Resit Project 2

Jan van Ruijven 2006698

January 26, 2024

Question1

1

The first greedy strategy is to go for streets whenever the odds seem adequete. An overview of this algorithm can be find in the appendix at algorithm 1. The second greedy algorithm is to go for combinations of three when there are two of the same dices. An overview of this algorithm can be found in the appendix at algorithm 2. The final algorithm consist of both the combination algorithm and the streets algorithm, by applying them both in that particular order. Important to note is that the dices are always sorted from low to high.

2

In this game there are $6 * 6 * 6 = 216$ possible states the game can be in and in each state the same list of 8 actions is available:

Throw first dice

Throw second dice

Throw last dice

Throw first and second dice

Throw first and last dice

Throw second and last dice

Throw all three dices

Throw no dice (stop the game)

A Q table is produces with 216 states and initialized with 8 zero's (one for each action). The reward for each action is equal to the change that is created after the action. So if the actions leads to an improvement the reward is positive and vice versa. This also means that the last action (doing nothing) results in a reward of 0.

3

Because each throw is independed of that what has happend before, gamma is set to one, so future rewards are just as important as rewards in the beginning of the game. For alpha, a relatively low value has to be chosen for proper learning (0.01 in this case), this is due to to fact that we are dealing with a game where the statistical average of a decicion is very importan. If alpha would be set to a high value, we might just end up with a Q table with values that are only high by change, not because they are consitantly outperforming the other options. Epsilon is set to 0.1, the algorithm should try new actions from time to time, but overall it is able to find the most suitable action, even when epsilon is very close to 0. This is due to the fact that the starting values are equal to 0, so taking wrong decicions will lead to negative values in the Q table, therefore the actions with a reward of 0 associated, will eventually be chosen.

4

After applying the greedy algorithms and letting the Q learning algorithm learn over 100.000 throws. The Q learning is ablo to outperform all of the greedies.

djfakl

Algorithms

Algorithm 1: Greedy steets Algorithm

```
Data: $D_i \quad i = 1,2,3 \quad C_j = [D_2 - D_1, D_2 - D_3]$
if $\sum_{j=1}^{2} C_j <= 1$ then
    Do nothing. We have combinations.
end
else if $C = [2,1]$ then
    Rethrow first dice
end
else if $C = [1,2]$ then
    Rethrow last dice
end
```

Algorithm 2: Greedy Combinations Algorithm

```
Data: $D_i \quad i = 1,2,3 \quad C_j = [Dice_2 - Dice_1, Dice_2 - Dice_3]$
if $\sum_{j=1}^{2} C_j <= 1$ then
    Do nothing. We have combinations.
end
else if $C[0] = 0$ then
    Rethrow last dice
end
else if $C[1] = [0]$ then
    Rethrow first dice
end
```