COMP 138 RL: Homework Template

Firstname Lastname

September 23, 2021

1 Goals

The goal of this assignment is to explore the inadequacies of the sample average method in non-stationary problems. Additionally, multiple feedback techniques will be used (ie, e-greedy, greedy, optimistic initial value, and Upper bound selection) on both stationary and non-stationary settings in order to illuminate the differences between the feedback techniques.

2 Introduction

We will define the stationary K-bandit problem as follows: There are k different actions to take. These actions will return some reward value which is chosen from a constant, normal Gaussian distribution with mean = 0 and standard deviation = 1. Action is chosen based on metrics derived from several approaches (sample mean, recency bias). Furthermore with multiple feedback techniques we will see a wide range of balance between exploration and exploitation which hopefully will provide insights into approaches in (non) stationary environments.

3 Conclusion

"I always thought something was fundamentally wrong with the universe" [**?**]

References