

IE613 Online Machine Learning
Assignment 2

Vivek Pandit 130010009



Department of Mechanical Engineering

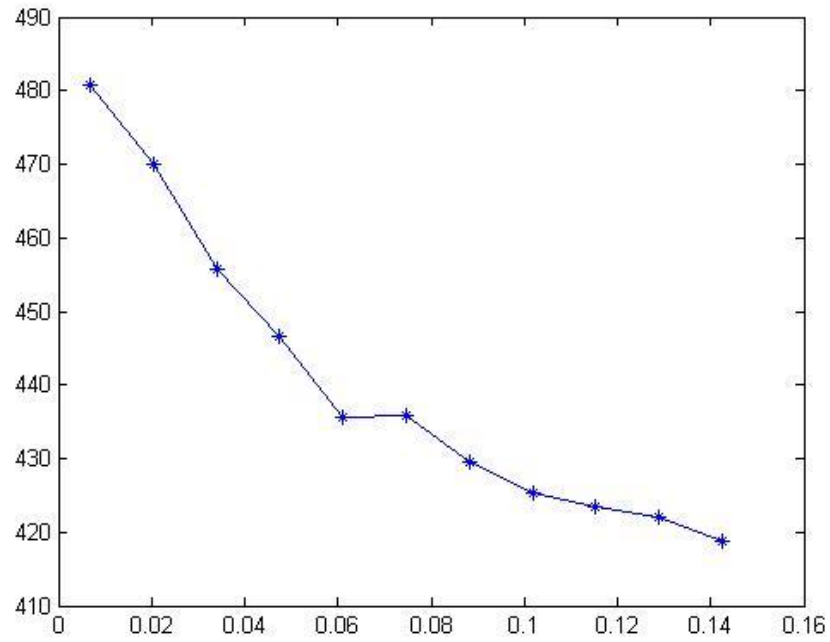
INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

March, 2017

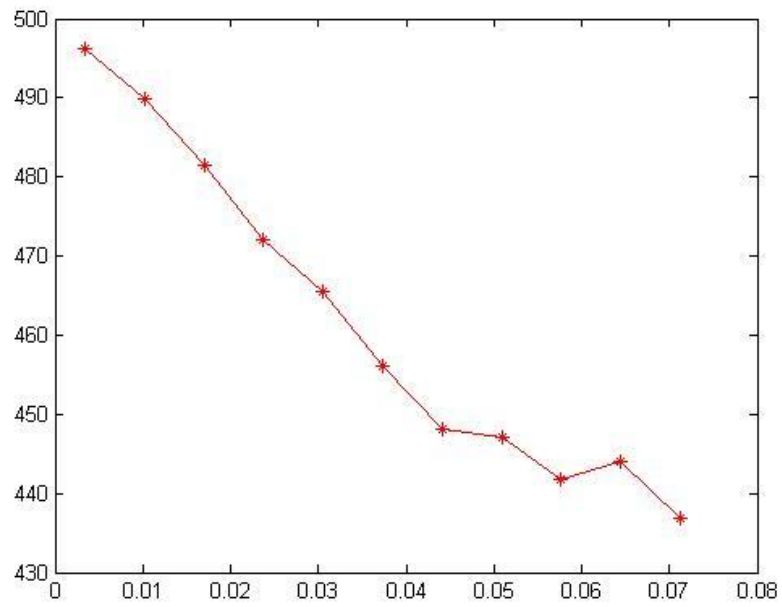
Q1

All plots are Pseudo_regret vs eta

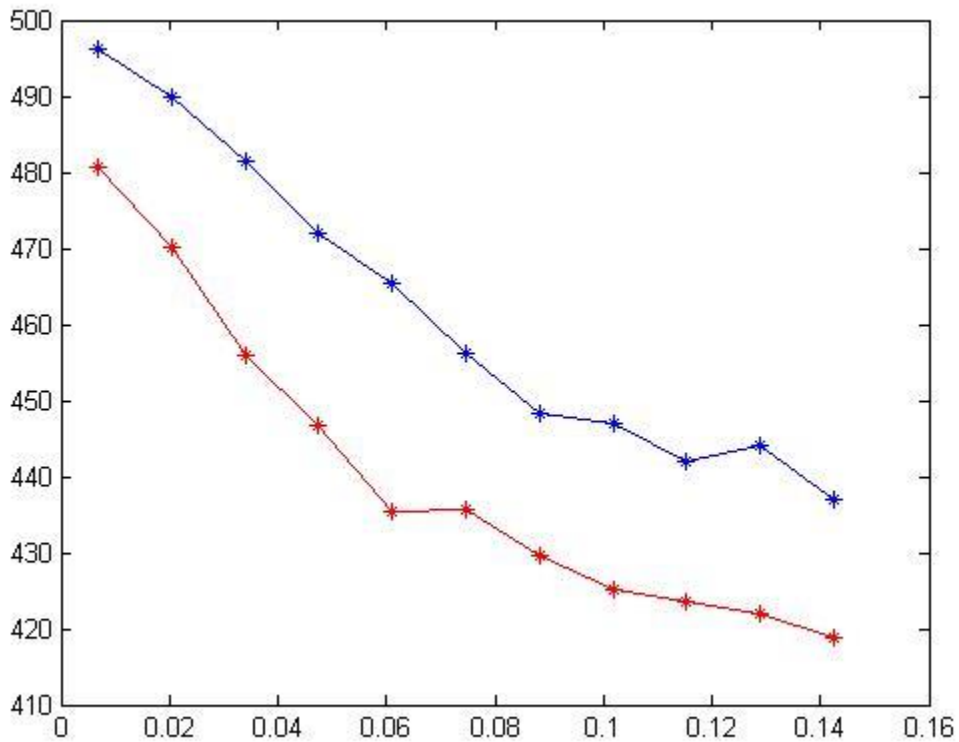
Plot for W



Plot for FTRL



Plot for both algorithms together



Blue line is the FTRL algorithm

Red line is the WM algorithm

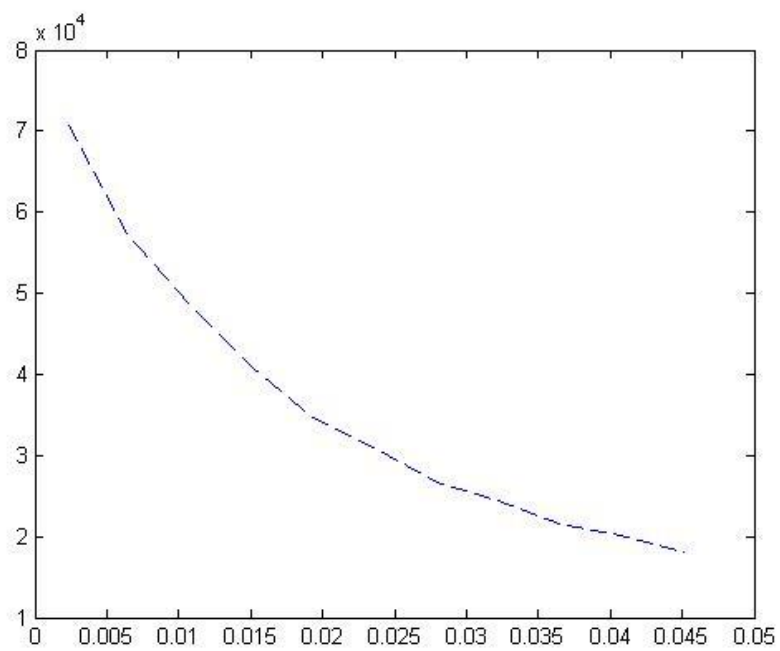
Value of $\eta^* = \sqrt{\log(d) / (2 \cdot T)}$

Clearly WM algorithm is better compared to FTRL algorithm. This is because FTRL requires minimization of an objective function, which translates to Randomised weighted majority. The weights are given out randomly and hence may not be a useful indicator compared to WM algorithm.

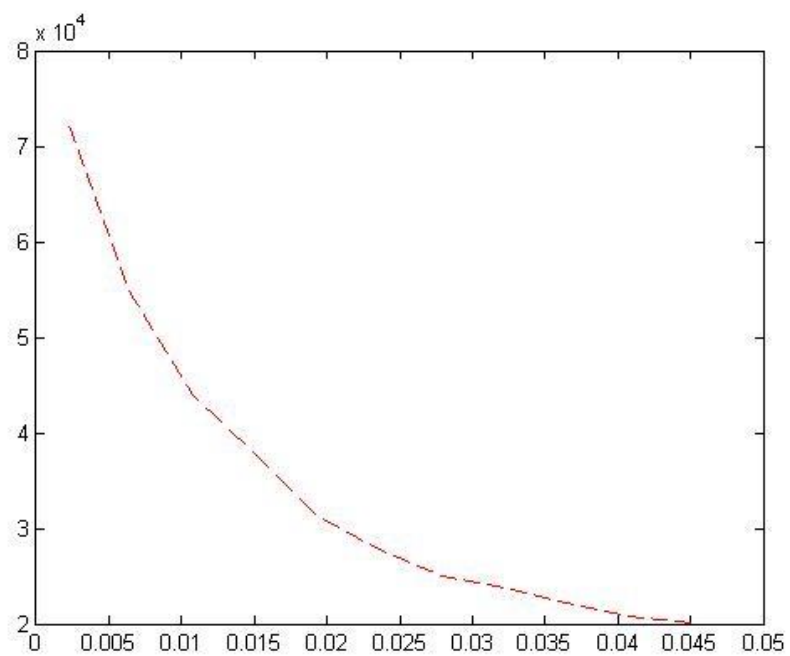
Q2.

All plots are for pseudo regret vs eta values

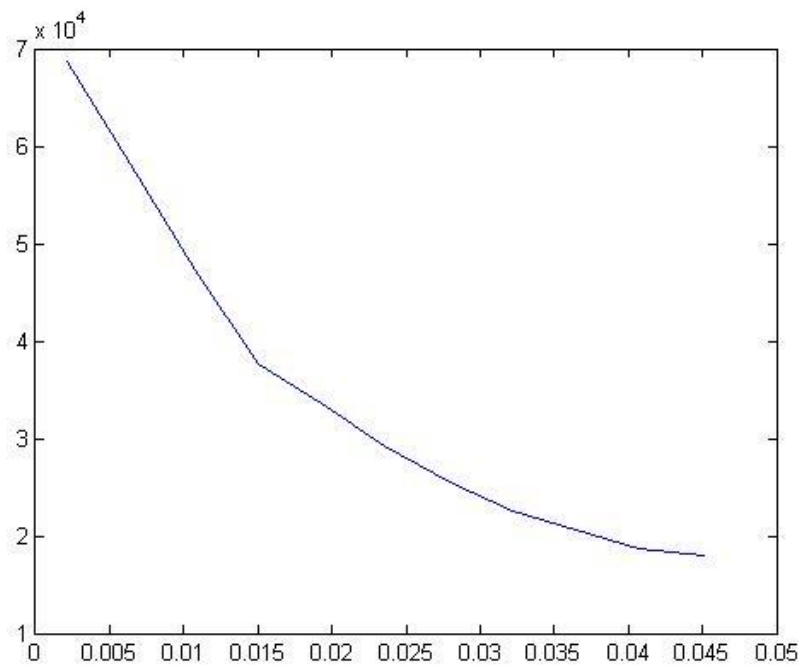
Plot for Exp3



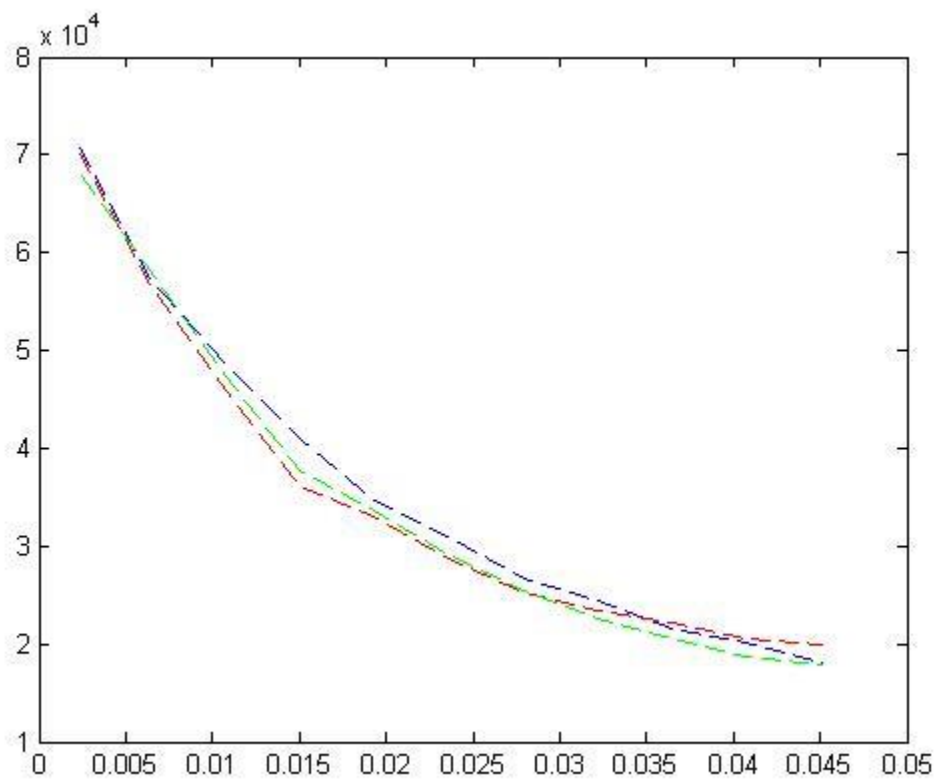
Plot for Exp3p



Plot for Exp_IX



Plot for all 3 algorithms



Blue dashed lines are for exp3 algorithm

Green dashed lines are for expIX algorithm

Red Dashed lines are for exp3p algorithm

Q3.

As learning rate increases, pseudo regret decreases for all the algorithms, as can be seen from the above plots.

At the region of lower learning rates, Exp3P algorithm performs the best as it has the lowest pseudo regret. This is because there are fluctuations of the loss estimates around the true losses, which are too large to permit bounding the true regret with high probability. Exp3P keeps these fluctuations under control by the use of beta.

As learning rate increases, ExpIX has the lowest pseudo regret.

One algorithm that I believe will outperform all the other algorithms is by having some prior distribution to the value of probability on which the arms have been sampled. If we assume a beta distribution for the value of w , and update the beta distribution depending on the reward gained, it will increase probability of those arms which are giving high rewards or low regrets.

Beta distribution has two parameters, alpha and beta.

Pseudo Code

Assume a beta distribution for all arms. Initial values for alpha and beta.

Sample a point from this distribution for each arm.

Pick the arm with the maximum value of this sampled point.

Take the loss and update alpha and beta of this arm such that its beta distribution becomes wider with small regrets i.e alpha and beta have some inverse relationship with regret.

Repeat over T rounds.

I haven't written the code for this. Hence cannot from a plot of this.