Assignment 3 - Report

1. **Different State Representations:** Yes, I did try different representations as mentioned below:
    a. Any given state is represented by x, y coordinates where x represents the horizontal distance between the upcoming pipe and the bird and y represents the vertical distance between the lower upcoming pipe and the y coordinate of bird.
    b. Discretizing the state representation in 2 different ways.
        i. We floor the x and y to multiple of 10.
        ii. We floor the x to multiple of 10 when nearest to pipe for more precise states and floored other values of states to multiple of 40. Similarly for y as well. However, this wasn't effective.
    c. Representing any given state by x, y, msec_to_climb where x and y are the same as in a. and msec_to_climb represents the remaining millisecond time left for climb to complete. Rationale behind this was that the bird when already in motion may not tend to jump and hence, different states.
    d. Same as in c. but the msec_to_climb is converted to 0-1 binary variable.

    Here, I have considered states every 10 frames. For some reasons, the choice state every 12 frames was performing very badly.

    In my case, b.(i) performed the best because of it's simplicity.

2. **Different exploration approaches:** I did try a couple of exploration techniques as mentioned below:
    a. After the score crossed 10 for fixed height pipes, the bird was making the same mistake at the same instant. I tried to randomize the action at this instant but it didn't prove effective for some reason. I used N(s,a) (number of times we have been in s and performed action a) to estimate if the repetitions of action is taking place.
    b. I tried to model α as 1/N(s,a) but then α became too small quickly and was preventing from learning.
3. **Different learning rates:** I tried 3 different learning rates:
    a. r: 1 if alive and -1000 if dead
    b. r: 10 if alive and -1000 if dead
    c. r: 1 if alive and -100 if dead

    The b. in my case performed the best. With r=1 for alive, the learning process was comparatively very slow. Also, in the case of -100 the mistakes committed by bird was taking time to get corrected because of less punishment.

4. **Time taken for training:** For the fixed pipe case, it took me 20 mins to train the bird for 10 pipes after which it started repeating itself. For random pipe case, I was not able to run the model

long enough for it to get stable. However, the maximum score I achieved was 14 and it took me 6 hrs for that to happen.

Given more time, I will try to properly do training for each of the cases mentioned above and then probably tweak the auto_player so that it performs better and converges quickly.

5. References:

a. https://github.com/chncyhn/flappybird-qlearning-bot
b. http://artint.info/html/ArtInt_265.html
c. http://sarvagyavaish.github.io/FlappyBirdRL/
d. http://www.mast.queensu.ca/~math472/FlappyQ.pdf