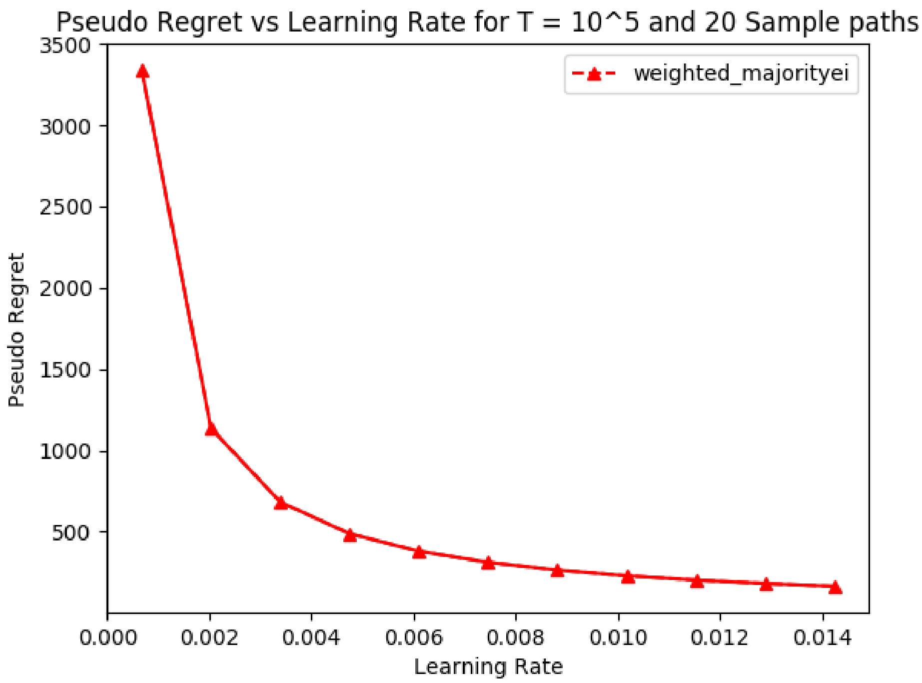
IE613: Assignment 1

Sumrit Gupta - 170040044

February 10, 2019

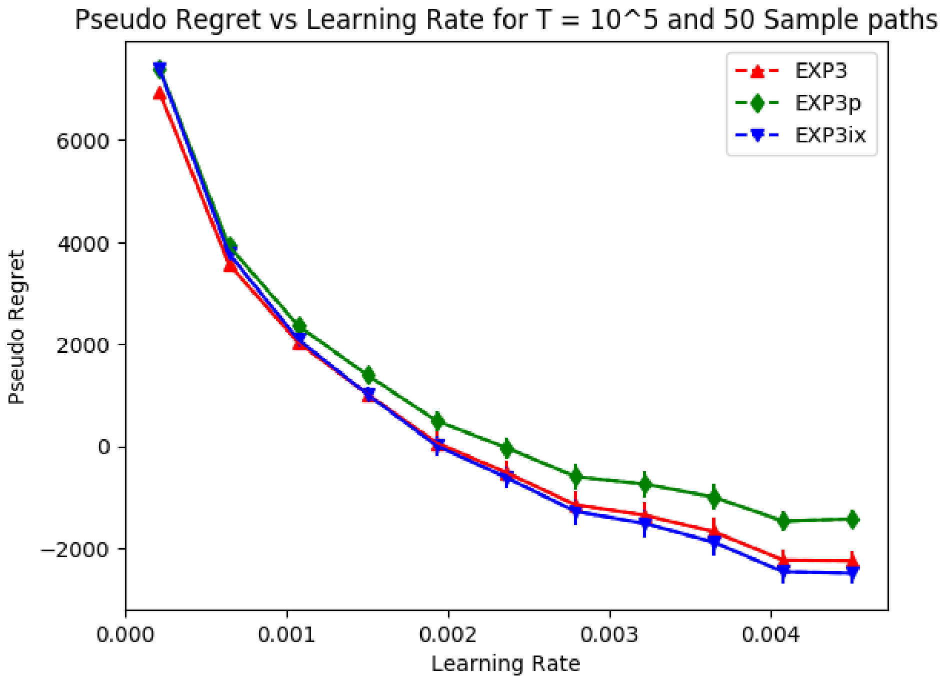
1 Solutions

Ans 1- Weighted Majority Algorithm



As we can see that the regret values decrease with increase in the value of learning rate. It varies from values close to 3400 to as low as 200. When the value of eta is low, the reduction in weight is not much, so the dot product of loss and weight is significant but as eta value is high the update in weight cause much reduction due to which regret is low.

Ans 2- Exp3 Algorithm



From the plots, we can say that as the value of eta is low, we are more exploring, so the best arm is not selected most of the times, hence the regret is high. As the value of eta increases there is a balance between exploration and exploitation, thus regret value improves. All the three algorithms perform better with increasing value of eta, however if the value of eta is taken very high (factor of 10 or 100), then there will be more exploitation and performance would be poor.

Ans 3- The Algorithm that performs best

EXP-IX performs better than the other algorithms as due to the addition of the term gamma to Pti in the denominator makes it less likely to choose the bad arms having low Pti again, thus, exploitation would be more and the estimate value wont get arbitrarily high probabilities.

Ans - 4. Suppose π is deterministic policy

$\Rightarrow$ $$I_t = f\left(x_{1A_1}, x_{2A_2} \ldots\ldots x_{t-1\,A_{t-1}}\right)$$

Define x_{ti}, for $t \in [1, 2 \ldots n]$ and $i \in [1, 2 \ldots K]$

$$x_{ti} = \begin{cases} 0 & \text{if } A_t = i \\ 1 & \text{else} \end{cases}$$

As policy will collect no reward (or 0 reward)

Yet, $$\max_{i \in [K]} \sum_{t=1}^{n} x_{ti} \geq \frac{1}{K} \sum_{t=1}^{n} \sum_{i=1}^{K} x_{ti} = \frac{n(K-1)}{K}$$

So, the reward is atleast $n\left(1 - \frac{1}{K}\right)$

Hence, for T rounds and K arms,

$$R_T(\pi, \nu) \geq T\left(1 - \frac{1}{K}\right)$$

Ans-5. $R_T(\Pi, \nu) = E\left[\sum_{t=1}^{T} \max_i x_{t,i} - \sum_{t=1}^{T} x_{t I_t}\right]$

Let $I_t \sim P_t$ where P_t is the prob. of each arm being chosen.

Let ν gives rewards $x_{t,i} = \begin{cases} 1 & \text{if } P_{t,i} = \min_i P_{t,i} \\ 0 & \end{cases}$

Then $\max_i x_{t,i} = 1 \quad \forall\ t \in [T]$

$$R_T(\Pi, \nu) = E\left[\sum_{t=1}^{T} 1 - \sum x_{t,I_t}\right]$$

$$= T - E\left[\sum_{t=1}^{T} x_{t,I_t}\right]$$

$$x_{t,I_t} = \mathbb{1}_{\{I_t = \arg\min_i P_{t,i}\}}$$

$$R_T(\Pi, \nu) = T - E\left[\sum \mathbb{1}_{\{I_t = \arg\min_i P_{t,i}\}}\right]$$

$$= T - \sum_{t=1}^{T} \min P_{t,i}$$

Now, $\sum_{t=1}^{K} P_{t,i} = 1 \Rightarrow \min P_{t,i} \leq 1/K \quad \forall\ t \in [T]$

$\Rightarrow \sum_{t=1}^{T} \min P_{t,i} \leq T/K$

$$\boxed{R_T(\Pi, \nu) \geq T\left(1 - 1/K\right)}$$

Ans-6. $E\left[\hat{x}(A, x_A)\right] = \sum_{i=1} P_i \hat{x}(i, x_i) = x_1$

$\hat{x} : [k] \times R \rightarrow R, \quad \forall \; x \in R^K$, and $A \sim p$

Let $X_1, X_2 \ldots X_M \in R^K$

let $\hat{x}(1, X_1) = \hat{x}(1, X_2) \ldots = \hat{x}(1, X_m) = \lambda$

So, $\sum_{i=2}^{k} P_i \hat{x}(i, X_{1i}) = x_1 - \lambda$

$(M > K)$

$\sum_{i=2}^{k} P_i \hat{x}(i, X_{2i}) = x_1 - \lambda$

$\vdots$

$\sum_{i=2}^{k} P_i \hat{x}(i, X_{Mi}) = x_1 - \lambda$

Since, $M > K \Rightarrow$ for consistency of eqns

$\hat{x}(i, X_1) = \hat{x}(i, X_2) \ldots = \hat{x}(i, X_M) = a_i$ (say)

$\forall \; i \in [2, 3 \ldots k]$

So, $\hat{x}(1, X_1) = \frac{1}{P_1}\left(x_1 - \sum_{i=2}^{k} a_i P_i\right)$

Let $a_1 = \frac{-1}{P_1}\left(\sum_{i=2}^{k} a_i P_i\right)$, $\left[\because \text{Given } \sum_{i=1}^{k} a_i P_i = 0\right]$

Then, $\hat{x}(1, X_1) = \frac{1}{P_1}(x_1 + a_1 P_1) = a_1 + \frac{x_1}{P_1}$

As $X_1, X_2 \ldots X_M$ were chosen randomly

so $\forall \; x \in R^K$

$\hat{x}(1, x) = a_1 + \frac{x}{P_1}$

and $\hat{x}(i, x) = a_i \quad (i \neq 1)$

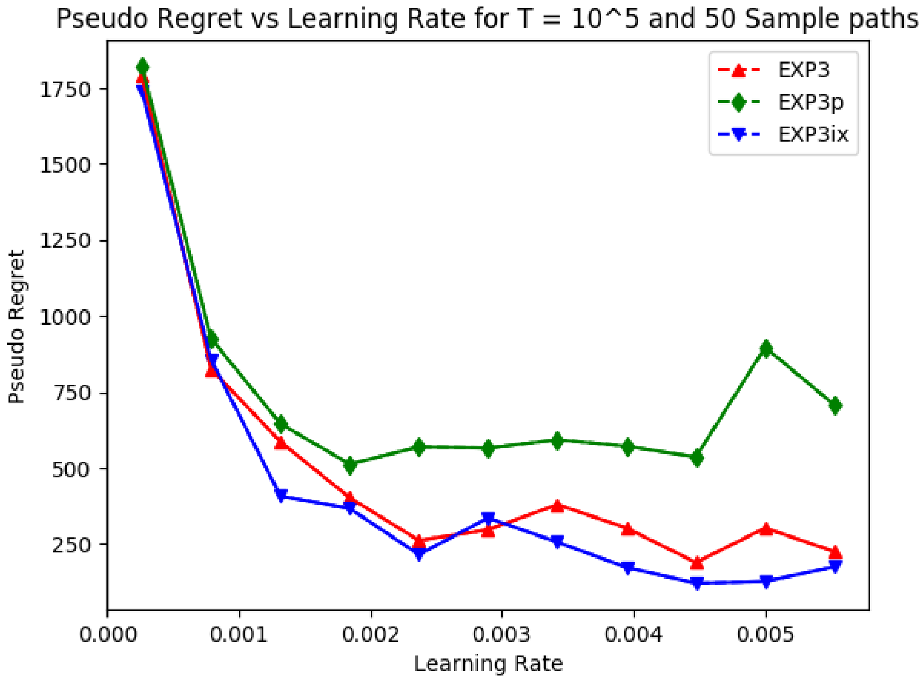
$\Rightarrow \hat{x}(i, x) = a_i + \left(\mathbb{1}_{\{i=1\}} x_1\right) / P_1 \quad \forall \; i \in [k] \; \& \; x \in R^k$

There exists $a \in \mathbb{R}^K$ s.t. $\sum a_i P_i = 0$

and $\boxed{\hat{x}(i, x) = a_i + \frac{\mathbb{1}_{\{i=1\}} x_1}{P_1}}$

Ans 7

The implementation for $\mu_1 = 0.5$ and $\mu_2 = 0.55$ of Exp3 gives the following plot of regret. Again Exp-IX performs the best and Exp3p shows slightly poorer results.



The implementation for $x_{t1} = \mathrm{I}(t <= T/4)$ and $x_{t2} = \mathrm{I}(t > T/4)$ of Exp3 gives the following plot of regret which shows that the regret remains almost same for different values of eta as we are fixing the observations x_{t1} and x_{t2}.