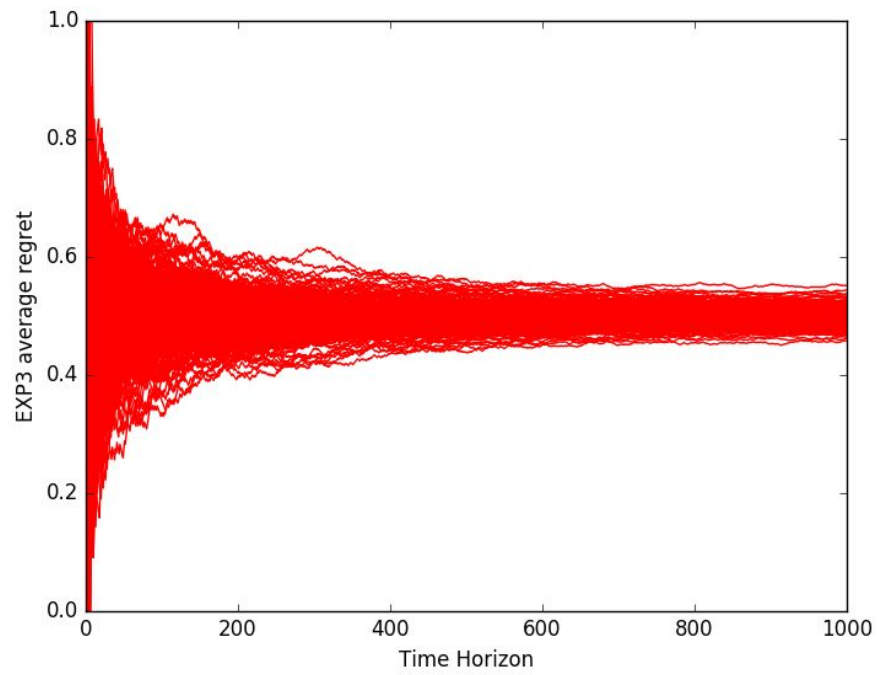
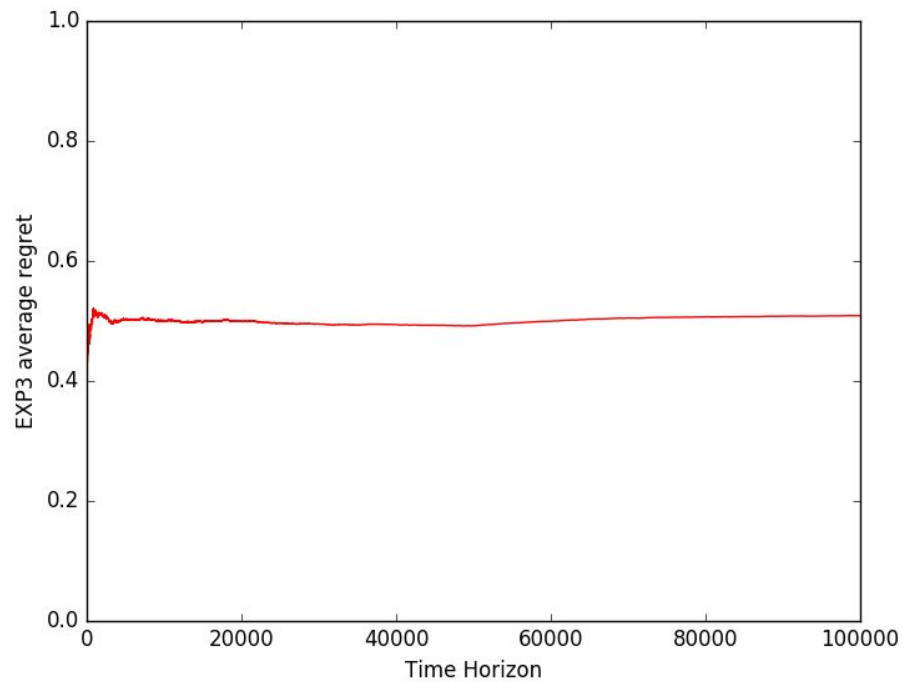
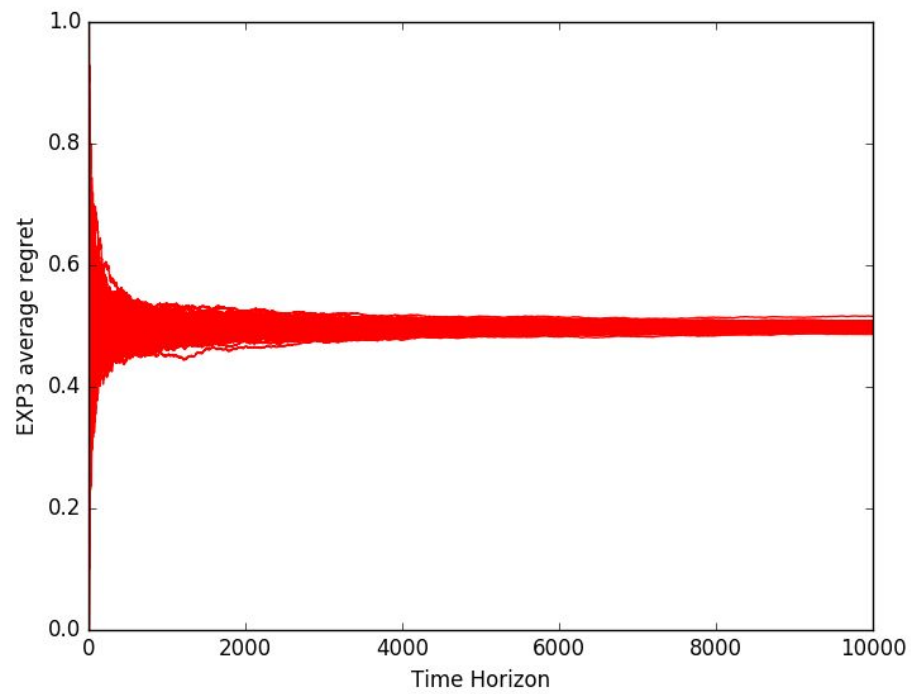


Assignment-2

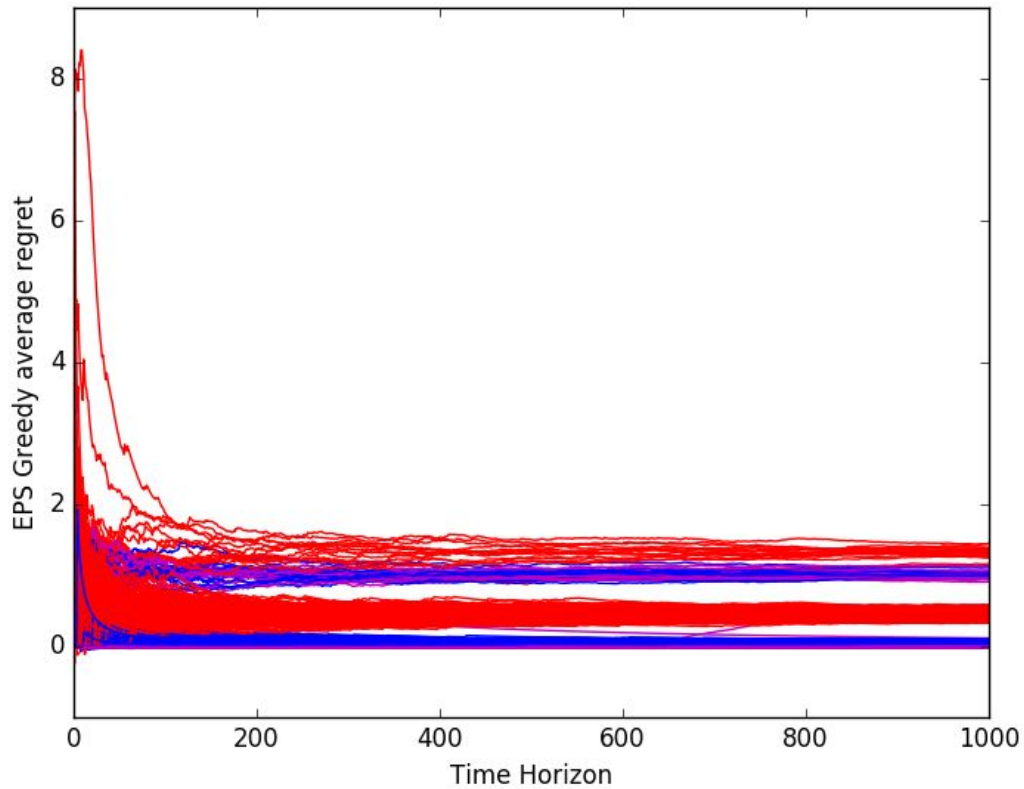
Exp3 Algorithm:





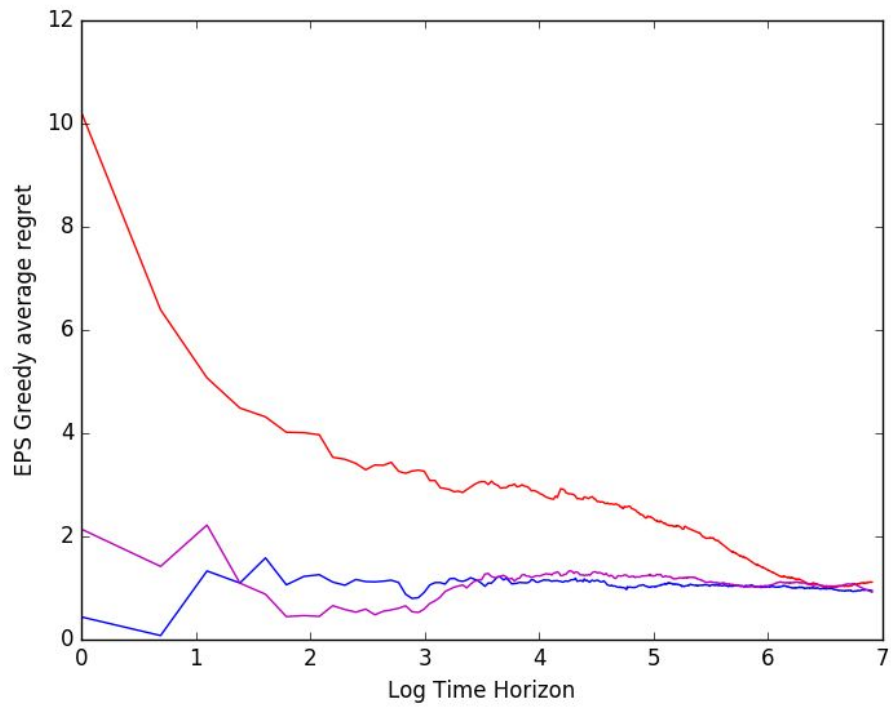
We can see that as the number of episodes increases, the average regret reaches a constant value

Epsilon-Greedy Algorithm:

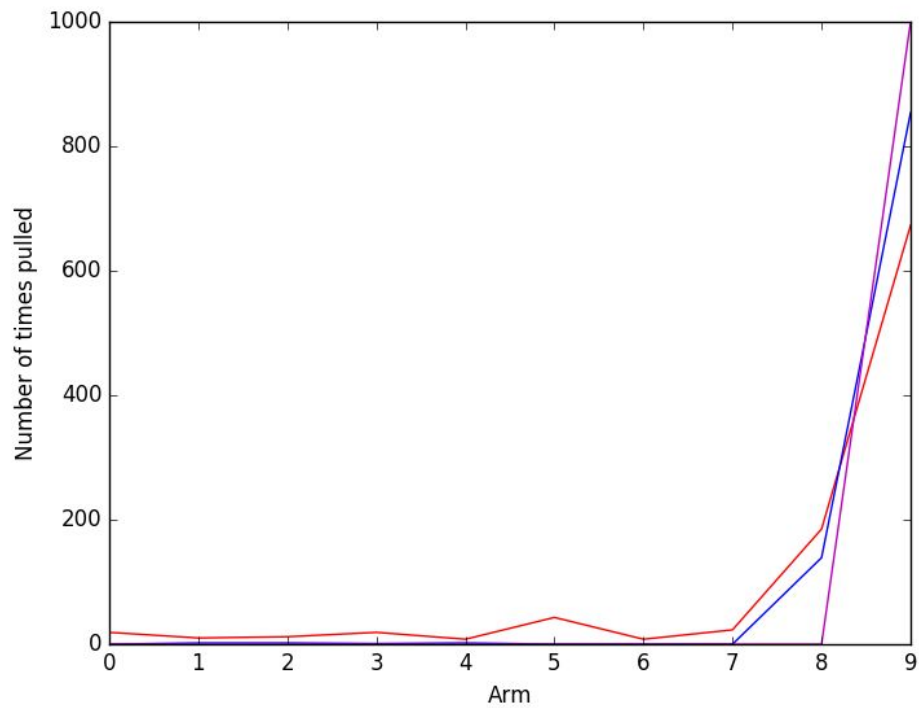


We can see that as the value of epsilon decreases and reaches zero, the algorithm converges to the stationary point faster.

As the number of episodes increases the regret comes closer to the stationary point



The convergence is faster with $\log(\text{time horizon})$



As we can see the arm which gives the maximum reward is pulled the most number of times, but the number of times this arm is pulled is more for the epsilon value which is closer to zero. As the value of epsilon increases the algorithm explores the other arms more frequently

Upper-Confidence-Bound Algorithm

