\*\*\* Double DQN = DQN + Double Q Learning

* What is the problem that you will be investigating? Why is it interesting?
* How common is overestimation with double Q learning and how can we reduce it?
* Does overestimation affect the performance of the algorithm?
* What is/are the plausible sources of overestimation?
* How to reduce overoptimism, resulting in more stable and reliable learning?
* To understand the “in-practice” aspects of implemented models/agents
* If relevant, what data, simulator or real-world RL domain will you be looking at? If you are collecting new datasets, how do you plan to collect them?
* Atari 2600 games: Alien, Space invaders, Time Pilot, Zaxxon, Wizard of Wor, Asterix
  + Using Arcade Learning Environment
* Collecting new datasets:
  + Function approximation?
  + Calculate the overestimations of DQN and Double DQN and compare the results
    - Test algorithm in a new domain
      * Compare DQN and Double DQN
      * Test different hyperparameters? (Double DQN tuned)
* What method, algorithm or theoretical analysis are you proposing? If there are existing implementations, will you use them, and how? How do you plan to improve or modify such implementations? If you are addressing a theoretical question, how do you plan to make progress?
* We will use Double DQN to create a policy to play specific games (choice of Atari or different video game domain?)
  + Need to start episodes from random starts along a human-played trajectory
  + Many existing implementations already exist, referring to the models in our selected paper and additional supporting surveyed literature.
  + Yes, we are planning to use them if provided, otherwise, we plan to re-implement the proposed algorithm.
  + We are planning to evaluate existing models compared to our modified model.
* Perhaps no longer a fair comparison: One option we have thought of for modification- we could do is to try to implement an actor-critic model that uses two networks compared to DDQN that only uses one network.
  + <https://www.freecodecamp.org/news/an-intro-to-advantage-actor-critic-methods-lets-play-sonic-the-hedgehog-86d6240171d/>
  + Actor critic models
  + Maybe find an easier method of improvement LOL :3, but maybe this could be pretty easy who knows
  + Another option we have thought of is to for modification- we could try different hyperparameters for the Double DQN?
  + Another option we have would be to try different policy methods and compare which perform better
* Epsilon greedy
* What literature have you already surveyed or will be examining to provide context and background?
* Deep Reinforcement Learning with Double Q-learning Hasselt et al. 2015 (our chosen paper)
* <https://towardsdatascience.com/atari-reinforcement-learning-in-depth-part-1-ddqn-ceaa762a546f>
* Sutton and Barlo, Reinforcement Learning 2nd Edition Ch. 6.7 Maximization Bias and Double Learning
* This paper uses the same algorithm in our paper but for transfer learning, maybe there are insightful things we can draw from this paper? Who knows? <http://web.stanford.edu/class/cs234/past_projects/2017/2017_Asawa_Elamri_Pan_Transfer_Learning_Paper.pdf>

<http://web.stanford.edu/class/cs234/past_projects/2017/2017_Asawa_Elamri_Pan_Transfer_Learning_Poster.pdf>

* How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

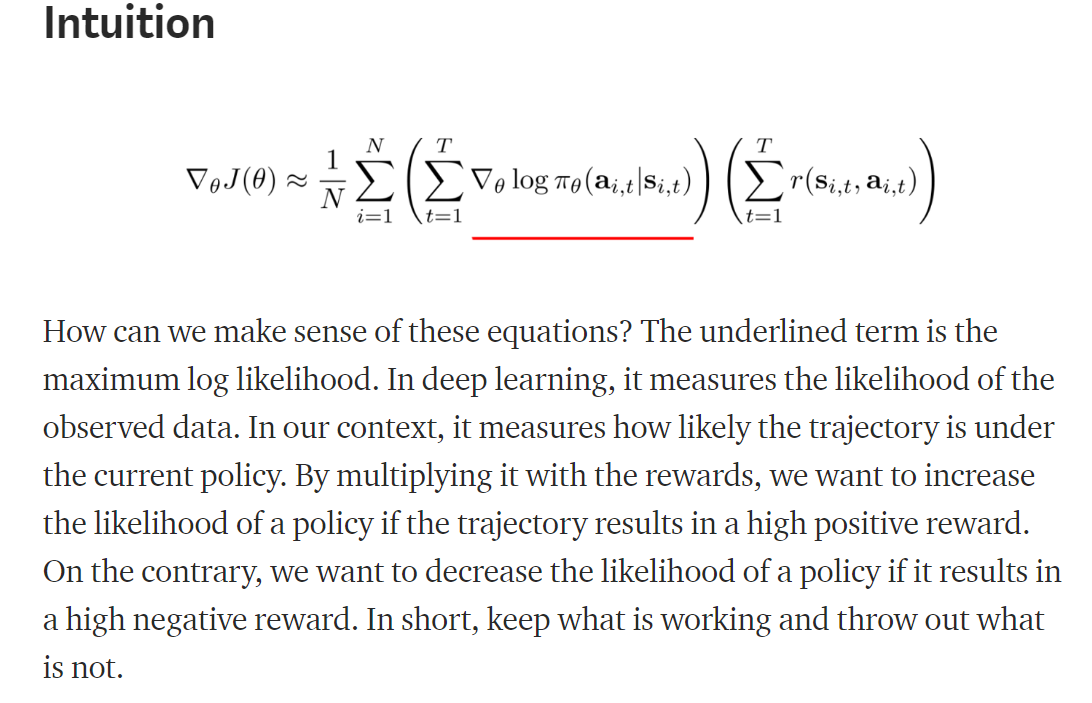
Evaluating the results will be

* This will probably lean more towards the performance metrics shown through graphs and figures. Perhaps tables would be beneficial if our problem has any statistical results to report.
* Start tests from random starts along a human trajectory and compare the terminal score with the score the agent started with
* Given that we have the data we can use provided data to compare the results
* Given that we don’t have the data we can collect the data ourselves by simulating episodes
  + Compare Double DQN vs. DQN vs. Human performance
    - DQN vs. Double DQN w/ same hyperparameters
    - DQN vs. Double DQN tuned
    - vs modified model (actor-critic) or other ideas?
  + Compare the value estimates of both networks to see if DQN is overestimating
    - See the value estimates for the policy vs. the actual score of the policy (see if the overestimate is leading to sub-optimal policies

**Appendix of random things/supportive material:**

Post on actor critic and maximum likelihood regarding why gradient log policy:

<https://medium.com/@jonathan_hui/rl-policy-gradients-explained-9b13b688b146>



Screen shot of what Double DQN is doing:

