

out-of-domain accuracy

0.834
0.832
0.830
0.828
0.826
0.824

- constant
- top all features
- top causal features
- top arguably causal features
- top anticausal features

0.83 0.84 0.85 0.86 0.87 0.88

in-domain accuracy

