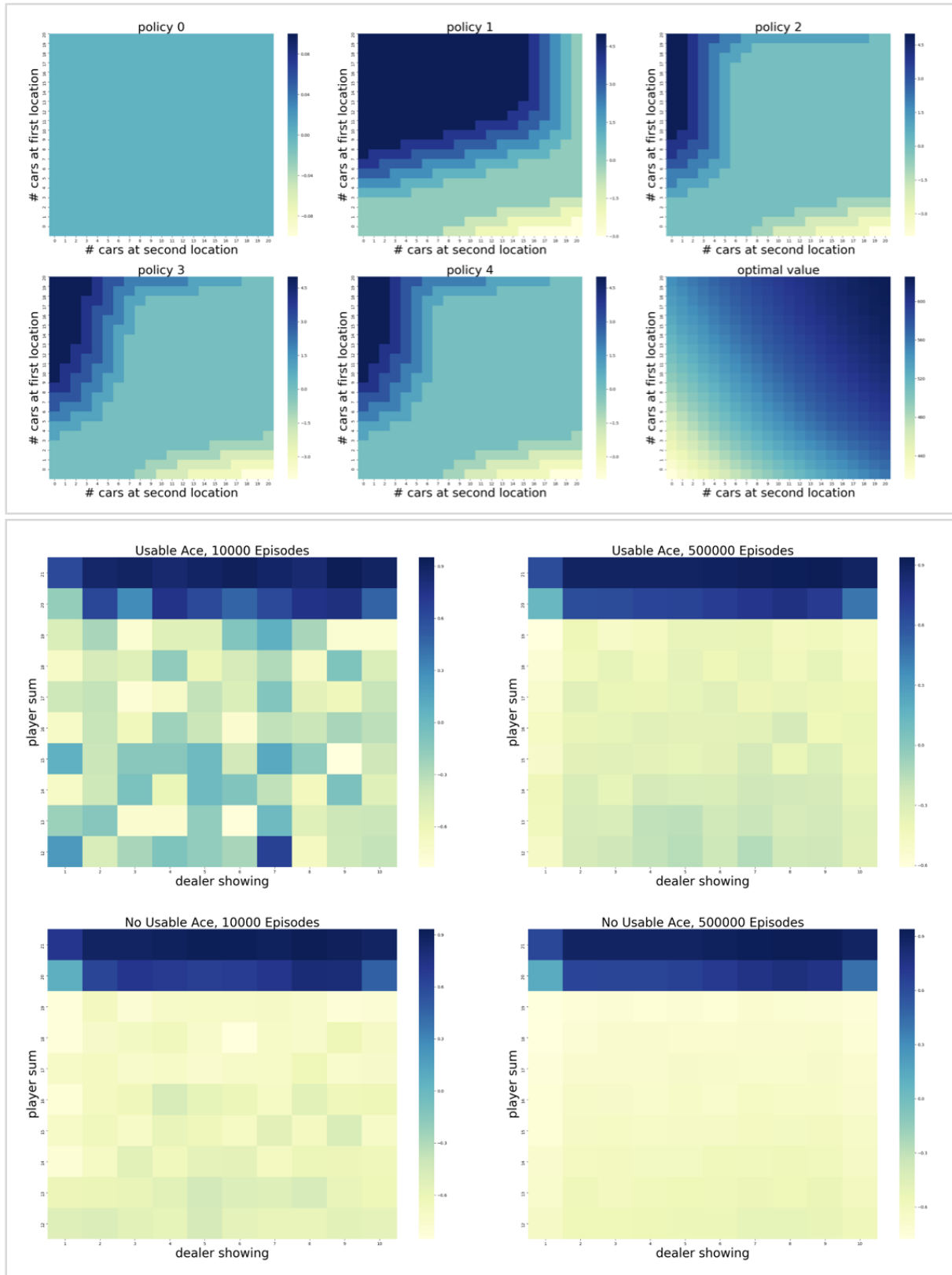


week08

Work

- finish *reinforcement learning : an introduction* chapter 3,4,5 and reproduce the given example based on existing template
- car rental problem & gambler problem



- optimization problem review and reproduce code in Matlab and Julia
- David Silver Slide remarked and upload to repo
- GitHub repo rearrange to better store

book_marked	week eight updated	5 hours ago
code-written-for-learning	repo rearranged	5 days ago
david-silver-course	update a little	4 days ago
rl-an-intro-reproduce	week eight updated	5 hours ago
weekly-summary	repo rearranged	5 days ago
.gitignore	add ignore	last month
README.md	repo rearranged	5 days ago

README.md		
-----------	--	--

deep-RL-weekly

- this repo is used for recording the learning process for deep RL.
- the exact learning trajectory of each week's update and work will be addressed in the weekxx.md (start from week08.md which is the first week after rearrange this repo)

Online Courses

- [CS294-112 at UC Berkeley : Deep Reinforcement Learning](#)
- [CMU10703 : Deep Reinforcement Learning and Control](#)
- [Deep RL Bootcamp at Berkeley](#)
- [David Silver RL Course at UCLYoutube](#)
- [CS234 at Stanford : Reinforcement Learning](#)
- [CS598 at UIUC : Statistical Reinforcement Learning](#)

Great Blog

- [Hongzi Mao](#)
- [Policy Gradient Algorithms](#)
- [An Outsider's Tour of Reinforcement Learning](#)

General Essay

- [OpenAI Spinning Up](#)
- [A beginner's guide to deep reinforcement learning](#)

ASSIST2019 Notes

- Session 3
 - Tony Quek – AI : A Networking and Communication Perspective
 - Federated Learning
 - Threshold-based
 - Stochastic Binarization Layer
 - Bo An – When AI meets Game Theory
 - RL and GT

- Computer Poker
- RL solves the Fraud Detection Problem
- Nan Jiang – Sample-efficient Exploration in Reinforcement Learning with Function Approximation
- Session 5
 - John C.S. Lui – An Online Learning Approach to Network Application Optimization with Guarantee
 - online is essential for the real world where the PARAMs are unknown
 - optimization with guarantee – minimum guarantee threshold – MAB
 - use least coins to get your estimates
 - deal with exploration and exploitation – short-term and long-term
 - two-layer reward – combined into combined reward
 - stochastic (level-1) and non-stochastic (level-2)
 - the total level-1 reward is guaranteed
 - Atilla Eryilmaz – Leveraging Side-Information for Learning and Optimization Under Uncertainty with Applications in Social and Communication Networks
 - Learning to Control Renewal Processes
 - UCB + interrupt Algorithm
 - Heavy-tailed distribution
 - Median of Mean Estimator
 - Online Free Trail Strategy
 - stochastic knapsack problem
 - Mengdi Wang – Reinforcement Learning in Feature Space: Matrix Bandit, Kernels, and Regret Bound
 - control problem + information theory
 - problems
 - how many samples are needed to learn a 90%-optimal policy
 - how many regret to pay when learning to control real time
 - information-theoretical limit
 - the optimal sampling-based algorithm
 - Q-learning is provably efficient
 - Bellman equation is equivalent to a bilinear saddle point problem (duality)
 - Dimension Reduction
 - state feature map
 - representing value function using linear combination of features
 - Reducing bellman equation using features
 - Hilbert space embedding of transition kernel
 - the MatrixRL algorithm

- Jia Liu – Combinatorial Bandits with Fairness Constraints
 - reward max v.s. fairness
 - MAB
 - Combinational MAB
 - Sleeping MAB
 - MAB with constraints