Domain shift robustness in deep learning

Semester project

October, 2023

The successful application of deep learning methods in fields such as medicine, biology or mobility research critically depends on the ability of the predictor to account for domain shifts, i.e., its ability to make robust predictions under change of enviroments and distribution. For instance, in computational biology, a model can be trained to predict the state of a cell type from its transcriptomic profile (one environment), and then be deployed to predict the state of another cell type, i.e., a second environment with possible domain shift wrt the distribution of the transcriptomic profile. If the predictor has been trained with a conventional cross-entropy loss, it might not be able to accurately make predictions.

The goal of this project is to derive a novel objective that protects against distributional shifts. Similarly to Gong et al. (2016) and Heinze-Deml and Meinshausen (2021), we shall exploit the fact that in prediction tasks the covariables can often be distinguished in conditionally invariant features C whose distribution $P(C|Y)$ does not change significantly across environments given an output Y, and style features S whose distribution $P(S|Y)$ might change in different environments. From a causal perspective, we model the relationship between features $X = (C, S)$ and outputs Y in an causal-anticausal framework where the response Y causes the features X and we are interested in predicting the anticausal direction $X \to Y$.

This is exemplified in Figure 1 which shows the causal models from Gong et al. (2016) and Heinze-Deml and Meinshausen (2021), respectively. Both models assume that the effect of an environment D on style features S is mediated via a latent noise variable Δ. Interventions on Δ are ultimately causing the distributional shift of X. Assuming that the core features are carrying the actual signal to predict a response Y, if one can identify the core features successfully, then distributional robust predictions can be made.

For this project, the student will implement deep neural network models for robust prediction under domain shift (Heinze-Deml and Meinshausen (2021), Tachet des Combes et al. (2020), Gong et al. (2016)) and evaluate them on several real-world data sets. We will first evaluate the method on data sets that are commonly used for classification, e.g., MNIST and CelebA, before extending it for regression tasks, i.e., when the response variables are continuous, and possibly time series settings.

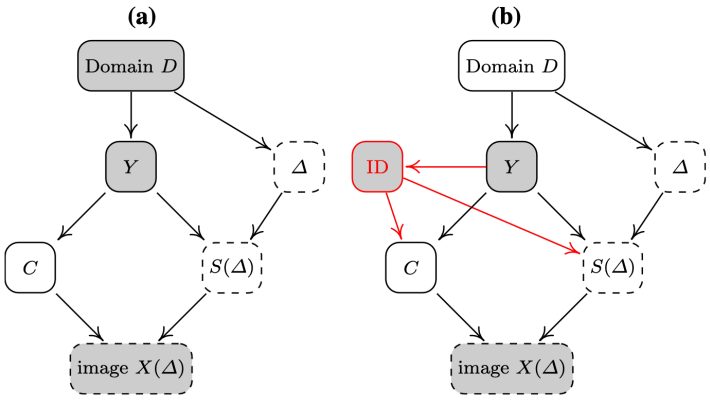

Figure 1: Data generating processes for the models of Gong et al. (2016) and Heinze-Deml and Meinshausen (2021), respectively. Transparent nodes are latent, shared nodes are observed. Heinze-Deml and Meinshausen (2021) introduce an additional ID variable that can be used to group observations.

Specifically, the student will

- conduct a literature study to acquire the needed background in deep learning and (causal) distributional robustness,
- implement the method using the Python libraries Flax and JAX (Bradbury et al. (2018), Babuschkin et al. (2020), Heek et al. (2023)),
- test the method on conventional benchmark data sets,
- help extending the method to regression and possibly time series settings.

Additional Information

- **What will you learn?**
  - Basic understanding of distributional robustness from a causal perspective
  - Fundamentals of deep learning
  - Google Deepmind JAX ecosystem (JAX, Flax, Optax)
- **Requirements**
  - capability to read, understand, and implement research papers
  - strong Python programming skills, experience with JAX-ecosystem helpful
  - background in statistics and deep learning
- **Difficulty** from easy to moderate
- **Supervisor** Simon Dirmeier (simon dot dirmeier at sdsc dot ethz dot ch) and Prof Dr Fernando Perez-Cruz

References

Babuschkin, Igor, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, et al. 2020. *The DeepMind JAX Ecosystem*. http://github.com/deepmind.

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, et al. 2018. *JAX: Composable Transformations of Python+NumPy Programs*. http://github.com/google/jax.

Gong, Mingming, Kun Zhang, Tongliang Liu, Dacheng Tao, Clark Glymour, and Bernhard Schölkopf. 2016. "Domain Adaptation with Conditional Transferable Components." In *Proceedings of the 33rd International Conference on Machine Learning*.

Heek, Jonathan, Anselm Levskaya, Avital Oliver, Marvin Ritter, Bertrand Rondepierre, Andreas Steiner, and Marc van Zee. 2023. *Flax: A Neural Network Library and Ecosystem for JAX*. http://github.com/google/flax.

Heinze-Deml, Christina, and Nicolai Meinshausen. 2021. "Conditional Variance Penalties and Domain Shift Robustness." *Machine Learning* 110 (2): 303–48.

Tachet des Combes, Remi, Han Zhao, Yu-Xiang Wang, and Geoff Gordon. 2020. "Domain Adaptation with Conditional Distribution Matching and Generalized Label Shift." In *Advances in Neural Information Processing Systems*.