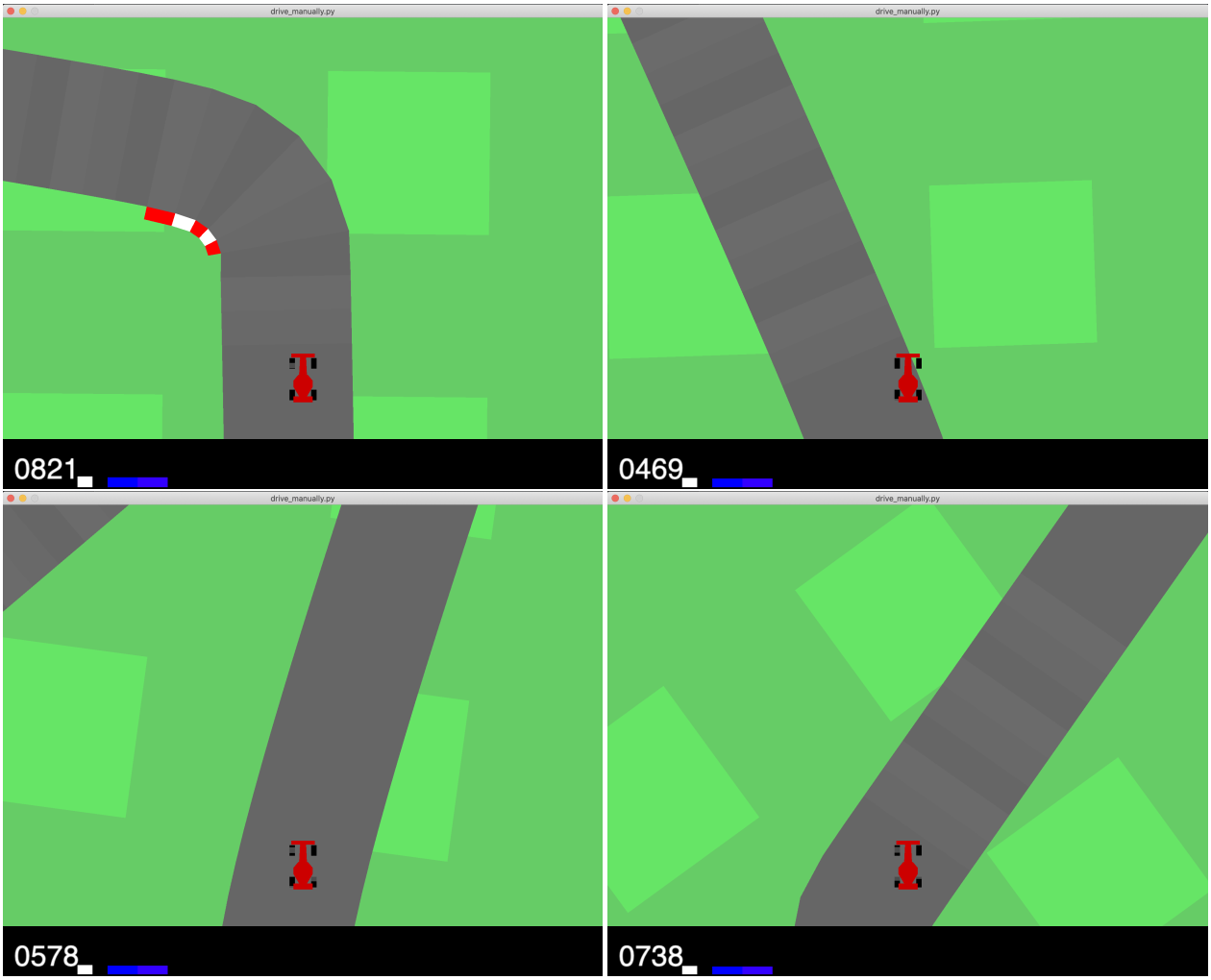
CSCI-GA 3033-090 Homework 2

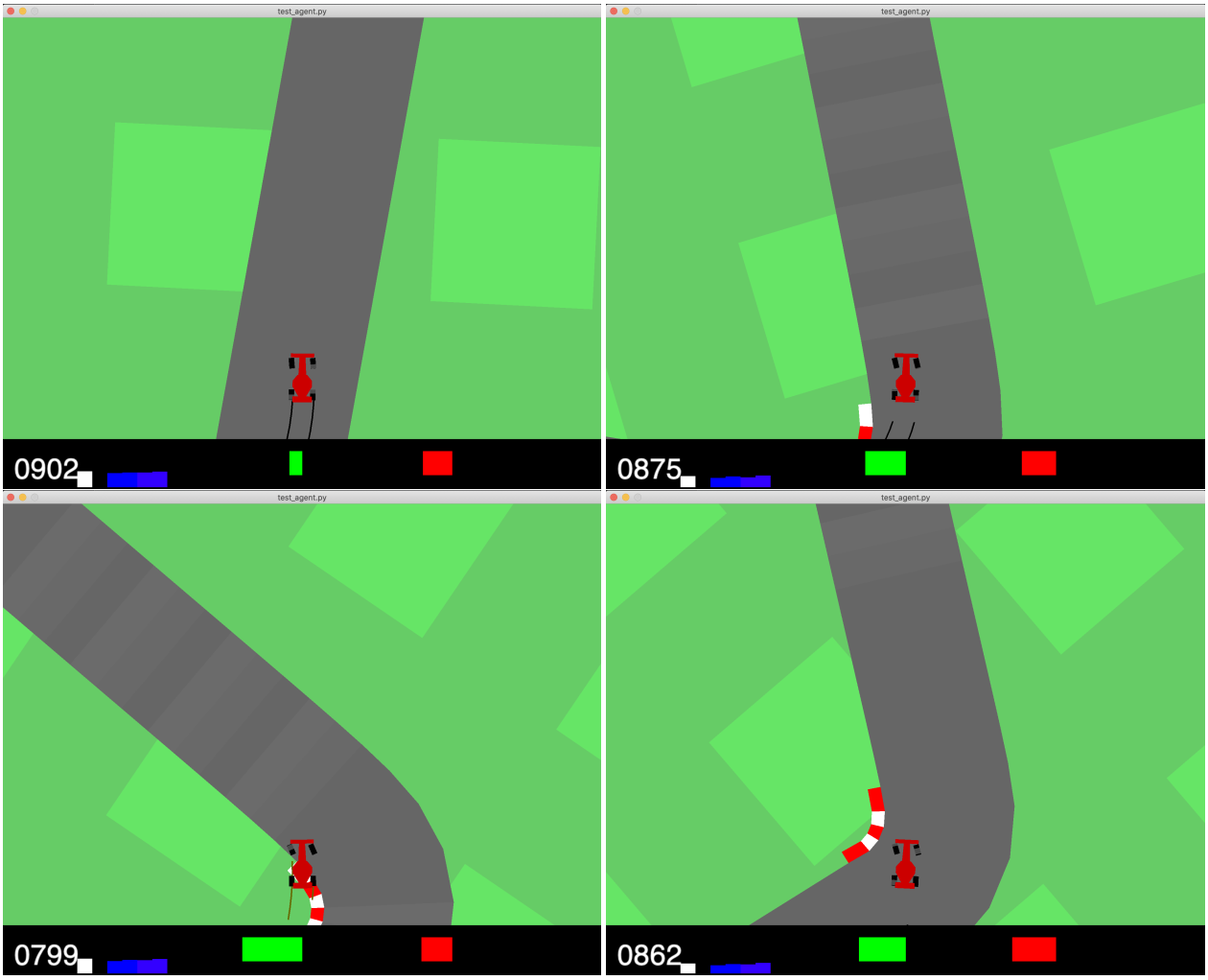
Nikhil Supekar (ns4486@nyu.edu)

October 6, 2020

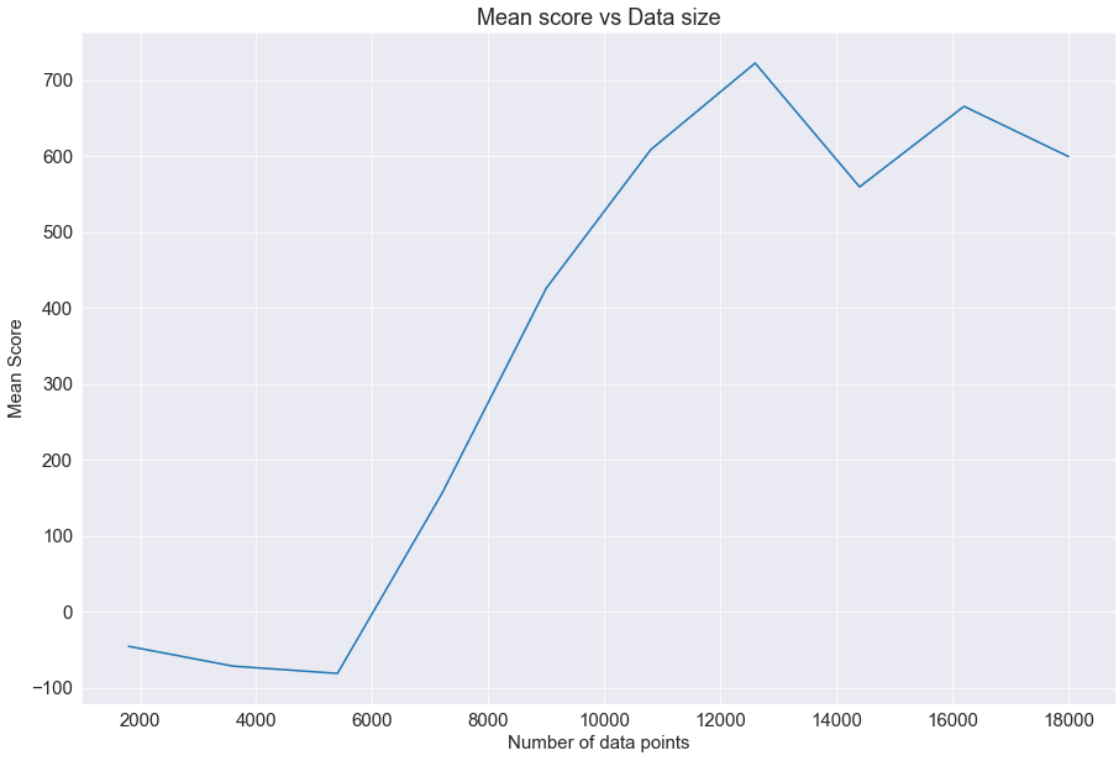
1. Part 1 - Question 2 - Data Collected



2. Part 1 - Question 3 - Behavior Cloning Output



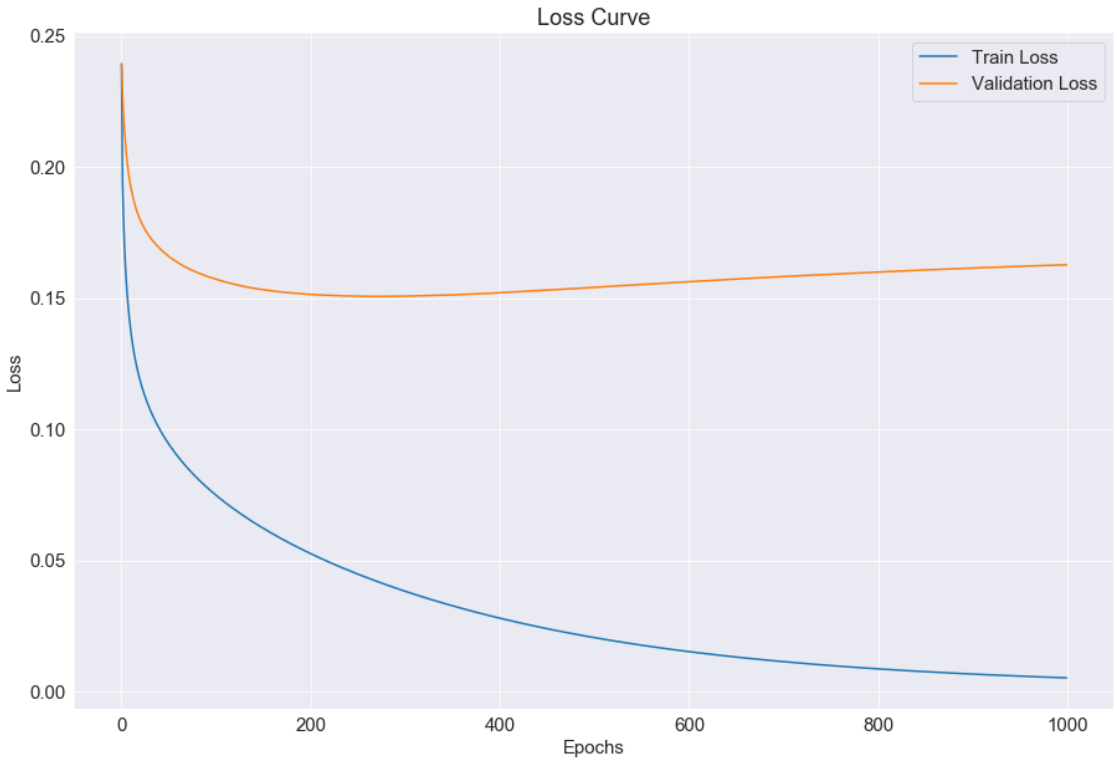
3. Part 1 - Question 4 - Data size for Behaviour Cloning



The curve suggests that around 10k - 12k samples are good enough to train an acceptable model.

[Disclaimer: The above curve has been generated by training models for only 50 epochs due to resource restrictions to accommodate more data set sizes. 50 was chosen since we see a considerable drop in training loss in this range which was a good tradeoff between resources and performance.]

4. Part 1 - Question 5 - Behaviour Cloning plots



Result generated by the best model:

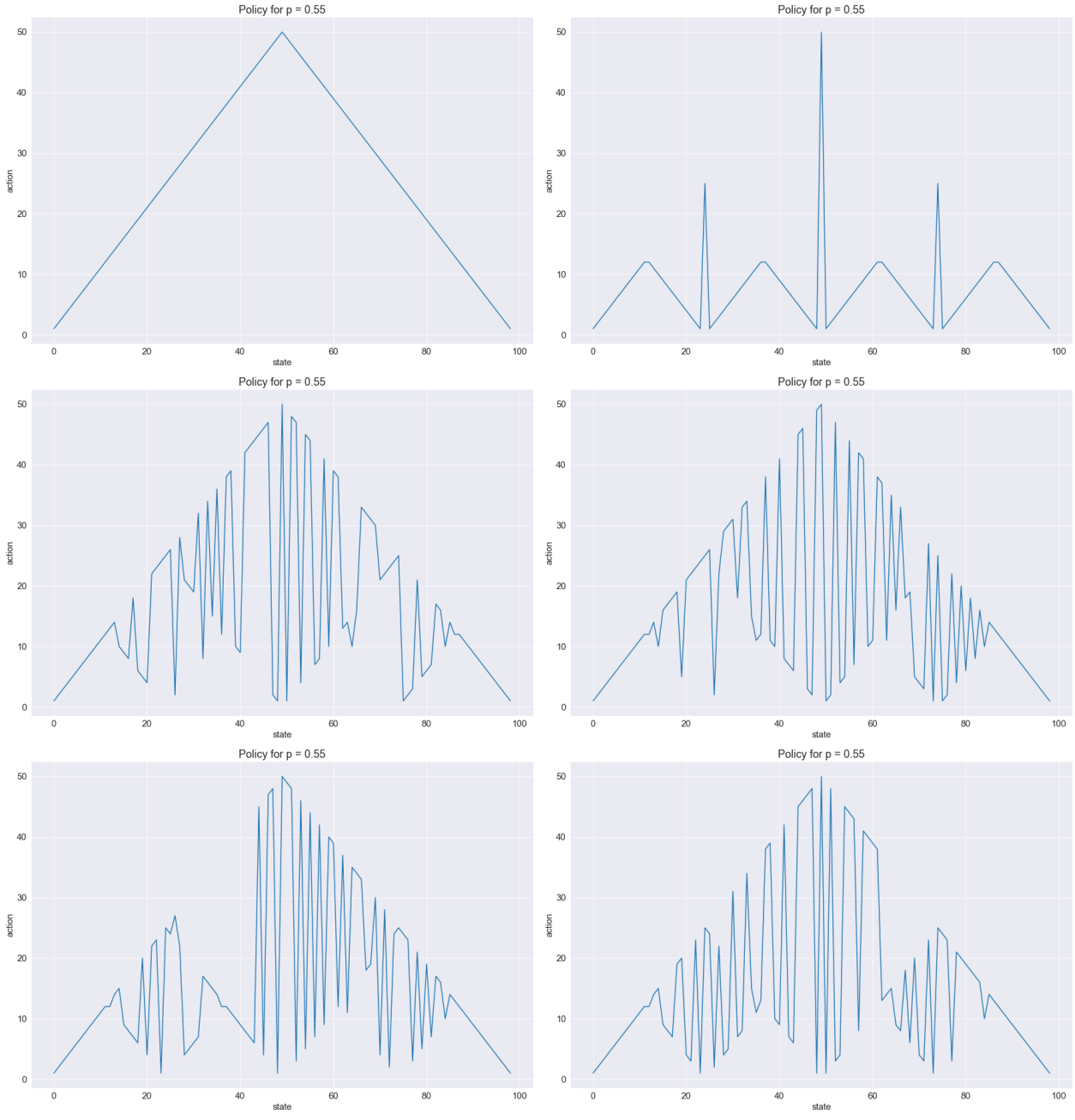

```
{
    "episode_rewards":[
        926.8999999999871,
        875.6785467127917,
        872.4085910652726,
        566.5666666666511,
        915.3999999999921,
        811.1627986347924,
        875.3385964912092,
        923.1999999999862,
        846.3882943143602,
        920.9999999999866,
        863.9287769783986,
        842.2728813559172,
        844.3444444444232,
        882.2943661971627,
        644.3794952681219
    ],
    "mean":840.7508972086034,
    "std":98.85162223457978
}
```

5. Part 1 - Question 6 - Tricks to improve Behaviour Cloning

1. Using standard models instead of hand-engineered conv models. Using ResNet-18 dramatically improved the performance of the agent.
2. Early stopping for model selection. As seen in the loss curve, we see that the validation loss starts increasing after around 200 epochs. The model was selected once the validation loss started increasing.
3. Conv models based on frame history. ResNet-18 was modified to accommodate historical frames for action prediction and improved mean score performance of the agent by 70.

6. Part 2 - Question 1 - Family of Policies

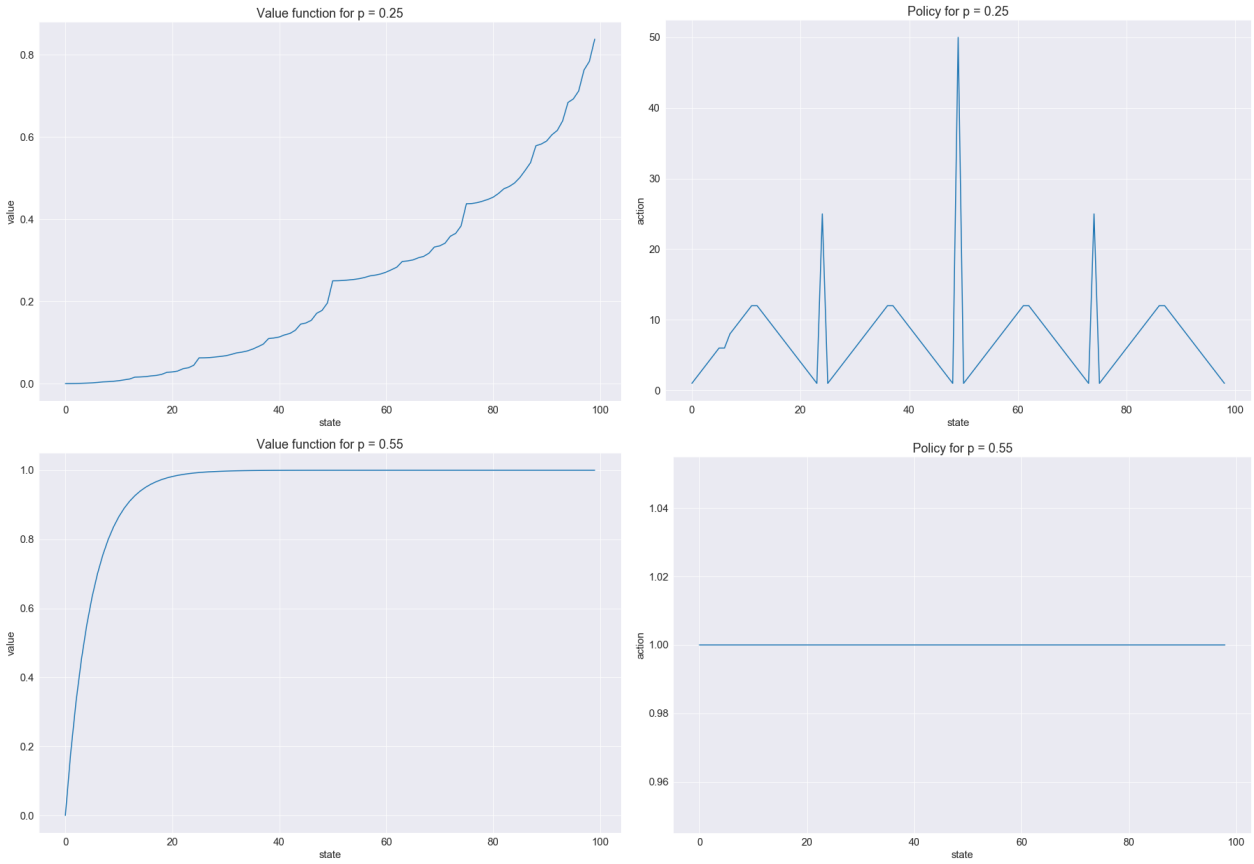
The family seems to resemble fractal triangles made by breaking a bigger triangle. This is illustrated in some of the policies discovered in further questions:



7. Part 2 - Question 2

The policy that bets only one on a capital of 51 can be considered good for $p_h = 0.4$ since the probability of getting tails is greater than 0 and losing in this state takes us to 50 where we would still have a shot at winning in the next turn. If we win that's anyway good and we can continue playing. In general, the pattern of this policy can be summarized as trying to get ourselves to a position where we would have a great chance of getting the reward.

8. Part 2 - Question 3



9. Part 2 - Question 4

$$Q_{i+1}(s,a) \longleftarrow \sum_{s'} T(s,a,s') \left[R(s,a,s') + \max_{a'} \{Q_i(s',a')\} \right]$$