Last edited: Just now

# **July 22nd, 2020**

* Completed David Silver Videos till Model-Free Prediction part (Lecture 4)
  + Monte Carlo Method for policy Evaluation
  + Temporal Difference Method for policy Evaluation
  + Lambda spectrum of TD Learning and n-steps Return
  + Bootstrapping and Sampling concepts in context with MC, TD, DP
  + Eligibility Traces for Credit assignment to causations. ****(Math is still not clear)****
* Gone through the documentation of Spinning Up of OpenAI for Installation and Algorithm Template.
* Installed Spinning Up in laptop at home. Will do the same at office workstation on next visit to the office.

# **July 27th, 2020**

* David Silver Video of Model-Free Control (Lecture 5)
  + Model-Free Policy Iteration with Action Value Function
  + Epsilon Greedy Exploration
  + GLIE Monte-Carlo Control ****(Need to confirm with Arthur if understanding is correct)****
  + Updating action value functions with Sarsa
  + On-Policy Control with Sarsa
* SUMO Setup and Learning
  + Hello World Example
  + OSMWebWizard Example
  + Quick Start Example
  + Imported the Environment created by Arthur

# **July 31st, 2020**

* David Silver Video of Model-Free Control (Lecture 5)
  + Convergence of Sarsa ****(Need to discuss/confirm step-size concept and statistic approximation theory)(Do we decay step-size as well?) (We do decay probability of epsilon-greedy exploration with time i.e. GLIE, if your understand of GLIE is correct)****
  + Windy Gridworld Example ****(Does GLIE decides exploration rate?)****
  + n-Step and Forward View Sarsa
  + Off-Policy Learning - Importance Sampling
  + Off-Policy Learning - Q Learning (Sarsamax)
  + Comparison of TD Learning with Dynamic Programming
* Revision of Model-Free Prediction and it's summarization
* Revision of Model-Free Control and some repetitions as the concepts were not clear in the first go.

# **August 3rd, 2020**

* David Silver Video of Value Function Approximation (Lecture 6)
  + Value Function Approximators Types
    - State In -> Value Out
    - State, Action In - > q-value Out
    - State In -> q-values of all actions for that state Out
  + Linear Combination and SGD in Value Function Approximation
  + MC and TD to generate training data for Value Function Approximators
  + Control with Value Function Approximation ****(Discuss the iteration graph with Arthur to get better understanding. At which step do we update weights, evaluation or improvement? Intuitively, it should be at improvement, but David Silver says it will be at evaluation. May be, the policy gets improved in improvement, but the q-values get improved in evaluation step and then q-values are used as target to update weights so weights would be improved in evaluation step.)****

# **August 5th, 2020**

* David Silver Video of Value Function Approximation (Lecture 6)
  + Action-Value function approximation
  + Batch Methods - Least Square Prediction
  + SGD with experience replay
  + Experience replay in Deep Q-Networks
* TRACI
  + Read about TRACI and its protocol
  + Traffic Light Simulation attributes
  + Explored runner.py of TraciTLS example