Can Blackjack Be Profitable?

An RL-based Approach to Generating Strategies in a Betting Game

Amy Kang, George Zhang

CS182: Artificial Intelligence, Fall 2016

What is Blackjack?

Popular casino card game

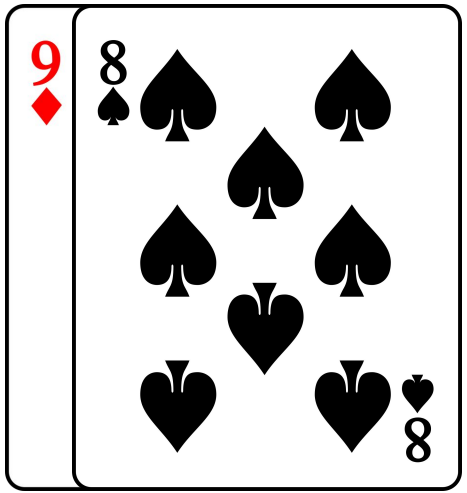
Basic gameplay

- Given two cards and can see dealer's upcard
- Decide whether to hit or stand
- Goal: beat dealer's hand without busting

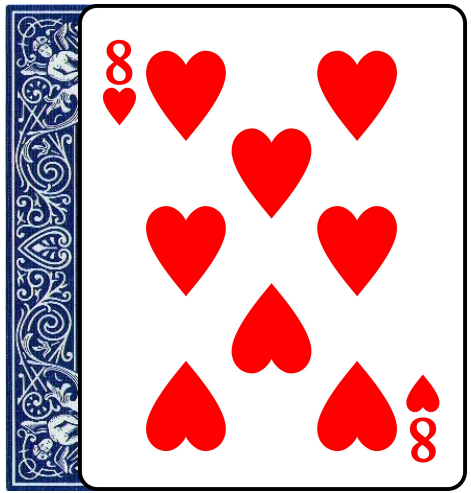
Gambler's Ruin

- Most variants have a house edge from 0.5% - 1%
- More beatable than most casino games
- Still will lose in the long run!

Sample Scenario



Player
Value: 17



Dealer
Value: 8

Actions

- **Hit**
- **Stand**
- **(Double)**
- **(Split)**

Our Goal

Most blackjack agents seek to maximize **win percentage.**

Ultimately, we want to train our agent to improve on the Basic Strategy by learning how to bet and how to maximize **earnings.**



Our Approach

Why Q-Learning?

Intuitive action-reward model, non-adversarial game

Online learning (integration with simulator), model-free

Relatively small state space: ~250 states with only 2-4 actions per state

Comparison to Basic Strategy

Basic Strategy is a policy over the player's and dealer's hand values generated from millions of computer simulations (IBM).

Our conjecture is that the optimal policy learned by our Q-learning agent will approach the Basic Strategy.

Game Representation

State Space

Player's hand value
Whether player's hand is "soft"
Dealer's hand value

Actions

Hit or stand

Rewards

On hit, Q-value of the subsequent state
On bust or stand, the payoff of the round according to bet

Current Results - Average Win Rate

Average Win Rate*		
Random Agent	Basic Strategy Agent	Our Q Learning Agent**

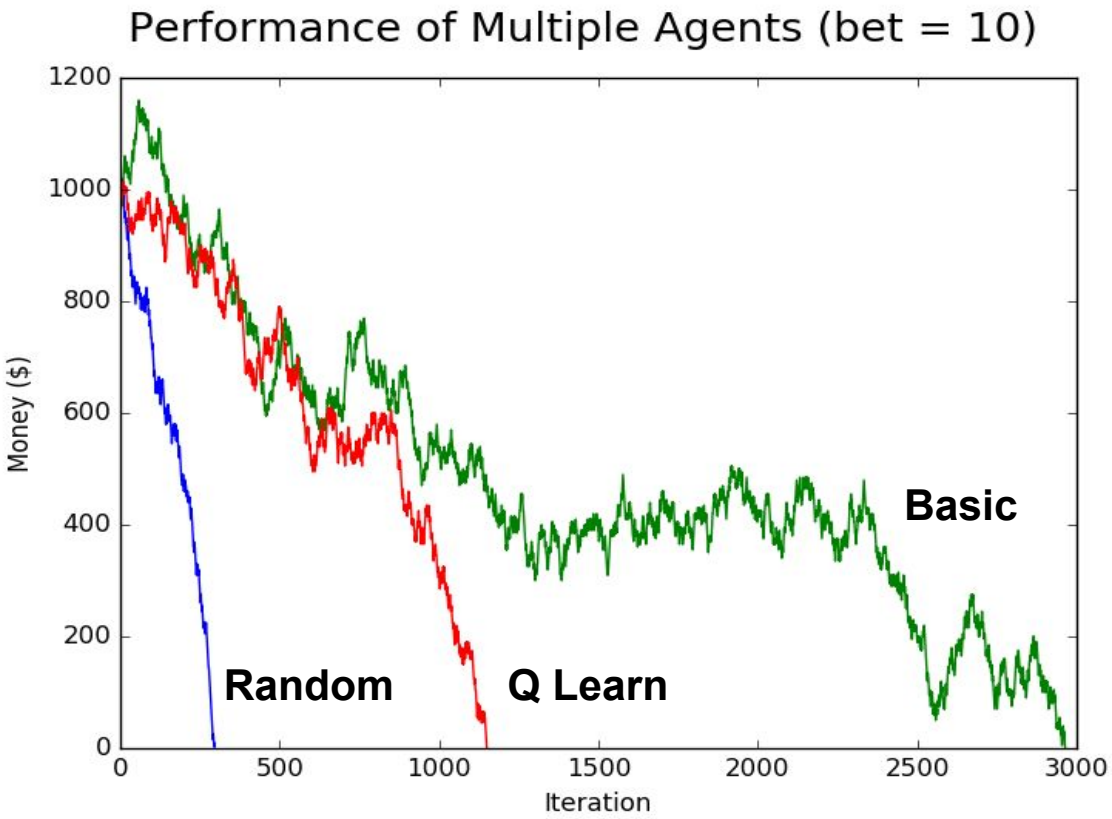
* Average win rate calculated over 100 games
** Trained on 1,000,000 rounds (ε = 0.4, α = 0.01, γ= 0.9)

Current Results - Average Win Rate

Average Win Rate*		
Random Agent	Basic Strategy Agent	Our Q Learning Agent**
28.88%	42.85%	37.88%

* Average win rate calculated over 100 games
** Trained on 1,000,000 rounds (ε = 0.4, α = 0.01, γ= 0.9)

Current Results - Money over Time



Basic Strategy vs. Q-Learning Actions

Table 1: Basic Strategy

Player Hand	Dealer Hand									
Hard	2	3	4	5	6	7	8	9	10	A
20	S	S	S	S	S	S	S	S	S	S
19	S	S	S	S	S	S	S	S	S	S
18	S	S	S	S	S	S	S	S	S	S
17	S	S	S	S	S	S	S	S	S	S
16	S	S	S	S	S	H	H	H	H	H
15	S	S	S	S	S	H	H	H	H	H
14	S	S	S	S	S	H	H	H	H	H
13	S	S	S	S	S	H	H	H	H	H
12	H	H	S	S	S	H	H	H	H	H
11	H	H	H	H	H	H	H	H	H	H
10	H	H	H	H	H	H	H	H	H	H
9	H	H	H	H	H	H	H	H	H	H
8	H	H	H	H	H	H	H	H	H	H
7	H	H	H	H	H	H	H	H	H	H
6	H	H	H	H	H	H	H	H	H	H
5	H	H	H	H	H	H	H	H	H	H
4	H	H	H	H	H	H	H	H	H	H

Soft	2	3	4	5	6	7	8	9	10	A
A, 9	S	S	S	S	S	S	S	S	S	S
A, 8	S	S	S	S	S	S	S	S	S	S
A, 7	S	S	S	S	S	S	S	H	H	H
A, 6	H	H	H	H	H	H	H	H	H	H
A, 5	H	H	H	H	H	H	H	H	H	H
A, 4	H	H	H	H	H	H	H	H	H	H
A, 3	H	H	H	H	H	H	H	H	H	H
A, 2	H	H	H	H	H	H	H	H	H	H
A, A	H	H	H	H	H	H	H	H	H	H

Table 2: Q Learning Actions

Player Hand	Dealer Hand									
Hard	2	3	4	5	6	7	8	9	10	A
20	S	S	S	S	S	S	S	S	S	S
19	S	S	S	S	S	S	S	S	S	S
18	S	S	S	S	S	S	S	S	S	S
17	S	S	S	S	S	S	S	S	S	S
16	S	S	S	S	S	S	S	S	S	S
15	S	S	S	S	S	S	S	S	S	S
14	S	S	S	S	S	H	S	S	S	S
13	S	S	S	S	S	H	S	S	S	S
12	S	S	S	S	S	S	S	S	S	S
11	H	H	H	H	H	H	H	H	H	H
10	H	H	H	H	H	H	H	H	H	H
9	H	H	H	H	H	H	H	H	H	S
8	H	H	H	H	H	H	H	H	S	S
7	H	H	H	H	H	H	H	H	S	S
6	H	H	H	H	H	H	H	S	S	S
5	H	H	H	H	H	S	H	S	S	S
4	S	H	H	H	H	S	H	S	S	S

Soft	2	3	4	5	6	7	8	9	10	A
A, 9	S	S	S	S	S	S	S	S	S	S
A, 8	S	S	S	S	H	S	S	S	S	S
A, 7	H	H	H	H	H	S	S	S	S	S
A, 6	H	H	H	H	H	H	H	S	H	S
A, 5	H	H	H	H	H	H	H	H	S	S
A, 4	H	H	H	H	H	H	H	H	S	S
A, 3	H	H	H	H	H	H	H	H	S	S
A, 2	H	H	H	H	H	H	H	S	H	S
A, A	H	H	H	H	H	H	H	H	H	S

Next Steps

Tweaking rewards to more closely approach basic strategy

Need to determine why we play conservatively and how to fix

Implementing doubling / splitting

Basic strategy also includes these, so we'll continue to use it as baseline

Expect strategic doubling / splitting can benefit player

Counting Cards

Card counting strategies take away from the house edge

Learning How to Bet

Strategic betting when counting may lead to greater payoffs when deck is favorable

Monte-Carlo Simulation

Expect to influence training time, as knowledge about reward propagated back faster

Questions?