Student ID: ____________

Name: ____________

Mark: ____________

COMP6247 Lab - Weeks 7 and 8

Adversarial learning

Due 29/03/2019

Consider a simple 2-player capture the flag game, which consists of 5 sites and 2 flags. At each round, Player 1 (the hider) chooses 2 sites to hide her flags. Note that each site can contain at most 1 flag. Player 2 (the seeker), without knowing the exact places of the flags, has to make a guess where the flags could be. To do so, she can choose 2 sites to check whether the flags are hidden there. If the seeker finds both flags, she gets 2 points, if she finds only 1 flag, then her reward is 1, and 0 otherwise. The goal of the seeker is thus to maximise her total points over time, while the hider aims to minimise the total reward of her opponent.

1. [*4 marks*] After 100 rounds, the hider summarises on how many times she hid the flags at each site. Her totals are $(47, 53, 31, 36, 34)$ for sites $1, 2, 3, 4$, and 5, respectively. The seeker, on hearing this, claims that this is incorrect. Who is right? What about $(102, 26, 24, 40, 8)$?

2. [*4 marks*] Suppose that the number of times the hider left the flags at each site is given by the distribution $(36, 24, 40, 63, 37)$. Provide the best fixed guess strategy (i.e., the strategy that does not change over time) of the seeker in hindsight.

3. [*12 marks*] Design a guessing algorithm for the seeker such that the seeker can guarantee her expected regret to be at most $O(\sqrt{T})$, where T is the number of plays, no matter how the hider tries to adversarially change the sites of the hidden flags. Please implement the algorithm (feel free to choose your favourite programming language) and run it against the following opponents:

   - Uniform random opponent (the 2 flags are uniformly randomly placed).
   - The opponent follows an epsilon-greedy policy.
   - The opponent follows an FPL policy.

   Please play against each opponent 500 rounds, and repeat this for 100 times. Plot the average reward of the seeker (i.e., the total reward divided by the number of plays) over number of plays. Can you observe convergence of this value at all?