

KL Divergence

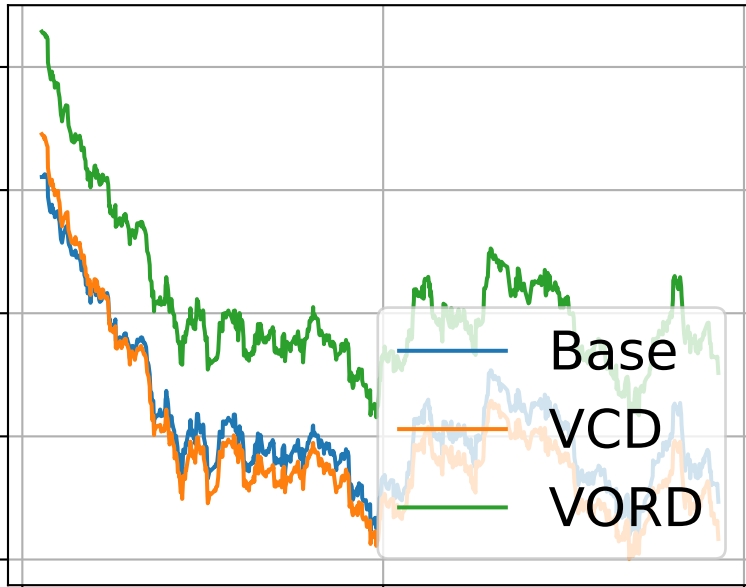
1.8
1.7
1.6
1.5
1.4

0

5000
Steps

10000

Base
VCD
VORD



The graph displays the KL Divergence on the y-axis (ranging from 1.4 to 1.8) against the number of training steps on the x-axis (ranging from 0 to 10,000). Three models are compared: Base (blue), VCD (orange), and VORD (green). All models show a general downward trend in KL Divergence as training progresses. The VORD model starts with the highest divergence (~1.83) and ends with the highest (~1.58). The Base model starts at ~1.71 and ends at ~1.46. The VCD model starts at ~1.75 and ends at ~1.42. A vertical line at 5,000 steps and a shaded region from 5,000 to 10,000 steps highlight the later stages of training.

Steps	Base	VCD	VORD
0	1.71	1.75	1.83
1000	1.58	1.57	1.68
2000	1.50	1.48	1.60
3000	1.48	1.47	1.58
4000	1.49	1.48	1.57
5000	1.46	1.44	1.55
6000	1.50	1.48	1.60
7000	1.52	1.50	1.62
8000	1.48	1.45	1.58
9000	1.50	1.48	1.60
10000	1.46	1.42	1.58