

# Imitation Learning

Andrija Djurisić

July 2020

Imitation learning is the umbrella term for family of Machine Learning methods where neural net attempts to learn from human demonstration.

## Behavioral cloning

Behavioral cloning treats imitation learning as a supervised learning problem. That is, it learns to map actions to states [1] the same way a neural network competing in the ImageNet challenge learns to map labels to images. By doing that we are learning the policy  $\pi_{\theta}(a_t|o_t)$ .

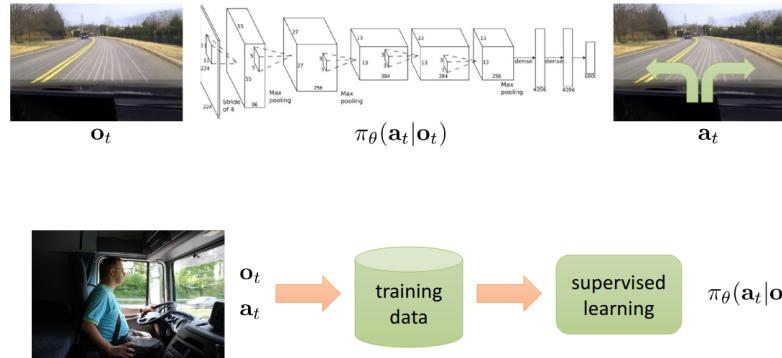


Figure 1: Behavioral cloning.

The theoretical reason why behavioral cloning might not do so well is that when we fit our policy to training data, learned policy won't be perfect. It will occasionally make mistakes and those mistakes will gradually create a bigger and bigger drift [2], thus putting the agent in the state with the properties that weren't encountered during the training.

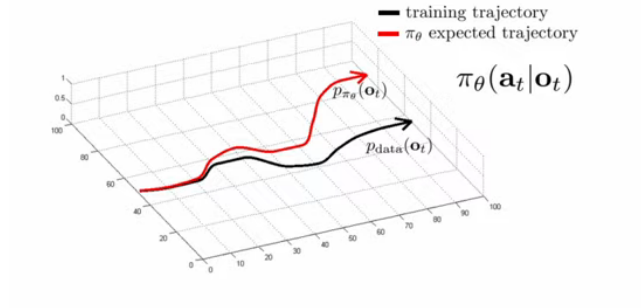


Figure 2: Drift caused by accumulating mistakes due to imperfect policy.

### Dagger

One way to get around the problem we described is to collect more data, especially in the areas where agent drifted away. The **Dataset Aggregation** is a four steps algorithm [3] that describes how to collect more data.

1. train  $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$  from human data  $\mathcal{D} = \{\mathbf{o}_1, \mathbf{a}_1, \dots, \mathbf{o}_N, \mathbf{a}_N\}$
2. run  $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$  to get dataset  $\mathcal{D}_\pi = \{\mathbf{o}_1, \dots, \mathbf{o}_M\}$
3. Ask human to label  $\mathcal{D}_\pi$  with actions  $\mathbf{a}_t$
4. Aggregate:  $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_\pi$

Figure 3: Dagger.

### Theoretical evaluation of Behavioral cloning

*Proof.*

□