IE 613

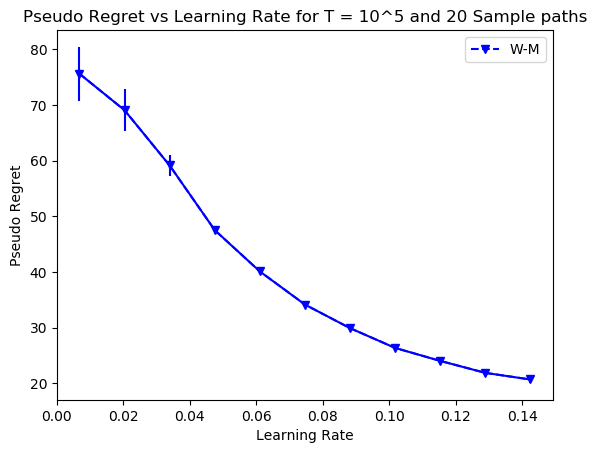
Online Learning Assignment - 1

Roll- 150100089

Name- Gampa Varun

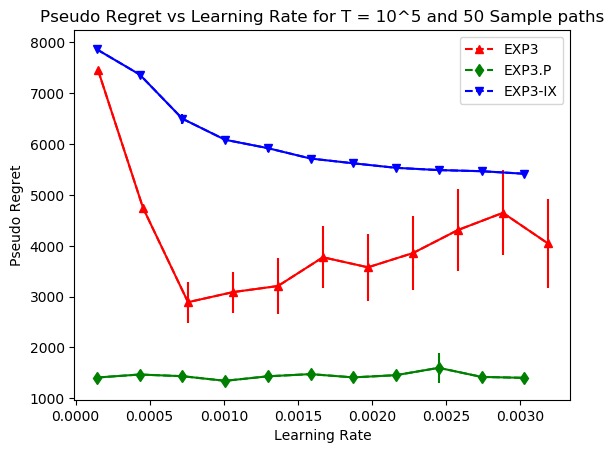
A1) Graphs of Pseudo regret of weighted-majority algorithm against the learning rate eta.

It is observed that regret is decreasing with increasing learning rates but is tending to a constant.



A2)

Graphs of Pseudo regret of EXP3, EXP3P, and EXP3IX against the learning parameter eta.



A3) The environment is such that earlier the ninth arm is better for generating rewards but after T = 50000, 10th arm is the best. EXP3P performed the best of all the three. It may have performed better because of the explicit exploration enforce upon it, as a result of that it is able to capture the variation in environment and as a result choose the tenth arm after T = 50000 whereas the other two algorithms couldn’t.

A7) The arm which generates rewards with u = 0.55 will be chosen more as the time progresses in the first case. In the second case if the exploration aspect is high enough then the algorithm will shft from arm 1 to arm 2 sometime after t = 25000.