

Table 1: Performance comparison of different model methods under various configurations. AVG represents average perplexity (lower is better) or average correct score rate. No-D: DPLR with no decay, Sc-D: DPLR with scalar decay, Ve-D: DPLR with vector decay.

Method	Pa	Loss	PPL ↓			Accuracy ↑						
			Wiki	LMB	AVG	BOQA	PIQA	Hella	Wino	ARC-e	ARC-c	OB
No-D	0.16	3.000	41.1	120.0	80.6	56.4	62.7	31.7	49.5	52.9	26.6	31
Sc-D-0	0.16	2.965	40.7	121.8	81.3	60.1	64.1	32.7	50.0	53.6	25.3	30
Sc-D-0.99	0.16	2.941	39.0	103.7	71.4	60.1	64.0	33.0	50.9	53.8	25.0	31
Ve-D-0	0.17	2.937	39.0	83.4	61.2	61.0	63.9	33.7	50.6	54.9	25.2	31
Ve-D-0.99	0.17	2.920	37.9	73.5	55.7	60.1	64.9	33.8	48.2	53.6	25.3	30
No-D	0.42	2.773	30.8	58.3	44.6	59.7	66.6	37.7	50.7	58.0	28.0	32
Sc-D-0	0.42	2.736	30.1	56.1	43.1	58.4	67.6	39.4	51.9	58.4	27.1	33
Sc-D-0.99	0.42	2.717	29.1	46.0	37.5	61.1	67.6	39.5	51.1	61.2	29.7	34
Ve-D-0	0.42	2.732	29.3	45.2	37.3	61.0	67.4	39.7	50.4	59.6	29.5	32
Ve-D-0.99	0.42	2.719	28.5	43.3	35.9	60.7	67.0	40.0	50.3	60.3	27.7	34