**RL Lab Assignment - 2**
**CS 414: Reinforcement Learning Lab**
**Course Instructor** : Prof. Prabuchandran
**Teaching Assistant** : Shruthi K.
**Date**: 20 Jan 2021

**INSTRUCTIONS**: You have to give clear and detailed plots and solution to each of the questions. **Submit one single pdf file containing solutions to all problems in google class room before 31st Jan, 11.30 am before the RL class. Submit the code as .py file or as python notebook along with the solution pdf. Only one member of the group has to submit the assignment. Name your pdf with** $rollno1\_rollno2\_rollno3$**.** For example $190010005\_190010006\_190010007.pdf$. Late submissions will not be graded. Students can discuss but must write their solutions based on their understanding independently. Do not use web resources or answers from your peers to obtain solutions. If anyone is involved in malpractice of any sort, then suitable disciplinary action will be taken.

This task is to implement 5 bandit algorithms on a testbed, compare and rank their performance and come up with a report on the performance evaluations with a clear description of the settings considered. When you compare your algorithms, clearly write your observations.

The test bed could be created along similar lines as in banditsComparison.pdf which is attached with the mail. The banditComparison.pdf assumes normal reward distribution for the arms. Recall in class we have discussed, Bernoulli reward distribution. For this assignment, you consider Bernoulli reward distributions and three different settings on the number of arms K=2,5 and 10. If any group is interested, they are free to explore normal reward distribution as well. In the case of normal reward distribution, for UCB algorithm refer to Figure 4 in auer.pdf with this mail and for Bayesian update in Thompson sampling refer to bayesNormal.pdf. Bonus points will be awarded if any group attempts normal reward distribution.

For performance evaluation metric for comparison, refer banditsComparison.pdf. It uses three criterions. You can use the same criterions for performance comparisons and also any other reasonable performance metrices to compare different algorithms.

The following are the five Bandit Algorithms whose performance needs to be compared.

- epsilon - greedy (fixed and variable)

- Softmax with different temperature parameters

- UCB

- Thompson sampling

- Reinforce Algorithm with and without baseline