David Sammel
Final Project
CS4701

Creating Blackjack AI Using Reinforcement Learning

Blackjack

- One of the most popular casino games in the world

- Try to maximize the sum of your cards without going over 21

- Beat the dealer to win

Reinforcement Learning

- An agent observes the state of the world

- Takes an action to maximize outcome

- Tunes the decision algorithm after outcome of the action is recorded

Project Set Up

- Designed a blackjack simulation

- One dealer, and any number of players to play against the dealer

- Some players play with basic strategies, two learn as they play

Learner

- The basic learning agent is called Learner

- State Considered: Own hand, Dealer's up-facing card

- Has a matrix of probabilities of hitting given the state

- Matrix entries are initialized to .5

Learner (cont.)

- Upon winning matrix is unchanged

- Player can lose in three ways
  - Player busts: lower probability of hitting next time
  - Dealer's hand is higher: raise probability of hitting next time
  - Dealer blackjack: Don't change matrix

Counting Cards

- A way to consider more state of the game
- Keep track of the cards left to be dealt in the deck
- If a low card is dealt (2-6) increase count, if a high card is dealt (10-Ace) decrease count
- High count favors the players, low count favors the dealer
- Strategy may change depending on the count

CountingLearner

- Similar to Learner, but it considers the count as part of the state

- Has 3 matrices, one for each of when the deck is high, low, or neutral

- Plays slightly differently depending on the count

Other Players

- ## Dealer's playstyle
  - Another player mimics the dealer's style
  - Hits when it's hand is below 17, stands on 17 or above

- ## DumbPlayer
  - Randomly guesses between hitting and standing with 50% probability of both
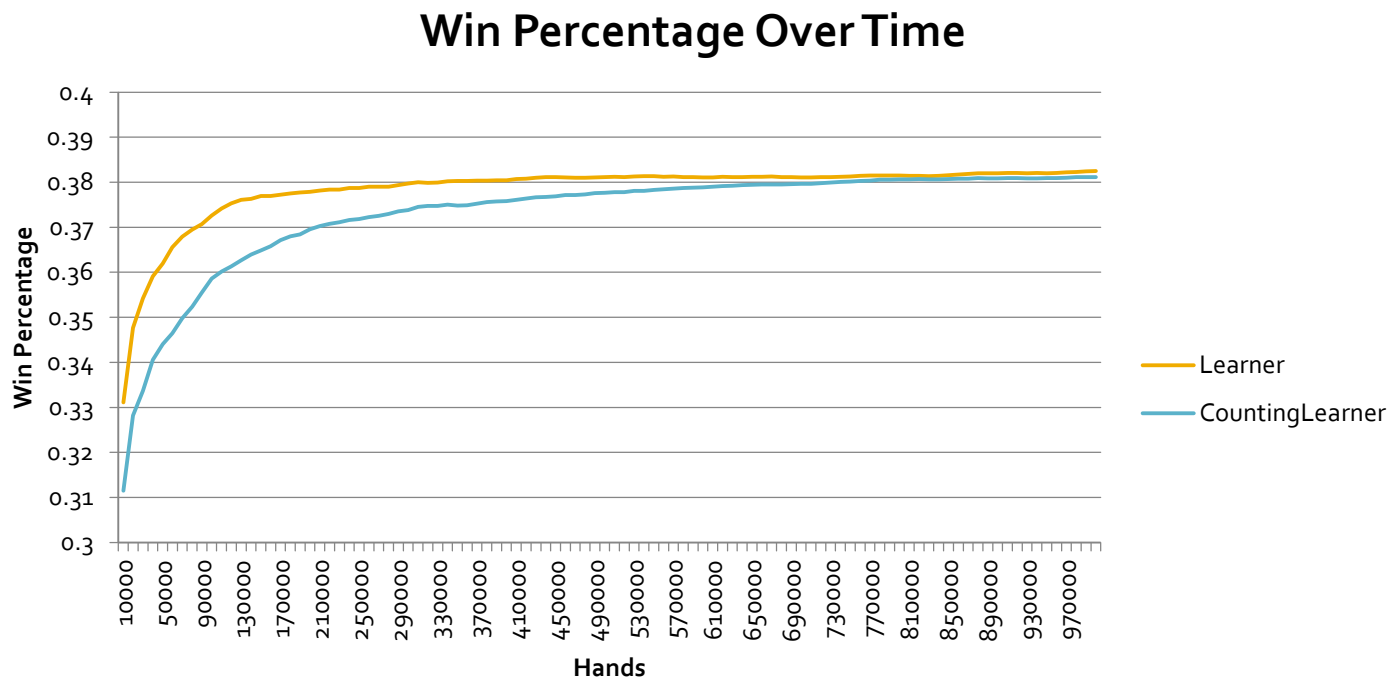 - Used as a baseline for the learning models

Results

- After playing 2 million hands:

Style	Dealer	Count (high)	Count (low)	Dumb	Learner	Counting Learner
Win pct.	41.4%	41.0%	42.1%	29.9%	38.3%	38.7%

Learning models perform well compared to the baseline model

Learning Rate

- Over the first 1 million hands, the model's accuracy rates:

Win Percentage Over Time

Conclusions

- The learning agents successfully learned a rule to play blackjack
- Win percentage over time increases for both models
- Card counting model outperforms the regular model
- Winning percentage much better than the baseline, approaching dealer's strategy