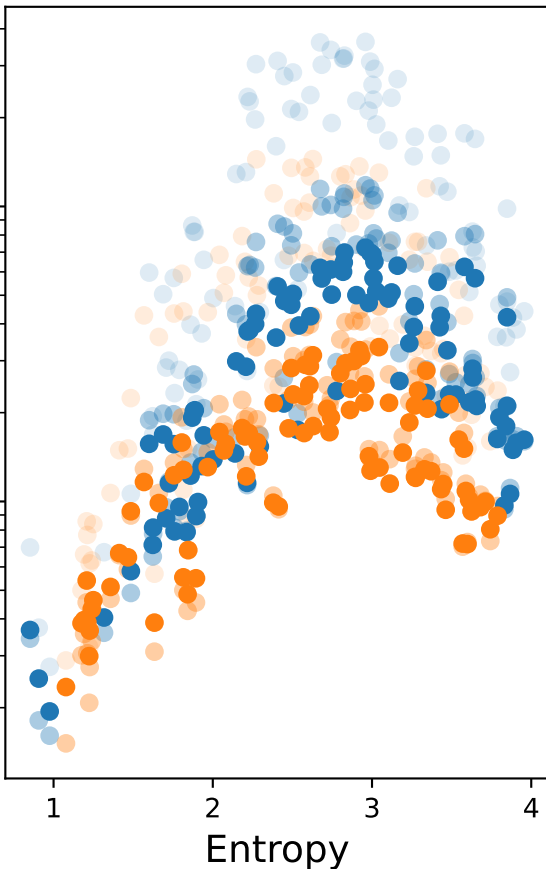


KL Divergence



- dense, model $d = 32$
- sparse, model $d = 32$
- dense, model $d = 128$
- sparse, model $d = 128$
- dense, model $d = 512$
- sparse, model $d = 512$