A Theory of PAC Learnability under Transformation Invariances

Han Shao, Omar Montasser and Avrim Blum

TTIC

November 6, 2022

Introduction

- (Group) transformation invariances are present in many real-world problems. E.g.,
  - Image classification is usually invariant to rotation/flip/color transformation.
  - Syntax parsing is invariant to exchange of noun phrases in a sentence.

Introduction

- (Group) transformation invariances are present in many real-world problems. E.g.,
  - Image classification is usually invariant to rotation/flip/color transformation.
  - Syntax parsing is invariant to exchange of noun phrases in a sentence.
- Data augmentation is one commonly used technique.
  - Add the transformed data into the training set.
  - Trains a model on the augmented data.

Introduction

- (Group) transformation invariances are present in many real-world problems. E.g.,
  - Image classification is usually invariant to rotation/flip/color transformation.
  - Syntax parsing is invariant to exchange of noun phrases in a sentence.
- Data augmentation is one commonly used technique.
  - Add the transformed data into the training set.
  - Trains a model on the augmented data.

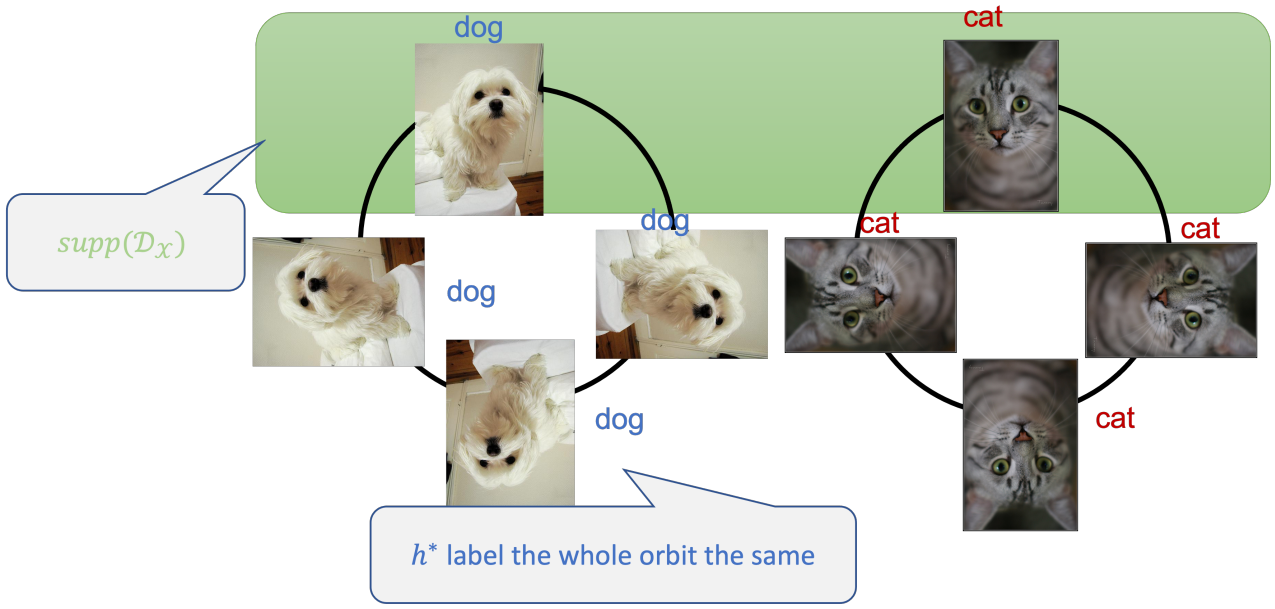
How does data augmentation perform theoretically?
What is the optimal algorithm in terms of sample complexity under transformation invariances?

Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.

Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.



Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.

Main results

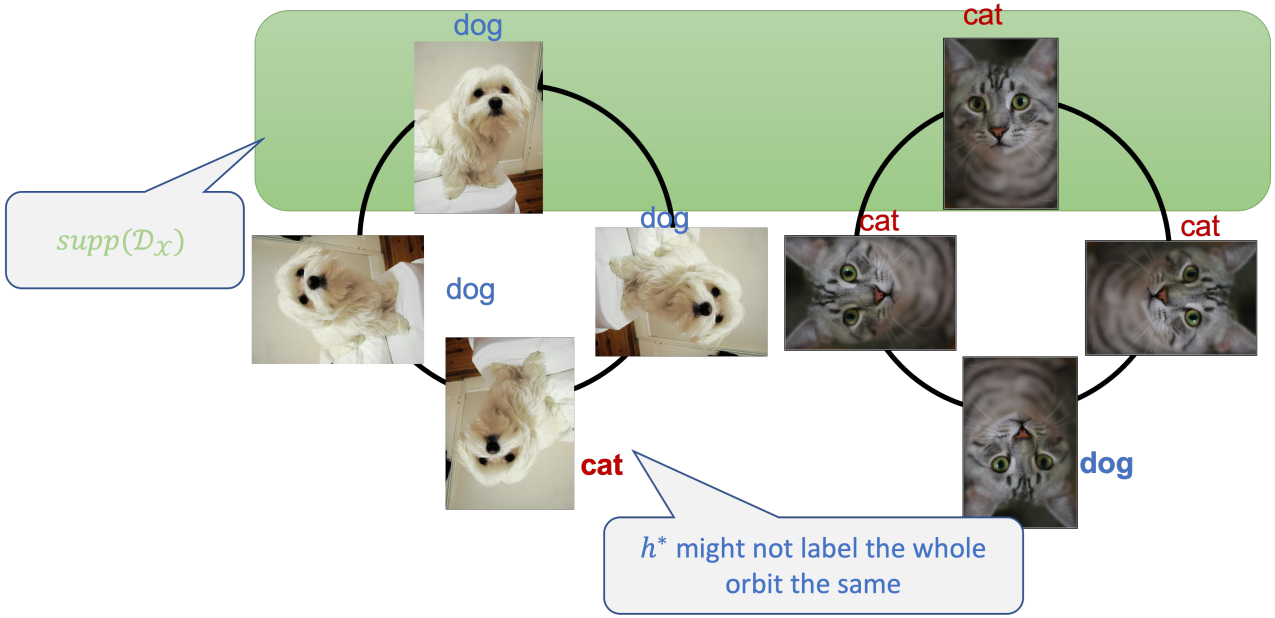
- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.

Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.
- **Relaxed realizable setting:** $\exists h^* \in \mathcal{H}$ such h^* has zero error over the support of the data distribution.

Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.
- **Relaxed realizable setting:** $\exists h^* \in \mathcal{H}$ such h^* has zero error over the support of the data distribution.



Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.
- **Relaxed realizable setting:** $\exists h^* \in \mathcal{H}$ such h^* has zero error over the support of the data distribution.
  - *DA can hurt.* Any algorithm not distinguishing the original data from the transformed data hurt. The optimal sample complexity of this family is characterized by $\mu(\mathcal{H}, \mathcal{G})$.

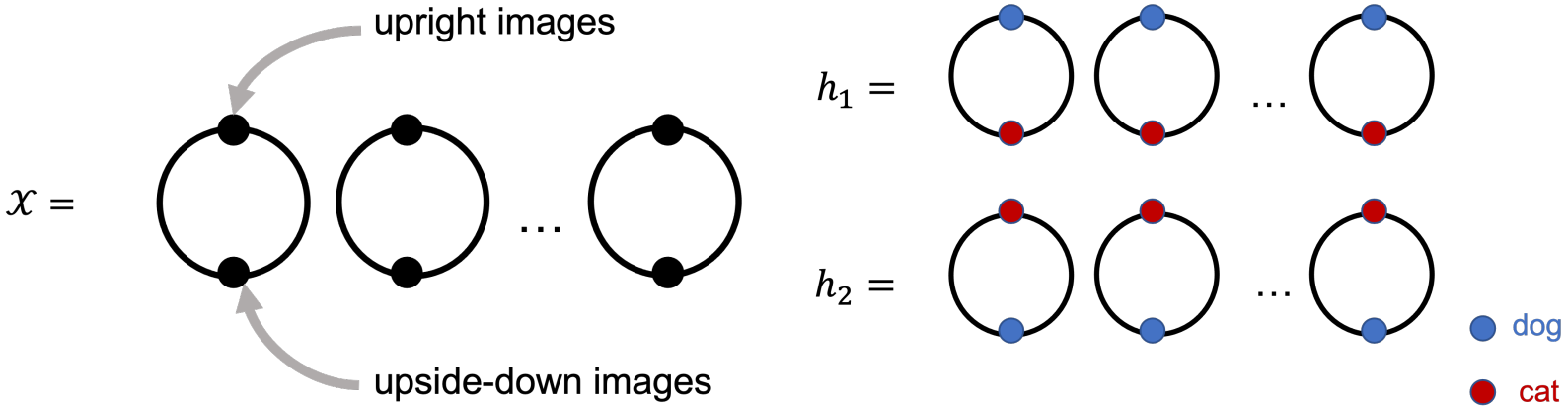
Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.
- **Relaxed realizable setting:** $\exists h^* \in \mathcal{H}$ such h^* has zero error over the support of the data distribution.
  - *DA can hurt.* Any algorithm not distinguishing the original data from the transformed data hurt. The optimal sample complexity of this family is characterized by $\mu(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.

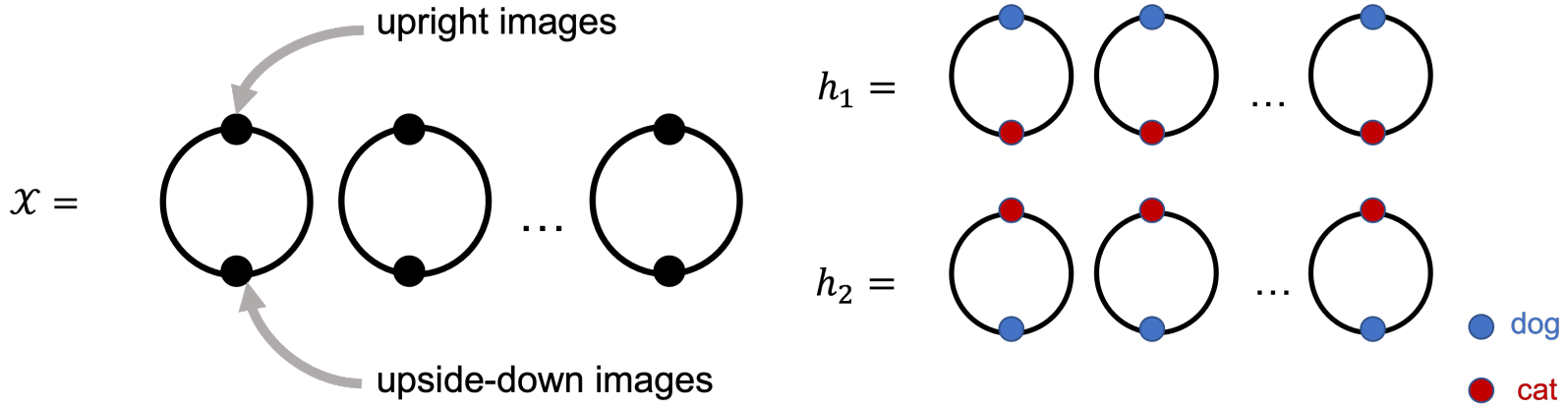
Main results

- **Invariantly realizable setting:** $\exists h^* \in \mathcal{H}$ s.t. h^* can correctly classify not only the natural data but also the transformed data.
  - *DA helps but is not optimal.* The sample complexity of DA is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{o}}(\mathcal{H}, \mathcal{G})$.
- **Relaxed realizable setting:** $\exists h^* \in \mathcal{H}$ such h^* has zero error over the support of the data distribution.
  - *DA can hurt.* Any algorithm not distinguishing the original data from the transformed data hurt. The optimal sample complexity of this family is characterized by $\mu(\mathcal{H}, \mathcal{G})$.
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.
- **Agnostic setting**
  - The optimal sample complexity is characterized by $\mathrm{VC}_{\mathrm{ao}}(\mathcal{H}, \mathcal{G})$.

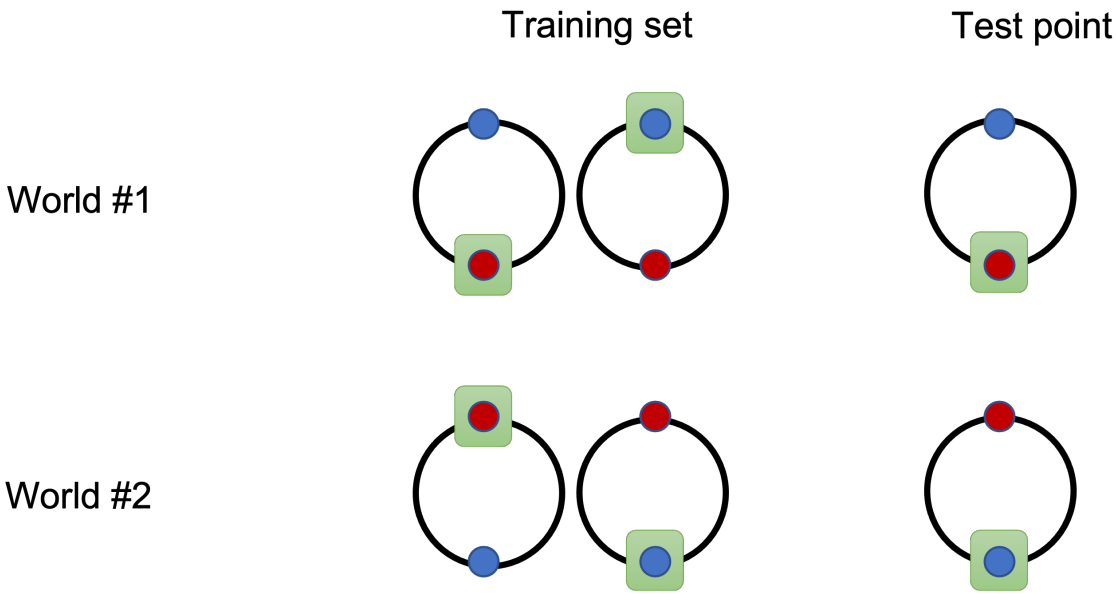
An example of DA hurts



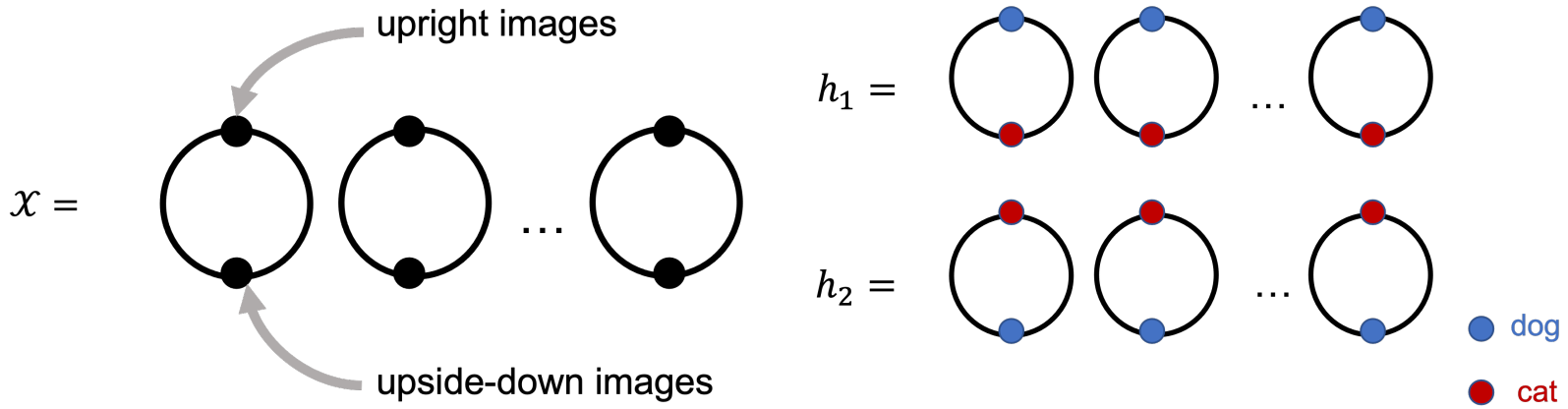
An example of DA hurts



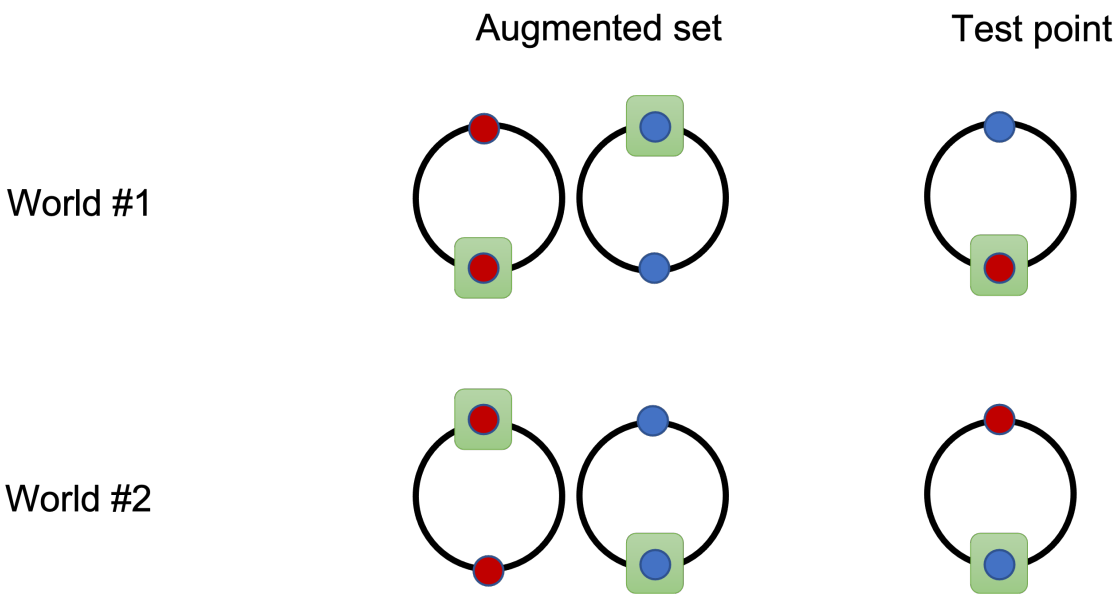
The natural data only has upright dogs and upside-down cats or only has upright cats and upside-down dogs.



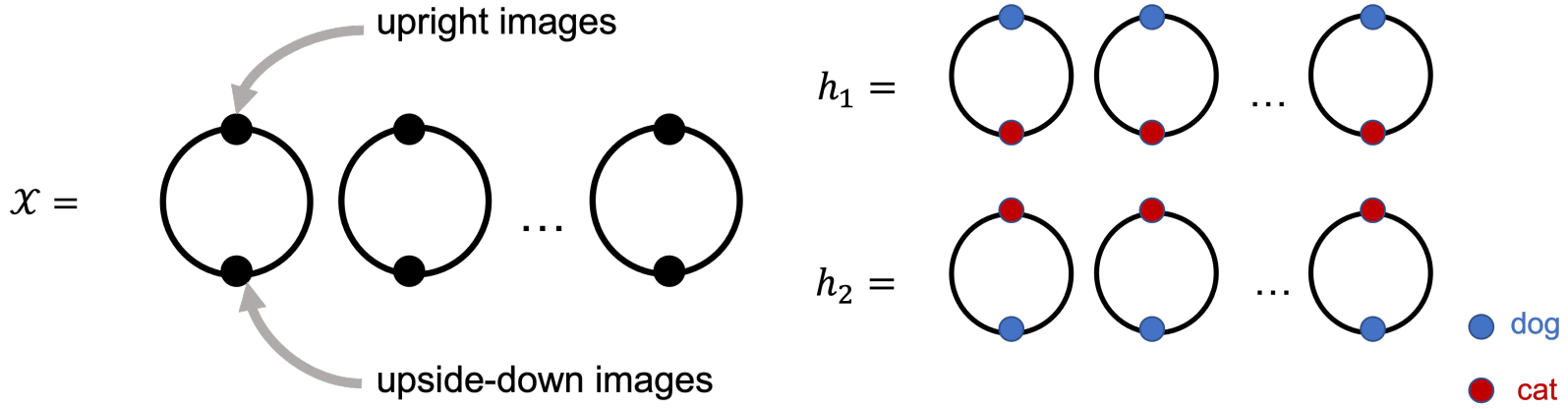
An example of DA hurts



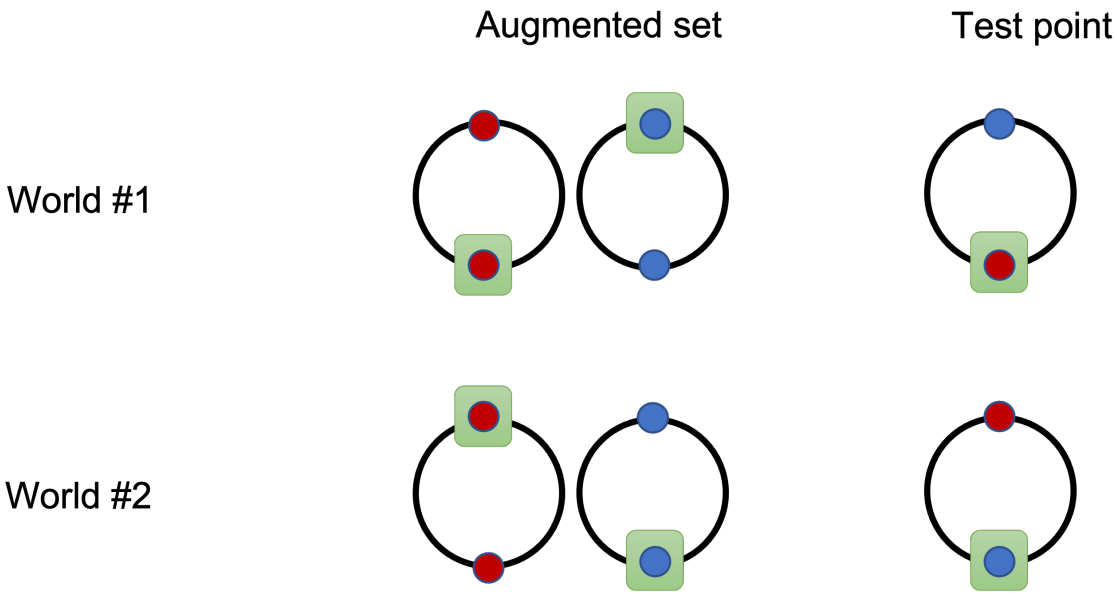
The natural data only has upright dogs and upside-down cats or only has upright cats and upside-down dogs.



An example of DA hurts



The natural data only has upright dogs and upside-down cats or only has upright cats and upside-down dogs.



Distinguishing between original and transformed data is important!

Come to our poster for more results!