

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- $p$ : DPLR with scalar decay (simple decay style with value  $p$ ), Ve-D: DPLR with vector decay (simple decay style with value  $p$ ).

Method	Pa	Loss	PPL ↓			Accuracy ↑							
			Wiki	LMB	AVG	BOQA	PIQQA	Hella	Wino	ARC-e	ARC-c	OB	
No-D	1.45	2.591	23.7	31.1	27.4	61.3	69.3	44.3	53.0	65.3	31.2	34	
Sc-D-0	1.45	2.523	23.1	26.6	24.8	61.5	70.0	47.1	53.1	65.4	33.1	35	
Sc-D-0.99	1.45	2.507	22.4	23.1	22.8	61.4	71.0	47.4	53.8	65.5	31.9	36	
Ve-D-0	1.47	2.508	22.5	22.3	22.4	60.8	69.6	48.1	53.5	66.9	32.7	36	
Ve-D-0.99	1.47	2.498	22.0	21.2	21.6	60.9	69.8	48.4	54.3	66.5	32.6	34	