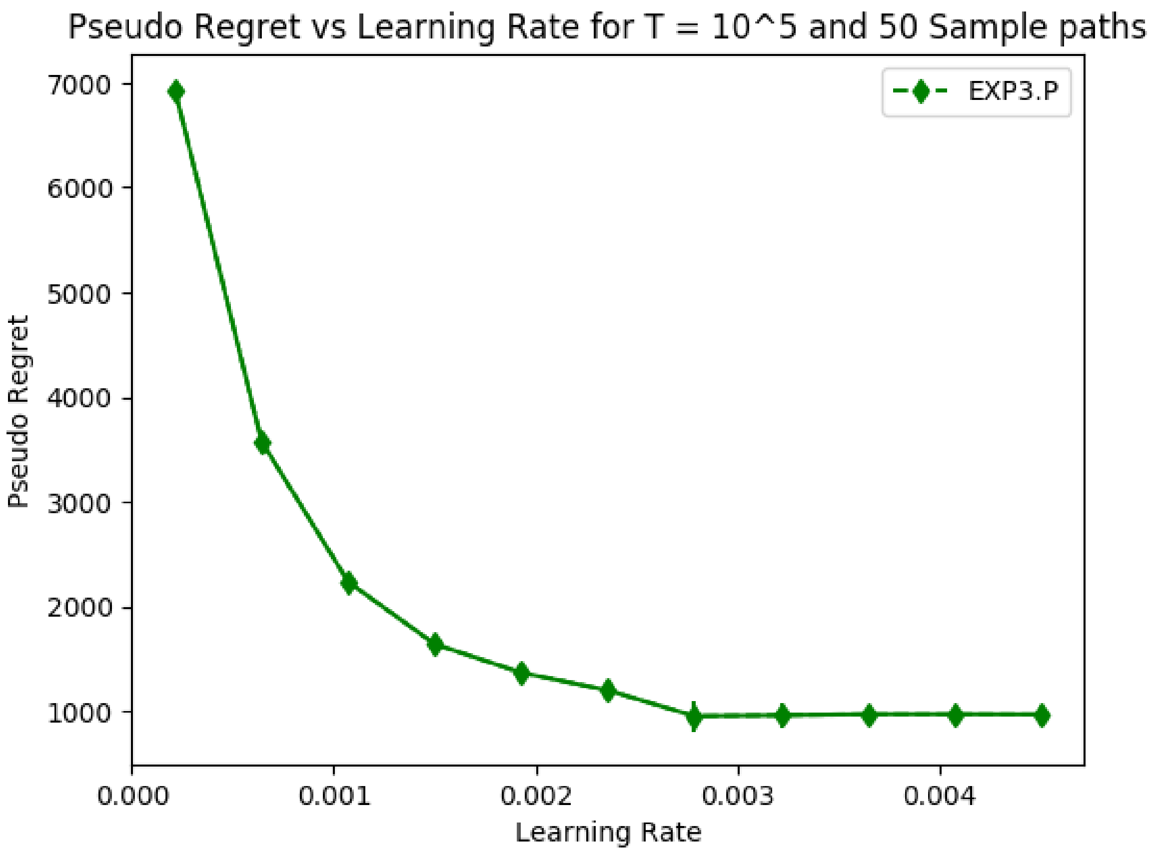

Figure 3: Pseudo Regret vs Learning Rate

Exp3IX

Parameter for Exp3 algorithm is

$$\eta = c * \sqrt{\frac{2 * \log K}{KT}}, \gamma = \eta/2$$

- Program file name - *Q2exp3ix.py*
- Plot of Pseudo Regret vs Learning Rate for $T = 10^5$ over 50 Sample paths and $c[0.1, 2.1]$

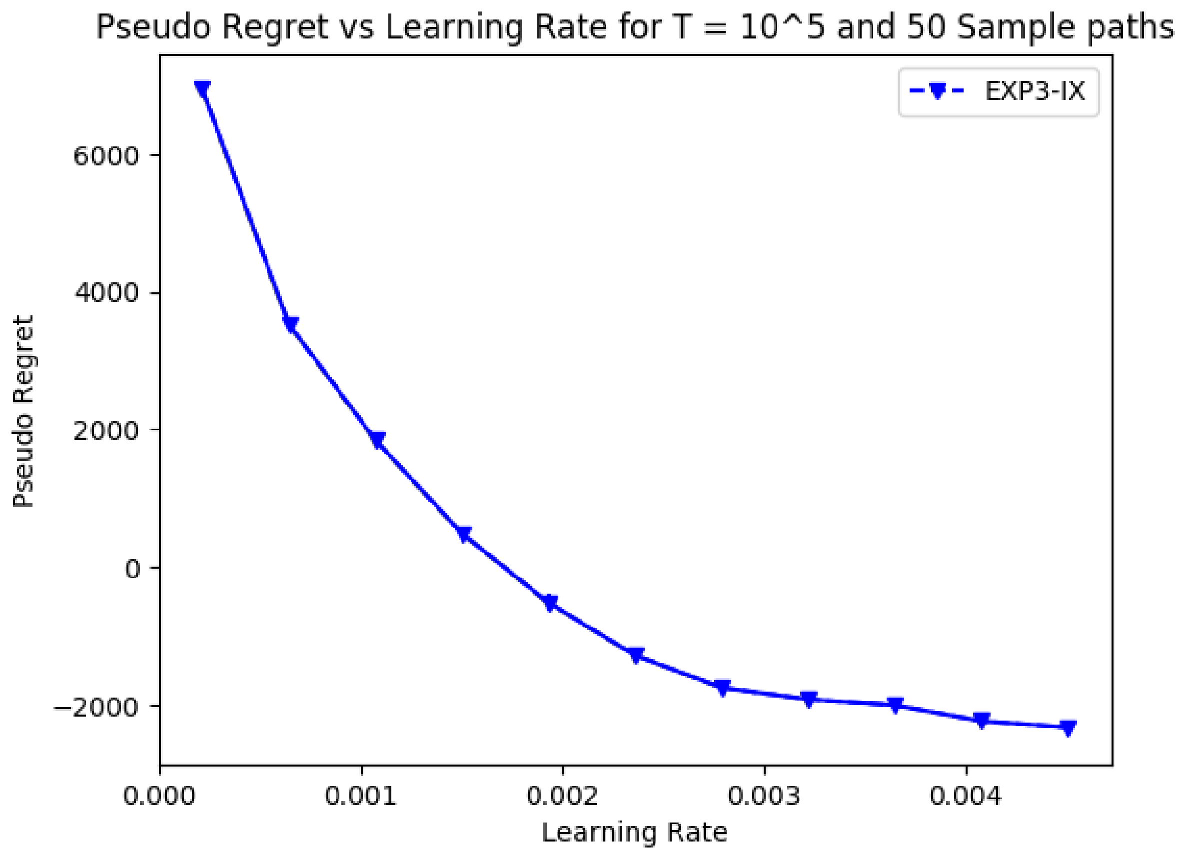

Figure 4: Pseudo Regret vs Learning Rate

Question-3

Performance of *Exp3-IX* is batter than *Exp3 and Exp3.P*

1) When We compare Figure2 and Figure4,this is clear, *Exp3-IX* performance batter than *Exp3.P*.(as the learning rate increase, regret for *Exp3-IX* decrease below zero and achieve negatives values, but in case of *Exp3.p* regret still positive for all learning rate.). thus the empirical performance of *Exp3-IX* batter.

2) The performance level of *Exp3-IX* (in terms of pseudo regret) and *Exp3* Nearly similar but variation of regret (over different values of learning rate) more for *Exp3* compare to *Exp3-IX*.Thus this conclude *Exp3-IX* is more robust than the *Exp3*

3) *implicit exploration* (IX) strategy of *Exp3-IX* algorithm make its performance batter and more robust than other two.

Question-4

Given that hypothesis class H under the realizability assumption. i.e. there exist h^* such that

$$h^*(x_t) = y_t$$

and we observe $z_t = y_t + v_t$, where $P(v_t = 1) \leq \gamma, \gamma = [0, 1/2)$

The following table show different values z_t with respect to y_t and v_t

y_t	v_t	z_t
1	1	0
1	0	1
0	1	1
0	0	0

from table this is clear when noise $v_t = 0$ we observe same as true level, also we know $P(v_t = 1) \leq \gamma$, thus under realizability assumption we always have one hypothesis h^* in our hypothesis class that always predict true level with probability at least $1 - \gamma$
$P(h^*(x_t) = y_t) \leq 1 - \gamma$.

Our Goal is to identify the h^* and other hypothesis in our class with appropriate weights thus we have minimum regret and good prediction over the noise v_t

Theorem1

For any $\gamma[0; 1 = 2)$, if we run $Algorithm - 1$ (see blow) with learning rate $\eta = \frac{1}{2}\log(\frac{1-\gamma}{\gamma})$ with respect to a set of experts $(f_1, ...f_N)$,then if for some $i \in (f_1, ...f_N)$, the labels yt (randomly generated by the environment) are such that $Pr(y_t \neq f_i^t) \leq \gamma$ for all t then

$$E\{\sum_{t=1}^{T} \mid \hat{y}_t - f_i^t \mid\} \leq \frac{\log(N)}{1 - 2\sqrt{\gamma(1-\gamma)}}$$

input: Number of experts N ; Learning rate $\eta > 0$
initialize: $\mathrm{w}^0 = (1, \ldots, 1) \in \mathbb{R}^N$; $Z_0 = N$
for $t = 1, 2, \ldots, n$
receive expert advice $(f_1^t, f_2^t, \ldots, f_N^t) \in \{0, 1\}^N$
environment determine y_t without revealing it to learner
define $\hat{p}_t = \frac{1}{Z_{t-1}} \sum_{i: f_i^t = 1} w_i^{t-1}$
predict $\hat{y}_t = 1$ with probability $\hat{p}_t$
receive label y_t
update: $w_i^t = w_i^{t-1} \exp(-\eta |f_i^t - y_t|)$; $Z_t = \sum_{i=1}^N w_i^t$

Question-5

The regret of A on each period of 2^m rounds is at most $\alpha\sqrt{2^m}$ Therefore, the total regret is at most

$$\sum_{m=1}^{\lceil \log_2 T \rceil} \alpha\sqrt{2^m} = \alpha \sum_{m=1}^{\lceil \log_2 T \rceil} (\sqrt{2})^m$$

Using formula of geometric progression, we get

$$= \alpha\sqrt{2}\left(\frac{1 - (\sqrt{2})^{\lceil \log_2 T \rceil - 1}}{1 - \sqrt{2}}\right)$$

$$= \alpha\frac{\sqrt{2}}{\sqrt{2} - 1}\left(\frac{\sqrt{T} - \sqrt{2}}{\sqrt{2}}\right)$$

$$= \alpha\frac{\sqrt{2}}{\sqrt{2} - 1}\left(\sqrt{\frac{T}{2}} - 1\right)$$

$$(since \sqrt{\frac{T}{2}} - 1 \leq \sqrt{T})$$

$$\leq \frac{\sqrt{2}}{\sqrt{2} - 1}\alpha\sqrt{T}$$

Hence proved.

Reference

1) Ben-David, Shai, Dávid Pál, and Shai Shalev-Shwartz. "Agnostic online learning." (2009).

2) Neu, Gergely. "Explore no more: Improved high-probability regret bounds for non-stochastic bandits." Advances in Neural Information Processing Systems. 2015.