Carnegie Mellon
School of Computer Science

Deep Reinforcement Learning and Control

Causality

Katerina Fragkiadaki

Causal Confusion - Image Classification

``hat”

Causal Confusion- Image Classification

``hat”

Causal Confusion- Image Classification

Access to more information makes the problem harder: our model needs to understand what part of the input is related to the label

``hat”

Causal Confusion- Imitation Learning

Assume we have access to (the same) expert driving trajectories in the two setups:

Scenario A

Scenario B

The yellow brake square is an indicator of whether the driver is braking.

Consider training behaviour cloning on these two tasks.

Q: Do we expect one setup to have lower test error (per single time step) than the other?

Do we expect one setup to learn much better policies for deployment?

Causal Confusion in Imitation Learning,De Haan et al

Causal Confusion- Imitation Learning

Access to more information makes the problem harder: our model needs to understand what part of the input is related to the label.

Assume we have access to (the same) expert driving trajectories in the two setups:

Scenario A

Scenario B

It is easy for BC in setup A to obtain low train and test error by learning to press the brake whenever the yellow indicator is on.

Q: How will this policy perform during deployment?

Causal Confusion in Imitation Learning,De Haan et al

Causal Confusion- Imitation Learning

Access to more information makes the problem harder: our model needs to understand what part of the input is related to the label.

Assume we have access to (the same) expert driving trajectories in the two setups:

Scenario A

Scenario B

It is easy for BC in setup A to obtain low train and test error by learning to press the brake whenever the yellow indicator is ON. How will this policy perform during deployment?

Q: Would GAIL under setup A solve the problem?

Potentially, at a cost of much more interactions with the environment

Causal Confusion in Imitation Learning,De Haan et al

Causal Confusion- Imitation Learning

Access to more information makes the problem harder: our model needs to understand what part of the input is related to the label.

``hat”

- Identifying what are the contributing features for the desired output (driving action of image label) is crucial both for learning good policies and learning classifiers that generalize, we need to learn to ignore.

Causal Confusion in Imitation Learning,De Haan et al

Causal Confusion- Imitation Learning

Access to more information makes the problem harder: our model needs to understand what part of the input is related to the label.

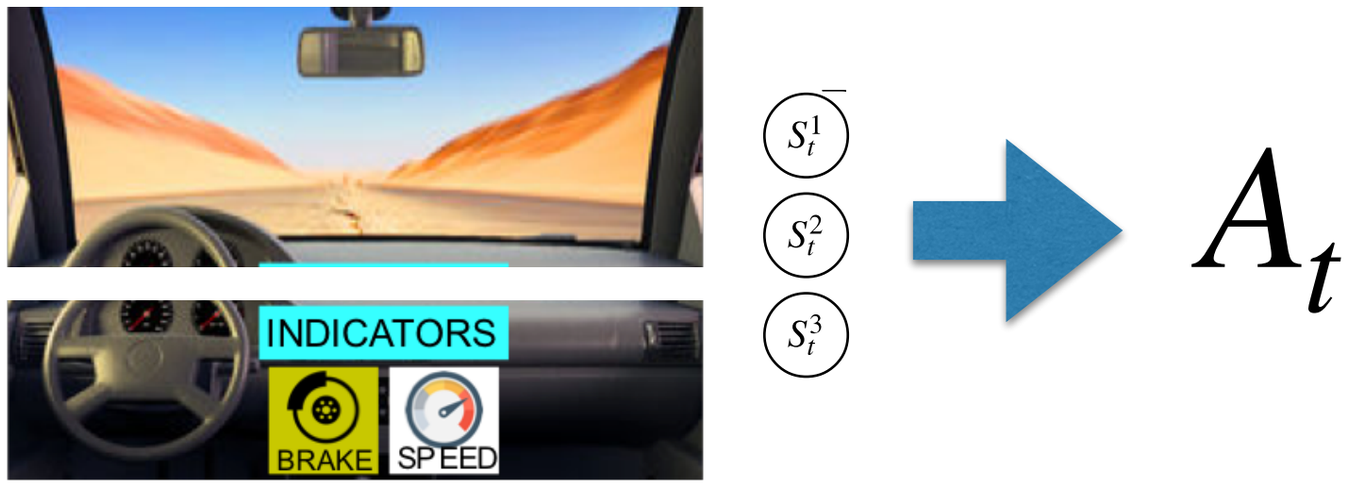
``hat”

- In the case of image classification, if you are supply enough data the classifier will eventually learn to attend to the right part.
- Q: In the case of BC, would more data solve the problem?
- because the objective is wrong (we don’t match trajectories, we match per timestep actions), supplying more data will not help.

Causal Confusion in Imitation Learning,De Haan et al

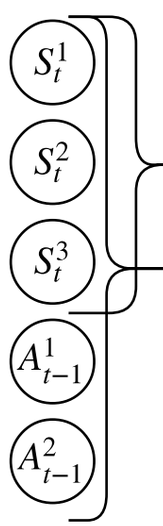
Disentangling Observations

- The disentangled state fetures are obtained with a variation on variational autoencoders, $\beta - VAE$



Causal Confusion in Imitation Learning,De Haan et al

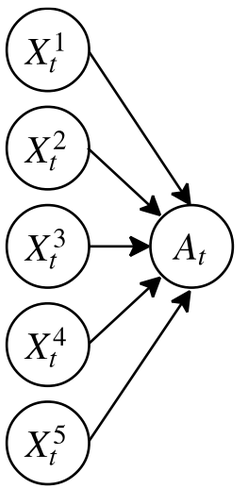
Confounding



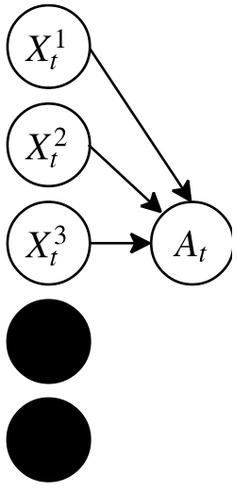
Original state

Confounded state

Task A:



Task B:
True causal graph

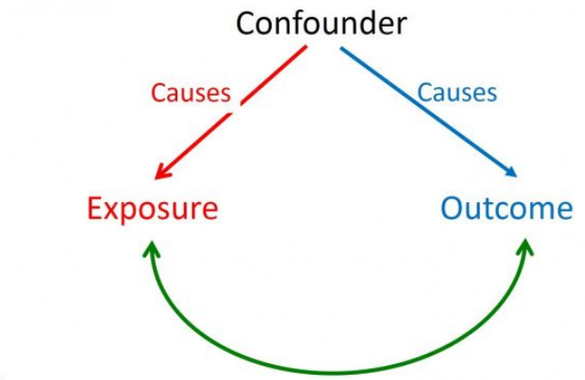


- You can easily confuse a behaviour cloning objective by supplying the past actions as part of the current state (used to regress to the next action).
- BC will prooly learn to copy paste the previous actions.
- Q: What is the problem with that?

Causal Confusion in Imitation Learning,De Haan et al

Confounding

Confounder: “an extraneous variable in an experimental design that correlates with both the dependent and independent variables”



AN OLD CLASSIC: MURDER AND ICE CREAM

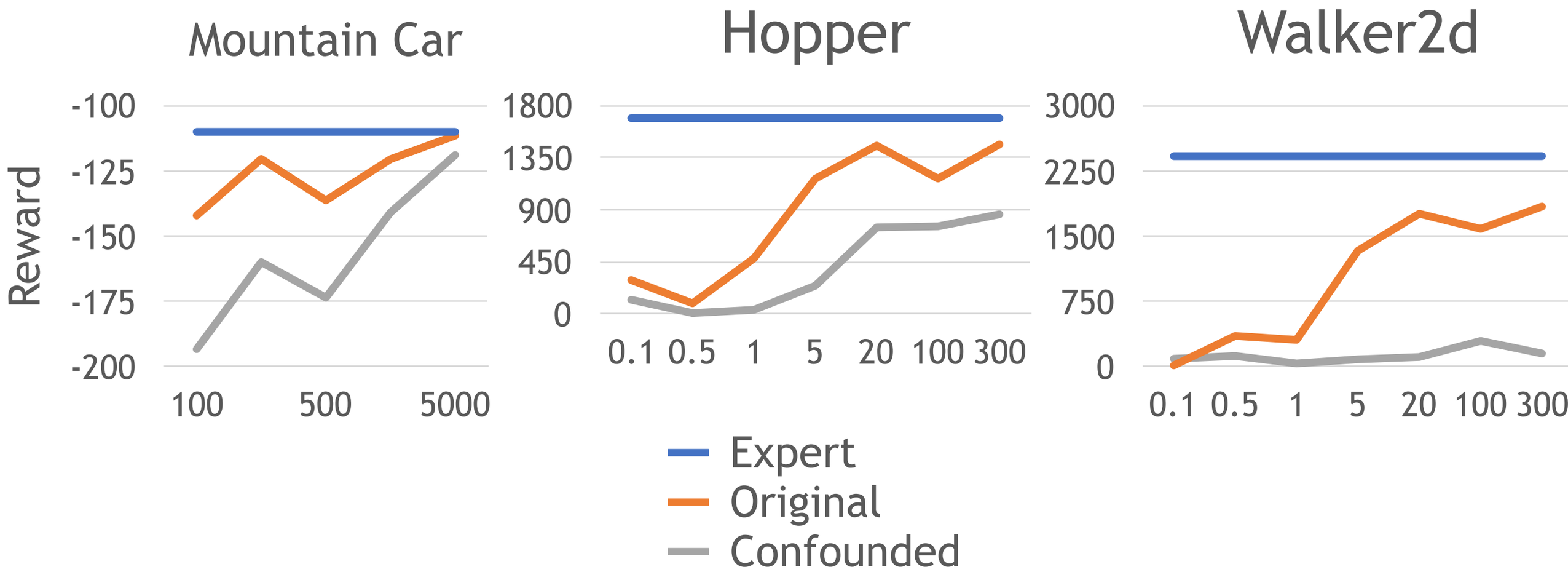
It is known that throughout the year, murder rates and ice cream sales are highly positively correlated. That is, as murder rates rise, so does the sale of ice cream. There are three possible explanations for this correlation:

Possibility #1: Murders cause people to purchase ice cream. One could imagine a world where this is true. Perhaps when one is murdered, they are resurrected as zombies who primarily feed on ice cream.

Possibility #2: Purchasing ice cream causes people to murder or get murdered. Again, one could imagine a world where this is true. Perhaps when one eats ice cream, those without ice cream become jealous and murder those with ice cream.

Possibility #3: There is a third variable—a confounding variable—which causes the increase in BOTH ice cream sales AND murder rates. For instance, the weather. When it’s cold and Wintery, people stay at home rather than go outside and murder people. They also probably don’t eat a lot of ice cream. When it’s hot and Summery, people spend more time outside interacting with each other, and hence are more likely to get into the kinds of situations that lead to murder. They are also probably buying ice cream, because nothing beats the sound of an ice cream truck on a blazing Summer day.

Confounding

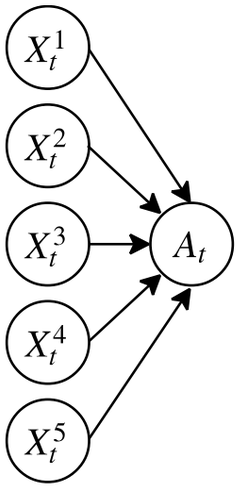


Causal Confusion in Imitation Learning,De Haan et al

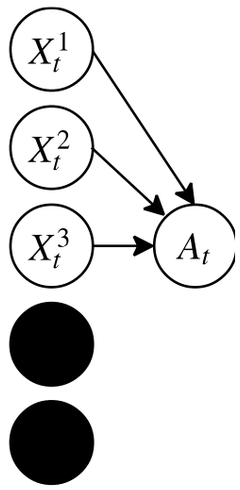
Finding the right causal structure

- There are 2^n combinations, hard to try one-by-one.

Task A:



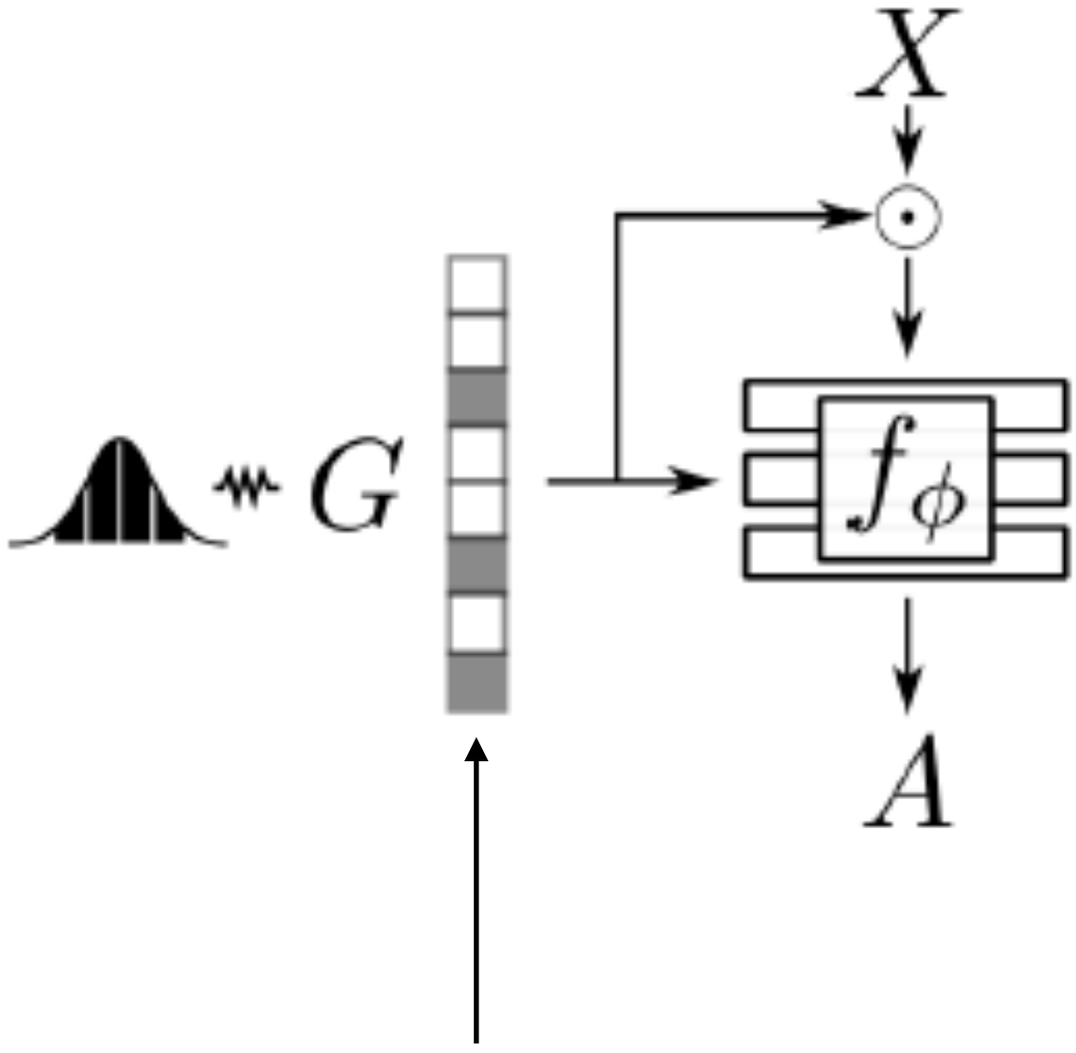
Task B:
True causal graph



Causal Confusion in Imitation Learning,De Haan et al

Learning a structure-parametrized policy

Learning a mapping from the mask, and the masked state features to the output action



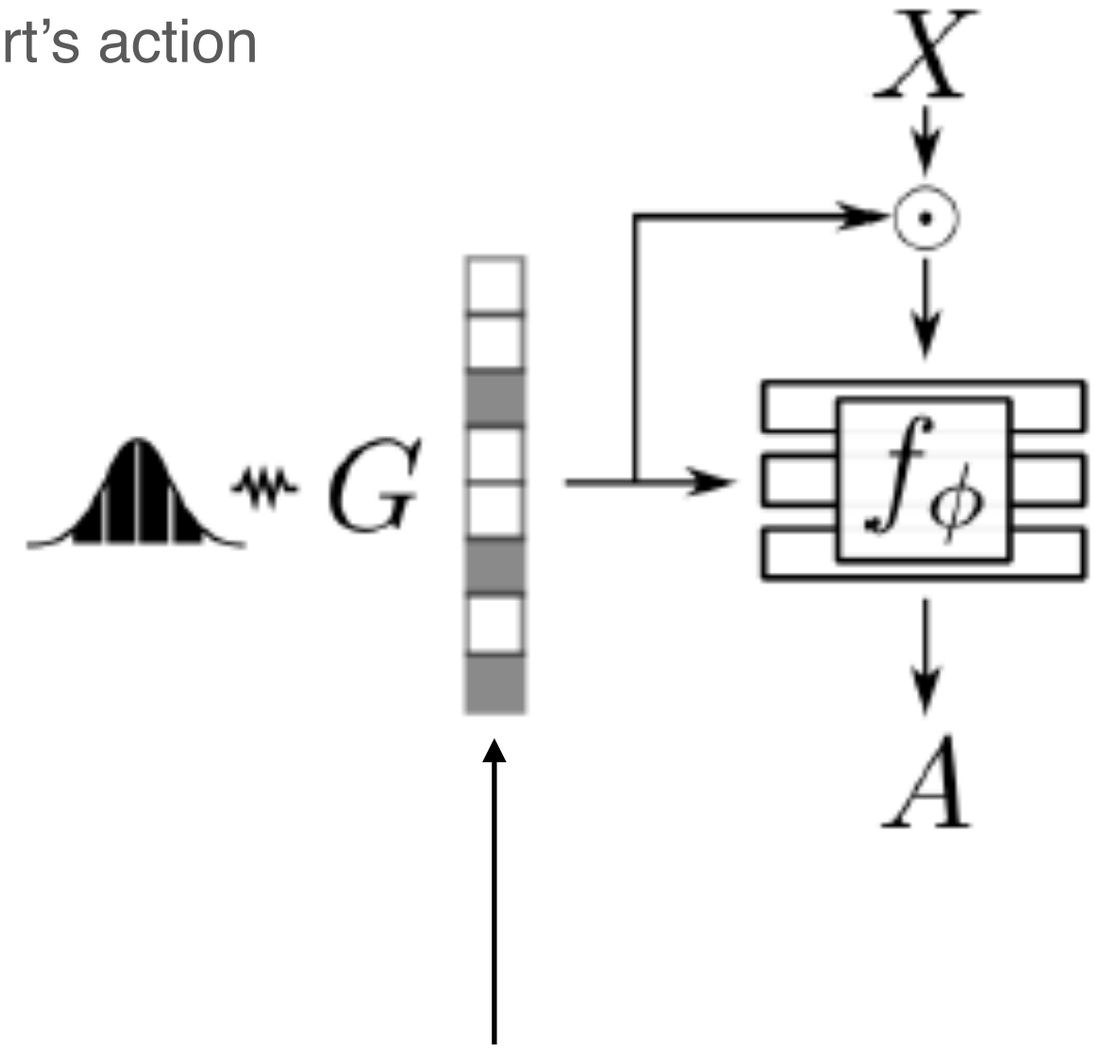
Sampling masking vectors and regressing to actions using the masked state we sampled.

Causal Confusion in Imitation Learning,De Haan et al

Learning a structure-parametrized policy

Input: the mask, and the masked state features

Output: the expert's action



Causal Confusion in Imitation Learning,De Haan et al

Finding the right causal structure

We need to find the right causal graph structure, in other words, the right state feature masking.

We will learn a mapping from graph structure to its likelihood.

We set the graph likelihood to be proportional to exponentiated reward: $p(G) \propto \exp\langle w, G\rangle$

$$p(G) = \prod_i p(G_i) = \prod_i \text{Bernoulli}(G_i|\sigma(w_i/\tau))$$

$$\begin{aligned}
\pi(G) &= \frac{\exp(\langle w, G\rangle + b)/\tau}{\sum_{G'} \exp(\langle w, G'\rangle + b)/\tau} \\
&= \frac{\exp\langle w, G\rangle/\tau}{\sum_{G'} \exp(\langle w, G'\rangle/\tau} \\
&= \frac{\prod_i \exp(w_i G_i/\tau)}{\sum_{G'} \prod_i \exp(w_i G'_i/\tau)} \\
&= \frac{\prod_i \exp(w_i G_i/\tau)}{\prod_i \sum_{G'_i} \exp(w_i G'_i/\tau)} \\
&= \prod_i \frac{\exp(w_i G_i/\tau)}{\sum_{G'_i} \exp(w_i G'_i/\tau)} \\
&= \prod_i \frac{\exp(w_i G_i/\tau)}{1 + \exp w_i/\tau} = \prod_i \text{Bernoulli}(G_i|\sigma(w_i/\tau))
\end{aligned}$$

Causal Confusion in Imitation Learning,De Haan et al

Finding the right causal structure

We need to find the right causal graph structure, in other words, the right state feature masking.

We will learn a mapping from graph structure to its likelihood.

We set the graph likelihood to be proportional to exponentiated reward:

$$p(G) = \prod_i p(G_i) = \prod_i \text{Bernoulli}(G_i | \sigma(w_i/\tau))$$

To obtain task rewards we will be deploying the correpsonding policy, obtaining trajectories and scoring their rewards.

Algorithm 2 Policy execution intervention

Input: policy network f_ϕ s.t. $\pi_G(X) = f_\phi([X \odot G, G])$
Initialize $w = 0, \mathcal{D} = \emptyset$.
for $i = 1 \ldots N$ **do**
  Sample $G \sim p(G) \propto \exp\langle w, G \rangle$.
  Collect episode return R_G by executing π_G.
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(G, R_G)\}$
  Fit w on $\mathcal{D}$ with linear regression.
end for
Return: $\arg\max_G p(G)$

Causal Confusion in Imitation Learning,De Haan et al

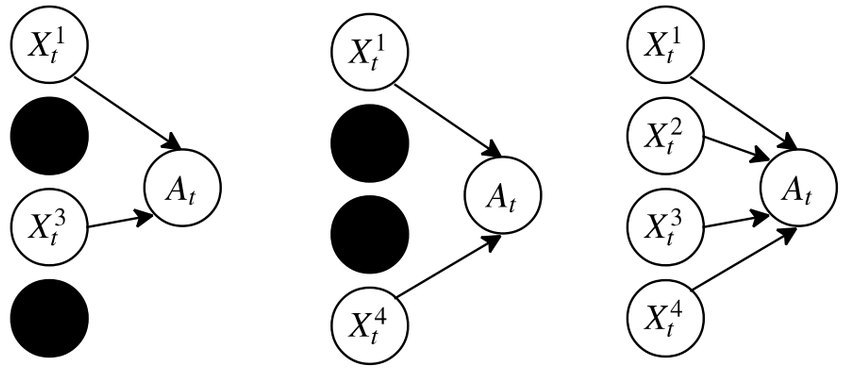
Finding the right causal structure

We need to find the right causal graph structure, else, the right state feature masking.

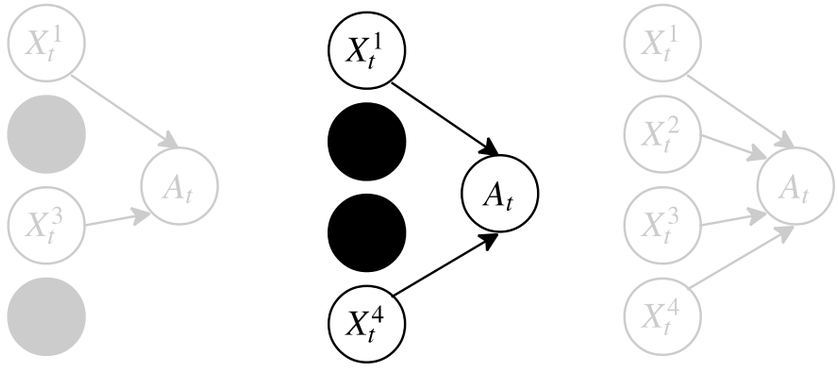
We will learn a mapping from graph structure to task reward

To obtain task rewards we will be deploying the correpsonding policy, obtaining trajectories and scoring their rewards.

1. Passive discovery: Find all graphs and policies consistent with data



2. Targeted intervention: Find true graph
   - Rewards
   - Expert queries



Causal Confusion in Imitation Learning,De Haan et al