Online Learning and Reinforcement Learning
Mid-Semester Exam
Course Instructor : Arun Rajkumar.
Duration : Until 11:59:59 PM on 18th June

INSTRUCTIONS: Please submit a single PDF titled with your name and roll number clearly specified. All answers must either be legibly written and scanned or typed. For programming based questions, paste code and the necessary plots obtained along with clear explanations. Plagiarism/copying of any form will lead to disciplinary action.

(1) State the Doubling trick formally. Explain what problem it solves and show how. (5 points)

(2) Consider the usual online learning protocol with a finite set of d experts. Consider an adversary who is always consistent with the majority of k out of d experts in the class (assume $k < d$ and k is odd). What is the worst case number of mistakes for any algorithm for this problem? Can you think of an algorithm which achieves the same? (10 points)

(3) Assume that you want to estimate the probability of a coin coming up heads. Let the true probability be μ.

- Assume $\mu = 0.75$. If you want to make a statement with at least 99% confidence that the estimate $\hat{\mu}$ satisfies $|\mu - \hat{\mu}| \leq 0.1$, how many times would you have to toss the coin? (4 points)
- If the true μ changed from 0.75 to 0.5, how does your answer from the first part change? (2 points)
- If the confidence changed from 99% to 90%, how does your answer from the first part change? (2 points)
- If the tolerance change from 0.1 to 0.01, how does your answer from the first part change? (2 points)

(4) An *epsilon*-greedy strategy for the stochastic multi-armed bandits set up exploits the current best arm with probability $(1-\epsilon)$ and explores with a small probability ϵ. Consider a problem instance with 10 arms where the reward for the i-th $(i = 1, \ldots, 10)$ arm is Beta distributed with parameters $\alpha_i = 5, \beta_i = 5 * i$.

- Implement the *epsilon*-greedy algorithm and compare it with the performance of the UCB and the EXP-3 algorithm. (5 points)
- Comment on your observations about the regret plots obtained in the previous part. (2.5 points)
- If you vary ϵ, how does the regret change? (2.5 points)