Meet-up 2019 | Doctorants & Industrie

CONTROLLED REPRESENTATION LEARNING: TOWARDS MORE RELIABLE MACHINE LEARNING

Victor BOUVIER[1,2]

Céline Hudelot[1], Myriam Tami[1], Clément Chastagnol[2], Philippe Very

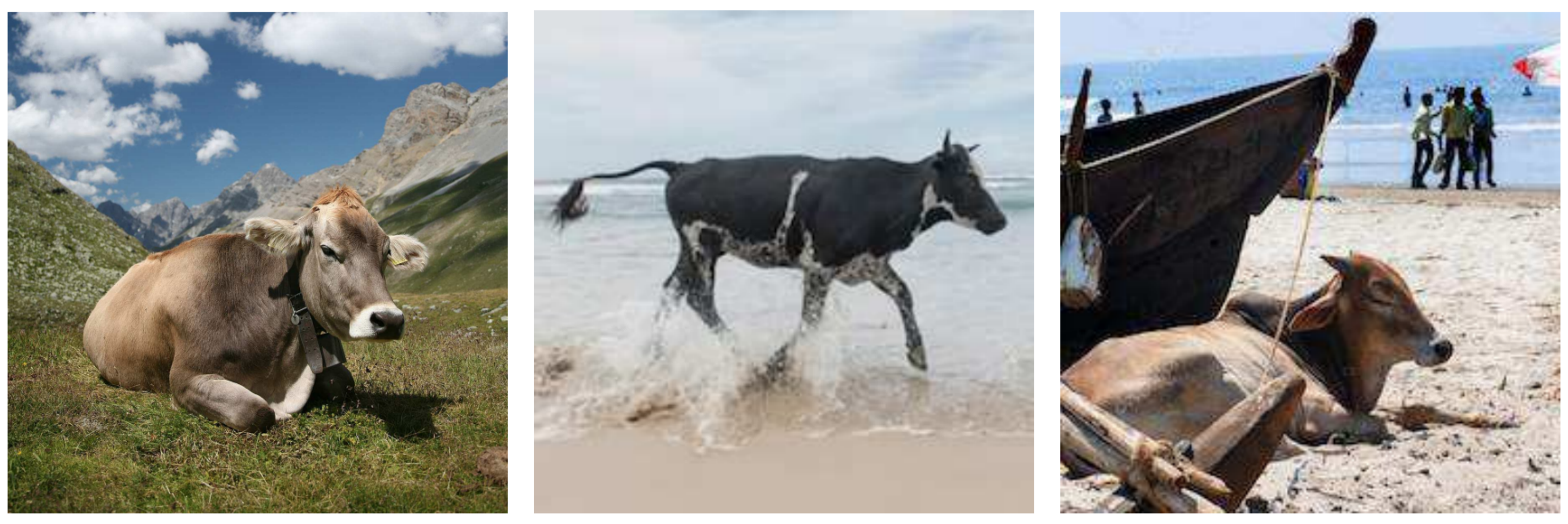
[1] CentraleSupélec, Université Paris-Saclay [2] Sidetrade

Abstract

Provided that a large amount of annotated data exists, deep learning allows valuable information to be derived from high dimensional data, such as text or images, opening the way to a wide range of applications in many industrial sectors. The main ingredient of this success is its ability to extract and infer on a huge amount of weak statistical patterns. Quite surprisingly, this strength can become a weakness in an industrial context; many factors may change in a production environment and their impact on deep models remains poorly understood. The research hypothesis of this thesis work is to search for stable properties under varying modalities. The promise of this approach is to infer on an intrinsic property of the task rather than on spurious, or too specific, statistical patterns. We studied the connections of our research hypothesis with domain adaptation - a well-established line of study which uses unlabelled data to adapt models to new environments. We have shown that our new point of view makes it possible to formulate a generic way of dealing with adaptation. Our claim is supported both theoretically and empirically and adds a step towards unifying two lines of study (Importance Sampling and Invariant Representation) by discussing the role of compression of the representations. In future work, we want to exhibit the limit of domain invariant representations for learning invariant predictors. More precisely, we believe that domain invariance is insufficient and provides only very sparse information (binary) to identify stable properties in the data. To enrich information, we aim to develop Active Learning strategies to quickly and cost-effectively identify sources of spurious correlations in a context where unlabelled production data are available. Ultimately, we wish to connect our research with counterfactual analysis, which has strategic applications in industry, particularly in decision-making. The potential applications of our research are extensive and include Transfer Learning, Robustness, Privacy and Fairness.

1 THE CURSE OF SUPERVISED LEARNING

High specialization degrades **True Generalization**



(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Figure 1: Recognition in Terra Incognita [3]

- Deep models learn **spurious** correlation
- Not robust to new environments or domains

Providing reliable model in production is a major challenge in Machine Learning

2 CONTROLLED REPRESENTATION LEARNING

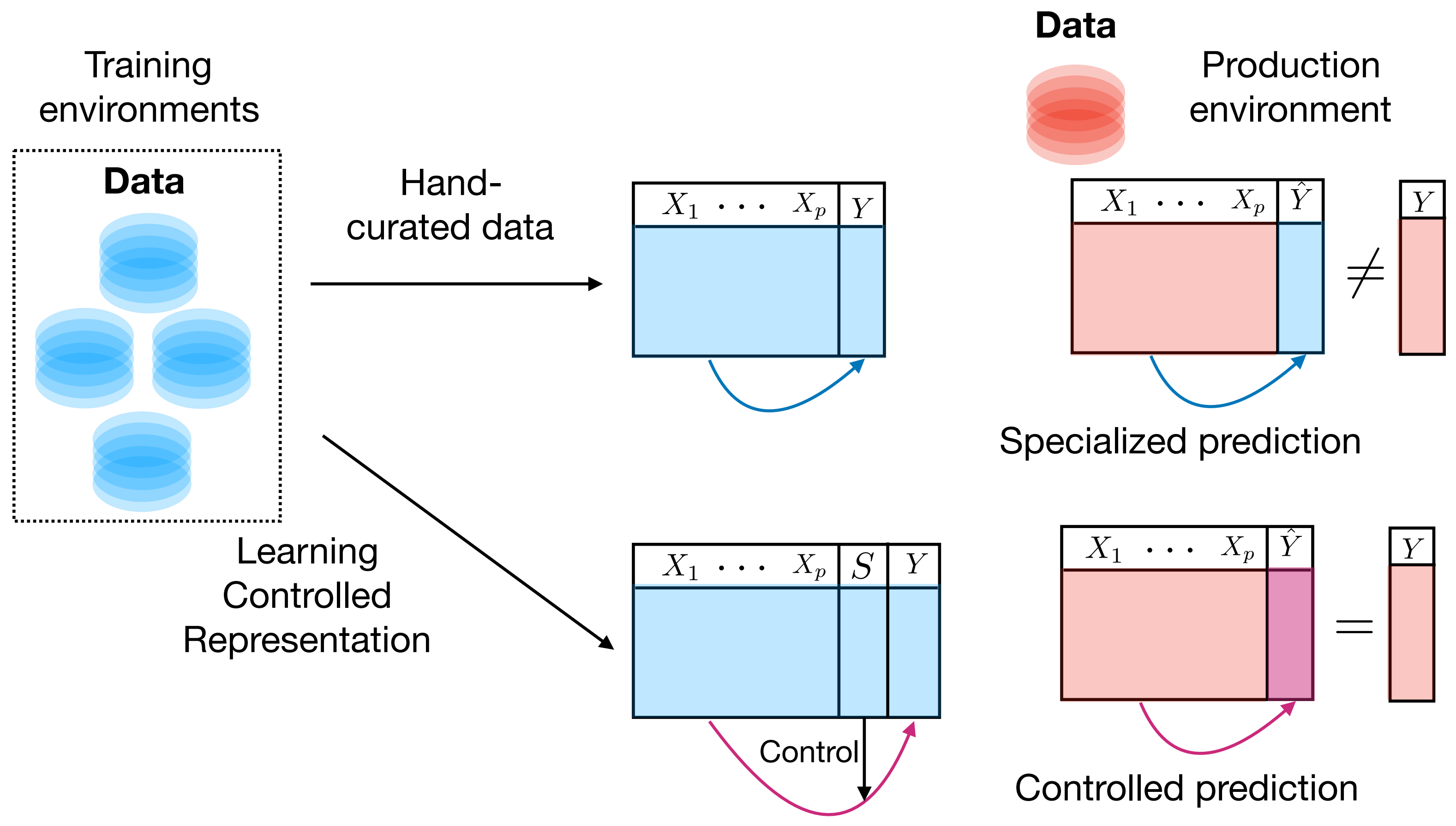


Figure 2: Supervised Learning VS Controlled Representation Learning

- Source environment ($\mathcal{S}$): **prior knowledge**
  What a reliable model must remain insensitive to?
- $\mathbb{E}[Y|Z]$ is invariant w.r.t. $\mathcal{S}$. [2].

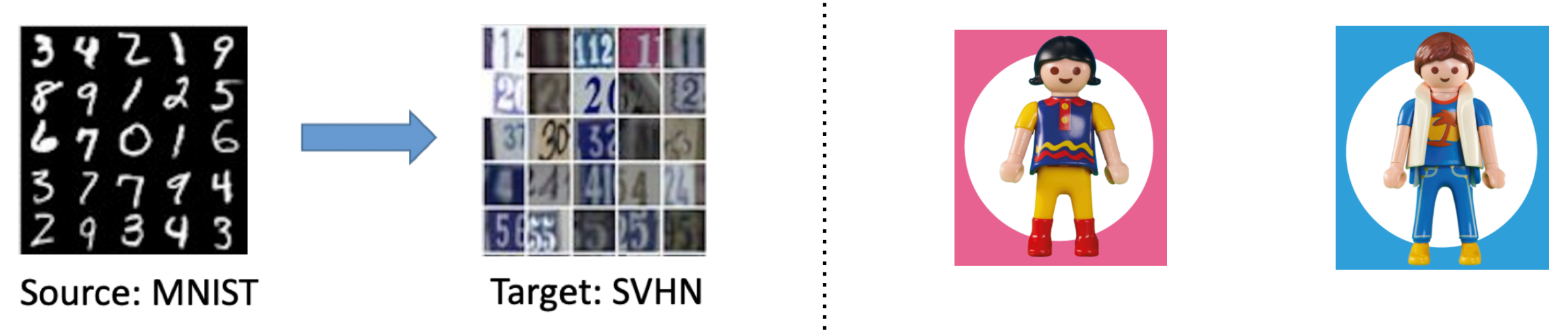
3 CONTRIBUTIONS

1. **Is a representation controlled?** [5]
   - Filtering strategy for control testing
   - Generic method for introducing bias in dataset
2. **How to learn a controlled representation?** [4]
   - Supervised and Unsupervised manner
   - Binary source case: have demonstrated better generalization properties than traditional methods.
3. **What is a good controlled representation?** [1]
   - The representations that best preserve the original feature space have the best guarantees
   - Role of *Weighted Representations*

4 APPLICATIONS

Find a representation $Z = \varphi(X)$ such that:

$\mathcal{S} = \{\text{source}, \text{target}\}$ $\mathcal{S} = \{\text{female}, \text{male}\}$



$\mathbb{E}^{\text{source}}[Y|Z] = \mathbb{E}^{\text{target}}[Y|Z]$ $\mathbb{E}^{\text{male}}[Y|Z] = \mathbb{E}^{\text{female}}[Y|Z]$

Domain Adaptation **Fair ML**

5 FUTURE WORK

1. Theoretical fundations of CRL
2. How to select controlled representations?
3. Beyond the binary source case
4. How to actively select samples?
5. Open source python library

```python
from csl import InvariantConditional

X, Y, S = data()

model = InvariantConditional()
model.fit(features=X, labels=Y, sources=S)
Z = model.control(X)
```

References

[1] Anonymous. Domain-invariant representations: A look on compression and weights. In *Submitted to International Conference on Learning Representations*, 2020. under review.

[2] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

[3] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 456–473, 2018.

[4] Victor Bouvier, Philippe Very, Céline Hudelot, and Clément Chastagnol. Hidden covariate shift: A minimal assumption for domain adaptation. *arXiv preprint arXiv:1907.12299*, 2019.

[5] Victor Bouvier, Philippe Very, Céline Hudelot, and Clément Chastagnol. Learning invariant representations for sentiment analysis: The missing material is datasets. *arXiv preprint arXiv:1907.12305*, 2019.

Acknowledgements

We would like to thank Sidetrade and ANRT for fundings.

Interface - Digicosme
CentraleSupelec
Victor Bouvier- vbouvier@sidetrade.com
CIFRE with Sidetrade