STAT 8101: Applied Causality (Spring'22) Week 12: Apr 12, 2022 - Apr 18, 2022

Invariance Risk Minimization

Han Lin hl3199@columbia.edu MS in Computer Science

- I read the non-linear IRM paper before this one, seems that I should do the opposite. This paper provides a good way of solving spurious correlations problem under multiple environments. In section 3, the authors formulate such goal as a constrained optimization problem, and then proposed a practical version of this algorithm (IRM-v1), which put the constraint as a penalty term, and use gradient norm to measure the optimality of the dummy classifier at each environment.
- For linear classifier w, the author further states how to choose the penalty term D. It's interesting to see IRM could be generalized with general loss function, and multiple outputs rather than binary, this might make it a useful tool for different problems (for example, solving causal confusions in imitation learning).
- To compare with iCaRL, Table1 in Lu et al. tests iCaRL and IRM on nonlinear classifier. IRM is pretty well on linear case, but could not perform very well on nonlinear case where iCaRL outperforms with a large margin. I reflected on such gap, and wonder which specific component (phase1, 2 or 3) mentioned in Lu et al's paper makes such difference. I guess it would be better if Lu et al. could do some ablation tests (for example, remove phase 2 to see whether the performance drop is significant).
- In the experiment section, the author tests on the colored MNIST dataset with non-linear classifier w. This gives some insight of how to use neural networks together with IRM.
- From a very practical point of view, I'm interested in how to choose the penalty weight λ in IRM-v1. Their code for colored MNIST first set $\lambda = 1$ for the first 100 iterations, and then set as 1e4 after 100 iterations. It seems that such setting could make the model first learn a reasonable classifier, then decrease its dependence on spurious correlations.