	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent	per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22	32	12000	3.00	128	3.00E-05														
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	90.0	0.840	1.07	0.847	1.07	32	12000	6.00	128	3.00E-05		1	0.9	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	60.1	0.839	1.08	0.838	1.08	32	12000	6.00	128	3.00E-05		1	0.6	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	40.1	0.840	1.12	0.841	1.12	32	12000	6.00	128	3.00E-05		1	0.4	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	30.1	0.832	1.18	0.835	1.18	32	12000	6.00	128	3.00E-05		1	0.3	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	20.1	0.816	1.28	0.827	1.28	32	12000	6.00	128	3.00E-05		1	0.2	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	10.1	0.778	1.49	0.790	1.51	32	12000	6.00	128	3.00E-05		1	0.1	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	6.1	0.738	1.67	0.748	1.68	32	12000	6.00	128	3.00E-05		1	0.06	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	3.1	0.689	1.90	0.698	1.89	32	12000	6.00	128	3.00E-05		1	0.03	1	1	magnitude								
B1		magnitude_1.0_*_1_1_null_0._3e-5_0._mag	2.1	0.663	2.02	0.681	1.98	32	12000	6.00	128	3.00E-05		1	0.02	1	1	magnitude								
B1	Magnitude Local	magnitude_1.0_*_1_1_null_0._3e-5_0._mag	1.1	0.629	2.22	0.644	2.17	32	12000	6.00	128	3.00E-05		1	0.01	1	1	magnitude								
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	90.0	0.853	1.18	0.857	1.18	32	12000	6.00	128	3.00E-05		1	0.9	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	60.1	0.852	1.17	0.852	1.18	32	12000	6.00	128	3.00E-05		1	0.6	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	40.1	0.844	1.17	0.849	1.17	32	12000	6.00	128	3.00E-05		1	0.4	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	30.1	0.838	1.18	0.843	1.18	32	12000	6.00	128	3.00E-05		1	0.3	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	20.1	0.819	1.20	0.829	1.21	32	12000	6.00	128	3.00E-05		1	0.2	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	10.1	0.783	1.29	0.793	1.30	32	12000	6.00	128	3.00E-05		1	0.1	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	6.1	0.748	1.38	0.756	1.38	32	12000	6.00	128	3.00E-05		1	0.06	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	3.1	0.694	1.53	0.706	1.52	32	12000	6.00	128	3.00E-05		1	0.03	1	1	magnitude					bert	bert-base	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_1_null_0._3e	2.1	0.671	1.60	0.681	1.57	32	12000	6.00	128	3.00E-05		1	0.02	1	1	magnitude					bert	bert-base	0.1	0.9
B3	Magnitude Local W/ Distillation	magnitude_with_distil_1.0_*_1_1_null_0._3e	1.1	0.637	1.76	0.650	1.71	32	12000	6.00	128	3.00E-05		1	0.01	1	1	magnitude					bert	bert-base	0.1	0.9
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	66.2	0.815	1.20	0.821	1.20	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	0				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	61.4	0.816	1.19	0.819	1.19	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	0.1				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	49.7	0.814	1.20	0.818	1.20	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	26.1	0.805	1.29	0.811	1.30	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	10				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	8.4	0.776	1.51	0.783	1.52	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	50				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	5.5	0.767	1.57	0.771	1.58	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	75				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	3.8	0.760	1.62	0.764	1.62	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	100				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	2.1	0.741	1.69	0.743	1.69	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	150				
C2	L0	l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_constan	1.0	0.722	1.79	0.731	1.78	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	250				
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	65.1	0.830	1.19	0.836	1.19	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	0	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	55.7	0.826	1.19	0.835	1.19	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	36.7	0.819	1.19	0.829	1.19	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	10	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	23.6	0.808	1.21	0.816	1.22	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	30	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	17.5	0.806	1.23	0.811	1.24	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	50	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	13.1	0.797	1.26	0.799	1.27	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	75	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	10.4	0.788	1.28	0.798	1.28	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	100	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	8.4	0.782	1.29	0.792	1.30	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	125	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	7.0	0.775	1.31	0.782	1.32	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	150	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	5.1	0.769	1.33	0.777	1.34	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	200	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	3.9	0.766	1.36	0.769	1.36	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	250	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	3.1	0.761	1.37	0.764	1.37	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	300	bert	bert-base	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	2.5	0.754	1.39	0.745	1.41	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	350	bert	bert-base	0.1	0.9
C4	L0	l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	1.7	0.760	1.39	0.750	1.41	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	450	bert	bert-base	0.1	0.9
C4	W/ Distillation	l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1_	1.1	0.736	1.44	0.745	1.44	32	12000	6.00	128	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	500	bert	bert-base	0.1	0.9
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	90.0	0.843	1.07	0.842	1.07	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.9	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	70.0	0.825	1.13	0.824	1.13	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.70	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	50.1	0.814	1.21	0.814	1.22	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.50	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	30.1	0.807	1.29	0.812	1.30	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.30	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	10.1	0.793	1.45	0.795	1.46	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.10	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	8.1	0.790	1.48	0.792	1.49	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.08	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	5.1	0.780	1.55	0.780	1.56	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.05	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	3.1	0.761	1.63	0.767	1.63	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.03	1	1	topK	constant	0						
D1		topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	2.1	0.746	1.71	0.753	1.71	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.02	1	1	topK	constant	0						
D1	Movement Local	topK_1.0_*_1_1_null_0._3e-5_1e-2_topK_c	1.1	0.713	1.82	0.728	1.81	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.01	1	1	topK	constant	0						
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	90.0	0.850	1.18	0.854	1.18	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.9	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	70.0	0.838	1.18	0.839	1.19	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.60	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	50.1	0.825	1.19	0.829	1.19	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.40	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	30.1	0.816	1.21	0.825	1.21	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.30	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	10.1	0.801	1.27	0.804	1.29	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.20	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	8.1	0.796	1.29	0.800	1.29	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.10	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	5.1	0.785	1.32	0.788	1.33	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.06	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3		topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	3.1	0.765	1.37	0.774	1.38	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.03	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3	Movement Local	topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	2.1	0.752	1.41	0.760	1.41	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.02	1	1	topK	constant	0			bert	bert-base	0.1	0.9
D3	W/ distillation	topK_with_distil_1.0_*_1_1_null_0._3e-5_1e	1.1	0.716	1.49	0.727	1.49	32	12000	6.00	128	3.00E-05	1.00E-02	1	0.01	1	1	topK	constant	0			bert	bert-base	0.1	0.9
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	90.1	0.840	1.07	0.840	1.07	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	1				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	72.5	0.834	1.07	0.843	1.07	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	10.00				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	28.6	0.823	1.16	0.825	1.17	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	100.00				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	19.1	0.816	1.23	0.821	1.23	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	150.00				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	12.8	0.812	1.29	0.817	1.30	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	200.00				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	5.9	0.800	1.41	0.803	1.42	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	300.00				
E2		l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	3.1	0.790	1.49	0.797	1.50	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	400.00				
E2	Soft Movement	l1_0._0.1_1_1_l1_*_3e-5_1e-2_sigmoied_th	1.9	0.782	1.54	0.787	1.56	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	500.00				
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	90.7	0.849	1.18	0.852	1.18	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	1	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	75.1	0.849	1.18	0.853	1.18	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	10.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	34.7	0.837	1.18	0.841	1.19	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	100.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	24.7	0.833	1.19	0.834	1.20	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	175.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	18.6	0.824	1.20	0.827	1.21	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	250.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	13.1	0.817	1.22	0.821	1.23	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	350.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	9.5	0.811	1.23	0.817	1.24	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	450.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	6.0	0.807	1.25	0.812	1.27	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	600.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	4.0	0.800	1.27	0.806	1.28	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	750.00	bert	bert-base	0.1	0.9
E4		l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	2.8	0.794	1.29	0.80	1.30	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	900.00	bert	bert-base	0.1	0.9
E4	Soft Movement	l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	2.1	0.791	1.31	0.797	1.32	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	1,050.00	bert	bert-base	0.1	0.9
E4	W/ Distillation	l1_with_distil_0._0.1_1_1_l1_*_3e-5_1e-2_s	1.7	0.788	1.32	0.79	1.33	32	12000	6.00	128	3.00E-05	1.00E-02	0	0.1	1	1	sigmoied	constant	0	l1	1,200.00	bert	bert-base	0.1	0.9

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/mnli	seed=92		dev		dev-MM							Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	ent	acc	ent		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-mnli	bert-base-uncased		0.845	1.22	0.849	1.22		32	12000	3.00	128	3.00E-05														

	bert-base-uncased/92/squad	seed=92		squad-1.1 - dev						Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	EM	F1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07		16	5400	3	3.00E-05														
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	90.0	78.98	87.55		16	5400	10	3.00E-05		1	0.9	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	60.1	78.84	87.32		16	5400	10	3.00E-05		1	0.6	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	40.1	78.40	87.00		16	5400	10	3.00E-05		1	0.4	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	30.1	77.33	86.25		16	5400	10	3.00E-05		1	0.3	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	20.1	75.71	84.80		16	5400	10	3.00E-05		1	0.2	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	10.1	67.71	78.51		16	5400	10	3.00E-05		1	0.1	1	2	magnitude								
B1		magnitude_1.0_*_1_2_null_0._3e-5_0._mag	6.1	58.14	70.74		16	5400	10	3.00E-05		1	0.06	1	2	magnitude								
B1	Magnitude Local	magnitude_1.0_*_1_2_null_0._3e-5_0._mag	3.1	40.08	54.49		16	5400	10	3.00E-05		1	0.03	1	2	magnitude								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	90.0	78.96	87.58		16	5400	10	3.00E-05		1	0.9	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	60.1	78.45	87.21		16	5400	10	3.00E-05		1	0.6	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	40.1	77.94	87.00		16	5400	10	3.00E-05		1	0.4	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	30.1	77.36	86.17		16	5400	10	3.00E-05		1	0.3	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	20.1	75.49	84.42		16	5400	10	3.00E-05		1	0.2	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	10.1	66.86	77.52		16	5400	10	3.00E-05		1	0.1	1	2	threshold								
B2		magnitude_global_1.0_*_1_2_null_0._3e-5_	6.1	57.93	70.63		16	5400	10	3.00E-05		1	0.06	1	2	threshold								
B2	Magnitude Global	magnitude_global_1.0_*_1_2_null_0._3e-5_	3.1	43.20	57.47		16	5400	10	3.00E-05		1	0.03	1	2	threshold								
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	90.0	82.71	89.65		16	5400	10	3.00E-05		1	0.9	1	2	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	60.1	82.48	89.37		16	5400	10	3.00E-05		1	0.6	1	2	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	40.1	81.83	88.93		16	5400	10	3.00E-05		1	0.4	1	2	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	30.1	80.65	88.07		16	5400	10	3.00E-05		1	0.3	1	2	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	20.1	78.57	86.50		16	5400	10	3.00E-05		1	0.2	1	2	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_1_2_null_0._3(	10.1	70.15	80.10		16	5400	10	3.00E-05		1	0.1	1	2	magnitude					bert	bert-base-	0.1	0.9
B3	Magnitude Local	magnitude_with_distil_1.0_*_1_2_null_0._3(	6.1	61.25	73.02		16	5400	10	3.00E-05		1	0.06	1	2	magnitude					bert	bert-base-	0.1	0.9
B3	W/ Distillation	magnitude_with_distil_1.0_*_1_2_null_0._3(	3.1	45.53	59.58		16	5400	10	3.00E-05		1	0.03	1	2	magnitude					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	90.0	82.72	89.68		16	5400	10	3.00E-05		1	0.9	1	2	threshold					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	60.1	82.58	89.55		16	5400	10	3.00E-05		1	0.6	1	2	threshold					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	40.1	81.58	88.66		16	5400	10	3.00E-05		1	0.4	1	2	threshold					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	30.1	80.53	87.98		16	5400	10	3.00E-05		1	0.3	1	2	threshold					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	20.1	77.80	85.93		16	5400	10	3.00E-05		1	0.2	1	2	threshold					bert	bert-base-	0.1	0.9
B4		magnitude_with_distil_1.0_*_1_2_null_0._3(	10.1	69.01	79.40		16	5400	10	3.00E-05		1	0.1	1	2	threshold					bert	bert-base-	0.1	0.9
B4	Magnitude Global	magnitude_with_distil_1.0_*_1_2_null_0._3(	6.1	60.63	72.59		16	5400	10	3.00E-05		1	0.06	1	2	threshold					bert	bert-base-	0.1	0.9
B4	W/ Distillation	magnitude_with_distil_1.0_*_1_2_null_0._3(	3.1	47.66	61.61		16	5400	10	3.00E-05		1	0.03	1	2	threshold					bert	bert-base-	0.1	0.9
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	66.8	74.97	84.56		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	0				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	58.4	75.16	84.65		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	40.7	74.30	83.91		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	10				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	17.7	71.88	81.76		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	75				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	14.8	70.74	80.89		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	100				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	9.8	69.86	80.01		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	175				
C2		l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	6.1	67.22	78.06		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	300				
C2	L0 Regu	l0_1.0_1.0_1_1_l0_*_3e-5_1e-1_l0_consta	2.9	61.03	73.09		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	600				
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	66.5	80.80	88.23		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	0	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	59.4	80.55	88.10		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	44.4	80.13	87.72		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	10	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	24.2	77.49	85.59		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	75	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	14.6	74.61	83.65		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	200	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	9.0	71.88	81.50		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	400	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	6.3	69.26	79.41		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	600	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	4.8	67.11	78.11		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	800	bert	bert-base-	0.1	0.9
C4	L0 Regu	l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	3.8	65.98	77.08		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1000	bert	bert-base-	0.1	0.9
C4	Distillation	l0_with_distil_1.0_1.0_1_1_l0_*_3e-5_1e-1	3.1	64.54	75.93		16	5400	10	3.00E-05	1.00E-01	1	1	1	1	l0	constant	2.197	l0	1200	bert	bert-base-	0.1	0.9
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	90.0	78.65	87.24		16	5400	10	3.00E-05	1.00E-02	1	0.9	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	70.0	76.11	85.02		16	5400	10	3.00E-05	1.00E-02	1	0.7	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	50.1	73.24	83.16		16	5400	10	3.00E-05	1.00E-02	1	0.5	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	30.1	72.46	82.38		16	5400	10	3.00E-05	1.00E-02	1	0.3	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	10.1	71.87	81.72		16	5400	10	3.00E-05	1.00E-02	1	0.1	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	8.1	71.44	81.21		16	5400	10	3.00E-05	1.00E-02	1	0.08	1	2	topK	constant	0						
D1		topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	5.1	69.36	79.59		16	5400	10	3.00E-05	1.00E-02	1	0.05	1	2	topK	constant	0						
D1	Movement Local	topK_1.0_*_1_2_null_0._3e-5_1e-2_topK_c	3.1	65.23	76.28		16	5400	10	3.00E-05	1.00E-02	1	0.03	1	2	topK	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	90.0	78.33	87.03		16	5400	10	3.00E-05	1.00E-02	1	0.9	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	70.0	75.72	84.93		16	5400	10	3.00E-05	1.00E-02	1	0.7	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	50.1	73.54	83.15		16	5400	10	3.00E-05	1.00E-02	1	0.5	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	30.1	72.28	82.28		16	5400	10	3.00E-05	1.00E-02	1	0.3	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	10.1	71.68	81.66		16	5400	10	3.00E-05	1.00E-02	1	0.1	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	8.1	71.53	81.49		16	5400	10	3.00E-05	1.00E-02	1	0.08	1	2	threshold	constant	0						
D2		topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	5.1	70.89	80.99		16	5400	10	3.00E-05	1.00E-02	1	0.05	1	2	threshold	constant	0						
D2	Movement Global	topK_global_1.0_*_1_2_null_0._3e-5_1e-2_	3.1	68.12	78.61		16	5400	10	3.00E-05	1.00E-02	1	0.03	1	2	threshold	constant	0						
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	90.0	82.81	89.73		16	5400	10	3.00E-05	1.00E-02	1	0.9	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	70.0	81.06	88.41		16	5400	10	3.00E-05	1.00E-02	1	0.7	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	50.1	79.84	87.58		16	5400	10	3.00E-05	1.00E-02	1	0.5	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	30.1	78.68	86.58		16	5400	10	3.00E-05	1.00E-02	1	0.3	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	10.1	75.57	84.29		16	5400	10	3.00E-05	1.00E-02	1	0.1	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	8.1	74.94	83.77		16	5400	10	3.00E-05	1.00E-02	1	0.08	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3	Movement Local	topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	5.1	72.04	81.43		16	5400	10	3.00E-05	1.00E-02	1	0.05	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D3	W/ Distillation	topK_with_distil_1.0_*_1_2_null_0._3e-5_1(	3.1	67.54	78.04		16	5400	10	3.00E-05	1.00E-02	1	0.03	1	2	topK	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	90.0	82.61	89.54		16	5400	10	3.00E-05	1.00E-02	1	0.9	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	70.0	81.55	88.63		16	5400	10	3.00E-05	1.00E-02	1	0.7	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	50.1	79.98	87.46		16	5400	10	3.00E-05	1.00E-02	1	0.5	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	30.1	78.86	86.80		16	5400	10	3.00E-05	1.00E-02	1	0.3	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	10.1	75.97	84.67		16	5400	10	3.00E-05	1.00E-02	1	0.1	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4		topK_global_with_distil_1.0_*_1_2_null_0._	8.1	75.44	84.20		16	5400	10	3.00E-05	1.00E-02	1	0.08	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4	Movement Global	topK_global_with_distil_1.0_*_1_2_null_0._	5.1	73.20	82.45		16	5400	10	3.00E-05	1.00E-02	1	0.05	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
D4	W/ Distillation	topK_global_with_distil_1.0_*_1_2_null_0._	3.1	70.03	80.02		16	5400	10	3.00E-05	1.00E-02	1	0.03	1	2	threshold	constant	0			bert	bert-base-	0.1	0.9
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	93.6	78.17	86.93		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	1				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	83.1	78.31	87.03		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	10				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	59.6	78.01	86.55		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	50				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	45.5	76.91	85.77		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	100				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	31.2	75.29	84.60		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	200				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	13.9	72.19	82.17		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	500				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	7.1	70.55	80.97		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	800				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	4.0	70.04	80.39		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	1100				
E2		l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	2.6	69.23	79.62		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	1400				
E2	Soft Movement	l1_0._0.1_1_2_l1_*_3e-5_1e-2_sigmoied_th	1.2	66.67	77.46		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	2300				
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	93.4	82.63	89.51		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	10	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	33.7	81.03	88.37		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	100	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	23.2	79.85	87.53		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	200	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	12.7	77.30	85.51		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied	constant	0	l1	500	bert	bert-base-	0.1	0.9

	bert-base-uncased/92/squad	seed=92		squad-1.1 - dev						Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	EM	F1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07		16	5400	3	3.00E-05														
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	8.4	76.22	84.55		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	800	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	6.2	74.99	83.80		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	1100	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	4.7	74.65	83.53		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	1400	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	3.7	73.67	82.90		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	1700	bert	bert-base-	0.1	0.9
E4	Soft Movement	l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	3.1	72.84	82.37		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	2000	bert	bert-base-	0.1	0.9
E4	W/ Distilation	l1_with_distil_0._0.1_1_2_l1_*_3e-5_1e-2_s	2.6	72.09	81.74		16	5400	10	3.00E-05	1.00E-02	0.00E+00	1.00E-01	1	2	sigmoied_	constant	0	l1	2300	bert	bert-base-	0.1	0.9

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

	bert-base-uncased/92/squad	seed=92		squad-1.1 - dev						Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	EM	F1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07		16	5400	3	3.00E-05														

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

bert-base-uncased/92/squad	seed=92		squad-1.1 - dev					Learning Rate		Scheduling				Pruning			Regularization		Distillation			
Model	EXP ID	Effective encoder remain weights %	EM	F1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07	16	5400	3	3.00E-05														

	bert-base-uncased/92/squad	seed=92		squad-1.1 - dev						Learning Rate		Scheduling				Pruning				Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	EM	F1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-squad	bert-base-uncased		80.42	88.07		16	5400	3	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev									Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88	32	11300	3.00	128	3.00E-05														
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	90.0	0.903	0.887	1.119	0.872	32	11000	10.00	128	3.00E-05		1	0.9	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	60.1	0.902	0.887	1.137	0.871	32	11000	10.00	128	3.00E-05		1	0.6	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	40.1	0.894	0.877	1.180	0.860	32	11000	10.00	128	3.00E-05		1	0.4	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	30.1	0.885	0.866	1.223	0.847	32	11000	10.00	128	3.00E-05		1	0.3	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	20.1	0.860	0.838	1.302	0.816	32	11000	10.00	128	3.00E-05		1	0.2	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	10.1	0.788	0.770	1.478	0.751	32	11000	10.00	128	3.00E-05		1	0.1	2	3	magnitude								
B1		magnitude_1.0_*_2_3_null_0_3e-6_0._mag	6.1	0.744	0.693	1.636	0.642	32	11000	10.00	128	3.00E-05		1	0.06	2	3	magnitude								
B1	Magnitude Local	magnitude_1.0_*_2_3_null_0_3e-6_0._mag	3.1	0.721	0.652	1.698	0.584	32	11000	10.00	128	3.00E-05		1	0.03	2	3	magnitude								
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	90.0	0.904	0.888	1.101	0.873	32	11000	10.00	128	3.00E-05		1	0.9	2	3	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	60.1	0.902	0.886	1.107	0.869	32	11000	10.00	128	3.00E-05		1	0.6	2	3	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	40.1	0.895	0.876	1.126	0.858	32	11000	10.00	128	3.00E-05		1	0.4	2	3	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	30.1	0.887	0.867	1.147	0.847	32	11000	10.00	128	3.00E-05		1	0.3	2	3	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	20.1	0.864	0.841	1.189	0.818	32	11000	10.00	128	3.00E-05		1	0.2	2	3	magnitude					bert	bert-base-	0.1	0.9
B3		magnitude_with_distil_1.0_*_2_3_null_0._3e	10.1	0.798	0.778	1.301	0.759	32	11000	10.00	128	3.00E-05		1	0.1	2	3	magnitude					bert	bert-base-	0.1	0.9
B3	Magnitude Local	magnitude_with_distil_1.0_*_2_3_null_0._3e	6.1	0.749	0.699	1.411	0.650	32	11000	10.00	128	3.00E-05		1	0.06	2	3	magnitude					bert	bert-base-	0.1	0.9
B3	W/ Distillation	magnitude_with_distil_1.0_*_2_3_null_0._3e	3.1	0.724	0.651	1.469	0.578	32	11000	10.00	128	3.00E-05		1	0.03	2	3	magnitude					bert	bert-base-	0.1	0.9
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	90.3	0.898	0.876	1.052	0.853	32	11000	6.00	128	3.00E-06	1.00E-02	1	1	1	1	l0	constant	2.197	l0	1				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	65.0	0.896	0.872	1.047	0.849	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	5				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	49.0	0.892	0.868	1.055	0.843	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	10				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	13.8	0.879	0.851	1.103	0.823	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	50				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	8.1	0.874	0.845	1.132	0.817	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	75				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	5.2	0.871	0.844	1.156	0.816	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	100				
C2		l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	2.6	0.864	0.836	1.193	0.809	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	150				
C2	L0	l0_1.0_1.0_1_1_l0_*_3e-6_1e-2_l0_constan	1.6	0.851	0.820	1.233	0.789	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	200				
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	92.3	0.908	0.889	1.084	0.870	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	1	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	70.2	0.907	0.888	1.081	0.868	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	5	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	55.8	0.906	0.886	1.082	0.867	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	10	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	24.4	0.892	0.868	1.080	0.844	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	50	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	18.0	0.888	0.862	1.084	0.837	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	75	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	13.9	0.884	0.857	1.082	0.830	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	100	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	9.1	0.881	0.854	1.090	0.827	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	150	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	6.3	0.875	0.847	1.098	0.819	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	200	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	4.6	0.876	0.850	1.106	0.824	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	250	bert	bert-base-	0.1	0.9
C4		l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	3.5	0.873	0.848	1.112	0.824	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	300	bert	bert-base-	0.1	0.9
C4	L0	l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	2.7	0.869	0.843	1.121	0.817	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	350	bert	bert-base-	0.1	0.9
C4	W/ Distillation	l0_with_distil_1.0_1.0_1_1_l0_*_3e-6_1e-2_	2.2	0.864	0.837	1.129	0.811	32	11000	6.00	128	3.00E-05	1.00E-02	1	1	1	1	l0	constant	2.197	l0	400	bert	bert-base-	0.1	0.9
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	90.0	0.912	0.896	1.016	0.881	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.9	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	70.0	0.908	0.892	1.025	0.876	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.70	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	50.1	0.906	0.889	1.053	0.873	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.50	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	30.1	0.901	0.884	1.108	0.867	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.30	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	10.1	0.891	0.873	1.182	0.855	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.10	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	8.1	0.889	0.870	1.210	0.852	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.08	2	3	topK	constant	0						
D1		topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	5.1	0.877	0.858	1.257	0.838	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.05	2	3	topK	constant	0						
D1	Movement Local	topK_1.0_*_2_3_null_0._3e-6_1e-2_topK_c	3.1	0.856	0.833	1.334	0.810	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.03	2	3	topK	constant	0						
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	90.0	0.915	0.900	1.085	0.886	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.9	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	70.0	0.912	0.897	1.086	0.881	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.60	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	50.1	0.910	0.894	1.086	0.878	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.40	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	30.1	0.907	0.890	1.088	0.874	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.30	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	10.1	0.897	0.879	1.121	0.862	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.20	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	8.1	0.892	0.874	1.128	0.856	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.10	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	5.1	0.881	0.861	1.158	0.841	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.06	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3		topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	3.1	0.861	0.838	1.207	0.815	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.03	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3	Movement Local	topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	2.1	0.846	0.821	1.252	0.796	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.02	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
D3	W/ distillation	topK_with_distil_1.0_*_2_3_null_0._3e-6_1e	1.1	0.814	0.788	1.338	0.761	32	11000	10.00	128	3.00E-06	1.00E-02	1	0.01	2	3	topK	constant	0			bert	bert-base-	0.1	0.9
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	77.7	0.910	0.894	1.018	0.878	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	1				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	52.2	0.911	0.895	1.021	0.879	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	10.00				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	13.1	0.904	0.887	1.079	0.871	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	100.00				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	7.6	0.901	0.884	1.118	0.866	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	150.00				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	4.9	0.896	0.879	1.142	0.861	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	200.00				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	2.7	0.890	0.872	1.193	0.853	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	300.00				
E2		l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	1.8	0.885	0.866	1.216	0.847	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	400.00				
E2	Soft Movement Pruning	l1_0._0.1_2_3_l1_*_3e-6_1e-2_sigmoied_th	1.4	0.877	0.857	1.244	0.838	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	500.00				
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	78.8	0.915	0.900	1.086	0.884	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	1	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	54.1	0.916	0.901	1.086	0.886	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	10.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	18.9	0.908	0.892	1.089	0.876	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	100.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	14.0	0.908	0.891	1.094	0.875	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	150.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	10.9	0.905	0.888	1.097	0.872	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	200.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	7.1	0.902	0.885	1.104	0.868	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	300.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	5.1	0.899	0.881	1.111	0.864	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	400.00	bert	bert-base-	0.1	0.9
E4		l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	3.9	0.896	0.878	1.120	0.859	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	500.00	bert	bert-base-	0.1	0.9
E4	Soft Movement Pruning	l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	3.2	0.893	0.875	1.123	0.856	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	600.00	bert	bert-base-	0.1	0.9
E4	W/ distillation	l1_with_distil_0._0.1_2_3_l1_*_3e-6_1e-2_s	2.6	0.891	0.873	1.129	0.854	32	11000	10.00	128	3.00E-06	1.00E-02	0	0.1	2	3	sigmoied_	constant	0	l1	700.00	bert	bert-base-	0.1	0.9

	bert-base-uncased/92/qqp	seed=92		dev										Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev										Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev									Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev										Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

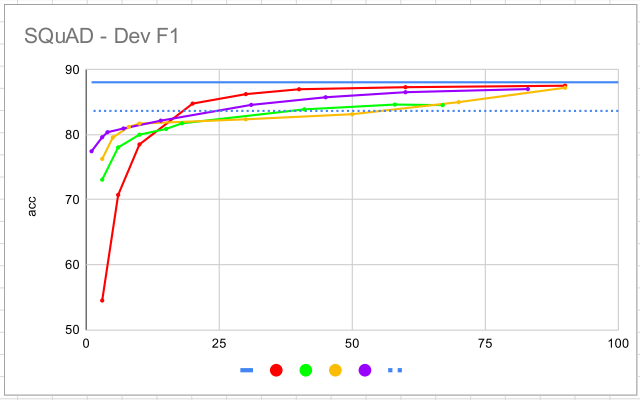
	bert-base-uncased/92/qqp	seed=92		dev									Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

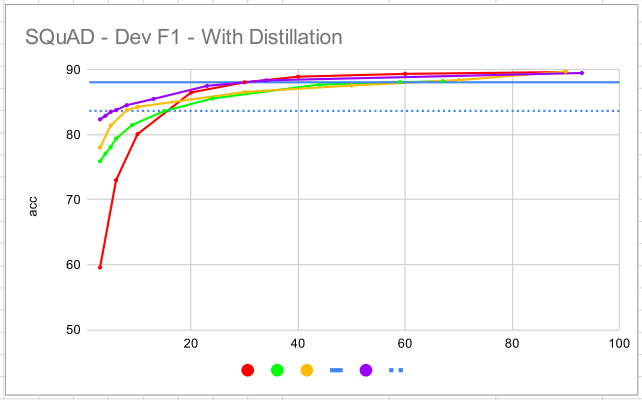
	bert-base-uncased/92/qqp	seed=92		dev									Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev										Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1		per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88		32	11300	3.00	128	3.00E-05														

	bert-base-uncased/92/qqp	seed=92		dev								Learning Rate		Scheduling				Pruning			Regularization		Distillation			
	Model	EXP ID	Effective encoder remain weights %	acc	acc& f1	ent	f1	per_gpu_train_batch_size	warmup_steps	num_train_epochs	max_seq_length	learning_rate	mask_scores_learning_rate	initial_threshold	final_threshold	initial_warmup	final_warmup	pruning_method	mask_init	mask_scale	regularization	final_lambda	teacher_type	teacher_name_or_path	alpha_ce	alpha_distil
	bert-base-uncased-finetuned-qqp	bert-base-uncased		0.91	0.90	1.10	0.88	32	11300	3.00	128	3.00E-05														

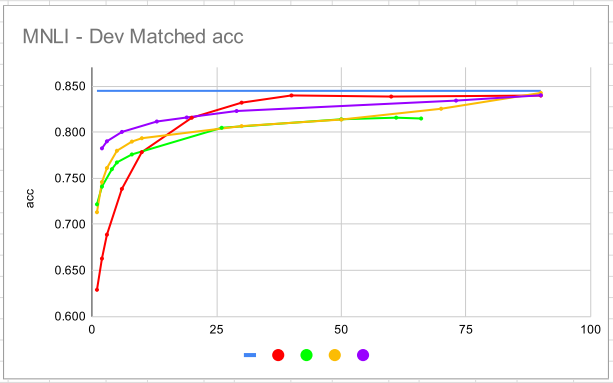
Category	HP Name	HP Explanation
	per_gpu_train_batch_size	Batch size per GPU/CPU for training.
	warmup_steps	Linear warmup over warmup_steps.
	num_train_epochs	Total number of training epochs to perform.
	max_seq_length	The maximum total input sequence length after WordPiece tokenization. Sequences longer than this will be trunc
	learning_rate	The initial learning rate for Adam.
Learning Rate	mask_scores_learning_rate	The Adam initial learning rate of the mask scores.
	initial_threshold	Initial value of the threshold (for scheduling).
	final_threshold	Final value of the threshold (for scheduling).
	initial_warmup	Run `initial_warmup` * `warmup_steps` steps of threshold warmup during which threshold stays at its `initial_thre
Scheduling	final_warmup	Run `final_warmup` * `warmup_steps` steps of threshold cool-down during which threshold stays at its final_thres
	pruning_method	Pruning Method (l0 = L0 regularization, magnitude = Magnitude pruning, topK = Movement pruning, sigmoied_thr
	mask_init	Initialization method for the mask scores. Choices: constant, uniform, kaiming.
Pruning	mask_scale	Initialization parameter for the chosen initialization method.
	regularization	Add L0 or L1 regularization to the mask scores.
Regularization	final_lambda	Regularization intensity (used in conjunction with `regulariation`.
	teacher_type	Teacher type. Teacher tokenizer and student (model) tokenizer must output the same tokenization. Only for distill
	teacher_name_or_path	Path to the already SQuAD fine-tuned teacher model. Only for distillation.
	alpha_ce	Cross entropy loss linear weight. Only for distillation.
Distillation	alpha_distil	Distillation loss linear weight. Only for distillation.

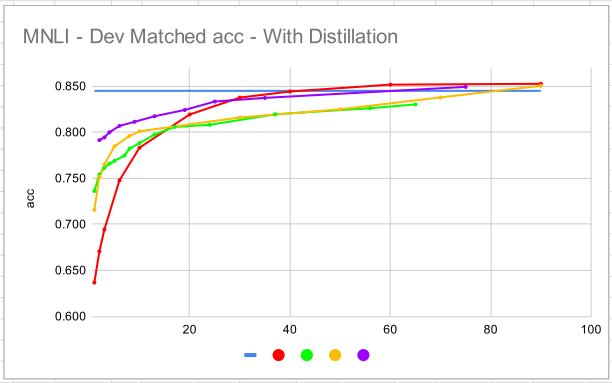




	A		B1		B2		B3		B4		C2		C4		D1		D2		D3		D4		E2		E4		
	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	
100	80.4	88.1																									83.67108917
99																											
98																											
97																											
96																											
95																											
94																											
93																									82.6	89.5	
92																											
91																											
90	80.4	88.1	79.0	87.5	79.0	87.6	82.7	89.7	82.7	89.7					78.6	87.2	78.3	87.0	82.8	89.7	82.6	89.5					
89																											
88																											
87																											
86																											
85																											
84																											
83																							78.3	87.0			
82																											
81																											
80	80.4	88.1																									
79																											
78																											
77																											
76																											
75																											
74																											
73																											
72																											
71																											
70	80.4	88.1													76.1	85.0	75.7	84.9	81.1	88.4	81.6	88.6					
69																											
68																											
67											75.0	84.6	80.8	88.2													
66																											
65																											
64																											
63																											
62																											
61																											
60	80.4	88.1	78.8	87.3	78.4	87.2	82.5	89.4	82.6	89.6													78.0	86.5			
59													80.5	88.1													
58											75.2	84.6															
57																											
56																											
55																											
54																											
53																											
52																											
51																											
50	80.4	88.1													73.2	83.2	73.5	83.1	79.8	87.6	80.0	87.5					
49																											
48																											
47																											
46																											
45																							76.9	85.8			
44													80.1	87.7													
43																											
42																											
41											74.3	83.9															

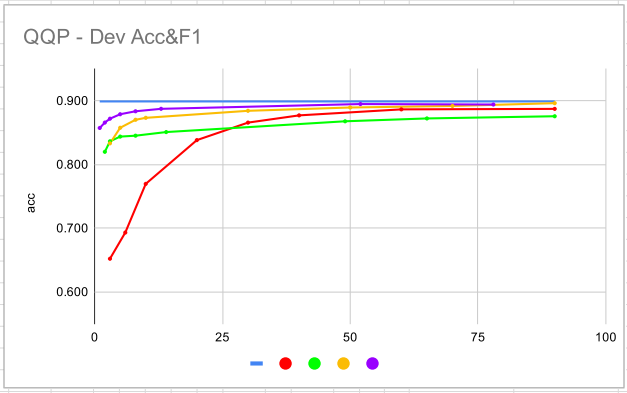
40	80.4	88.1	78.4	87.0	77.9	87.0	81.8	88.9	81.6	88.7																			
39																													
38																													
37																													
36																													
35																													
34																											81.0	88.4	
33																													
32																													
31																									75.3	84.6			
30	80.4	88.1	77.3	86.3	77.4	86.2	80.7	88.1	80.5	88.0					72.5	82.4	72.3	82.3	78.7	86.6	78.9	86.8							
29																													
28																													
27																													
26																													
25																													
24													77.5	85.6															
23																											79.8	87.5	
22																													
21																													
20	80.4	88.1	75.7	84.8	75.5	84.4	78.6	86.5	77.8	85.9																			
19																													
18											71.9	81.8																	
17																													
16																													
15											70.7	80.9	74.6	83.6															
14																									72.2	82.2			
13																											77.3	85.5	
12																													
11																													
10	80.4	88.1	67.7	78.5	66.9	77.5	70.2	80.1	69.0	79.4	69.9	80.0			71.9	81.7	71.7	81.7	75.6	84.3	76.0	84.7							
9													71.9	81.5															
8															71.4	81.2	71.5	81.5	74.9	83.8	75.4	84.2					76.2	84.6	
7																									70.5	81.0			
6			58.1	70.7	57.9	70.6	61.2	73.0	60.6	72.6	67.2	78.1	69.3	79.4													75.0	83.8	
5													67.1	78.1	69.4	79.6	70.9	81.0	72.0	81.4	73.2	82.4					74.6	83.5	
4													66.0	77.1											70.0	80.4	73.7	82.9	
3			40.1	54.5	43.2	57.5	45.5	59.6	47.7	61.6	61.0	73.1	64.5	75.9	65.2	76.3	68.1	78.6	67.5	78.0	70.0	80.0	69.2	79.6	72.8	82.4			
2																													
1	80.4	88.1																					66.7	77.5			83.67108917		

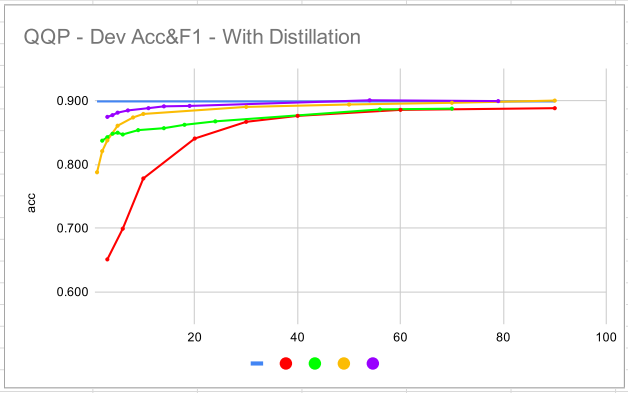




	A		B1		B2		B3		B4		C2		C4		D1		D2		D3		D4		E2		E4	
	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM	acc	acc-MM
90	0.845	0.849	0.840	0.847			0.853	0.857							0.843	0.842			0.850	0.854	0.847	0.849	0.840	0.840		
89	0.845	0.849																								
88	0.845	0.849																								
87	0.845	0.849																								
86	0.845	0.849																								
85	0.845	0.849																								
84	0.845	0.849																								
83	0.845	0.849																								
82	0.845	0.849																								
81	0.845	0.849																								
80	0.845	0.849																								
79	0.845	0.849																								
78	0.845	0.849																								
77	0.845	0.849																								
76	0.845	0.849																								
75	0.845	0.849																							0.849	0.853
74	0.845	0.849																								
73	0.845	0.849																					0.834	0.843		
72	0.845	0.849																								
71	0.845	0.849																								
70	0.845	0.849													0.825	0.824			0.838	0.839	0.839	0.840				
69	0.845	0.849																								
68	0.845	0.849																								
67	0.845	0.849																								
66	0.845	0.849									0.815	0.821														
65	0.845	0.849											0.830	0.836												
64	0.845	0.849																								
63	0.845	0.849																								
62	0.845	0.849																								
61	0.845	0.849									0.816	0.819														
60	0.845	0.849	0.839	0.838			0.852	0.852																		
59	0.845	0.849																								
58	0.845	0.849																								
57	0.845	0.849																								
56	0.845	0.849											0.826	0.835												
55	0.845	0.849																								
54	0.845	0.849																								
53	0.845	0.849																								
52	0.845	0.849																								
51	0.845	0.849																								
50	0.845	0.849									0.814	0.818			0.814	0.814			0.825	0.829	0.828	0.827				
49	0.845	0.849																								
48	0.845	0.849																								
47	0.845	0.849																								
46	0.845	0.849																								
45	0.845	0.849																								
44	0.845	0.849																								
43	0.845	0.849																								
42	0.845	0.849																								
41	0.845	0.849																								
40	0.845	0.849	0.840	0.841			0.844	0.849																		
39	0.845	0.849																								
38	0.845	0.849																								
37	0.845	0.849											0.819	0.829												
36	0.845	0.849																								
35	0.845	0.849																							0.837	0.841
34	0.845	0.849																								
33	0.845	0.849																								
32	0.845	0.849																								
31	0.845	0.849																								
30	0.845	0.849	0.832	0.835			0.838	0.843							0.807	0.812			0.816	0.825	0.814	0.821				
29	0.845	0.849																					0.823	0.825		
28	0.845	0.849																								
27	0.845	0.849																								

26	0.845	0.849									0.805	0.811													
25	0.845	0.849																						0.833	0.834
24	0.845	0.849											0.808	0.816											
23	0.845	0.849																							
22	0.845	0.849																							
21	0.845	0.849																							
20	0.845	0.849	0.816	0.827			0.819	0.829																	
19	0.845	0.849																					0.816	0.821	0.824	0.827
18	0.845	0.849																							
17	0.845	0.849											0.806	0.811											
16	0.845	0.849																							
15	0.845	0.849																							
14	0.845	0.849																							
13	0.845	0.849											0.797	0.799									0.812	0.817	0.817	0.821
12	0.845	0.849																							
11	0.845	0.849																							
10	0.845	0.849	0.778	0.790			0.783	0.793					0.788	0.798	0.793	0.795			0.801	0.804	0.807	0.804				
9	0.845	0.849																							0.811	0.817
8	0.845	0.849									0.776	0.783	0.782	0.792	0.790	0.792			0.796	0.800	0.800	0.805				
7	0.845	0.849											0.775	0.782												
6	0.845	0.849	0.738	0.748			0.748	0.756															0.800	0.803	0.807	0.812
5	0.845	0.849									0.767	0.771	0.769	0.777	0.780	0.780			0.785	0.788	0.795	0.792				
4	0.845	0.849									0.760	0.764	0.766	0.769											0.800	0.806
3	0.845	0.849	0.689	0.698			0.694	0.706					0.761	0.764	0.761	0.767			0.765	0.774	0.777	0.781	0.790	0.797	0.794	0.801
2	0.845	0.849	0.663	0.681			0.671	0.681			0.741	0.743	0.754	0.745	0.746	0.753			0.752	0.760			0.782	0.787	0.791	0.797
1	0.845	0.849	0.629	0.644			0.637	0.650			0.722	0.731	0.736	0.745	0.713	0.728			0.716	0.727						





	A	B1	B2	B3	B4	C2	C4	D1	D2	D3	D4	E2	E4
90	0.899	0.887		0.8		0.87		0.896		0.900			
89	0.899												
88	0.899												
87	0.899												
86	0.899												
85	0.899												
84	0.899												
83	0.899												
82	0.899												
81	0.899												
80	0.899												
79	0.899												0.9
78	0.899											0.8	
77	0.899												
76	0.899												
75	0.899												
74	0.899												
73	0.899												
72	0.899												
71	0.899												
70	0.899						0.8	0.892		0.897			
69	0.899												
68	0.899												
67	0.899												
66	0.899												
65	0.899					0.87							
64	0.899												
63	0.899												
62	0.899												
61	0.899												
60	0.899	0.887		0.8									
59	0.899												
58	0.899												
57	0.899												
56	0.899						0.8						
55	0.899												
54	0.899												0.9
53	0.899												
52	0.899											0.8	
51	0.899												
50	0.899							0.889		0.894			
49	0.899					0.8							
48	0.899												
47	0.899												
46	0.899												
45	0.899												
44	0.899												
43	0.899												
42	0.899												
41	0.899												
40	0.899	0.877		0.87									
39	0.899												
38	0.899												
37	0.899												
36	0.899												
35	0.899												
34	0.899												
33	0.899												
32	0.899												
31	0.899												
30	0.899	0.866		0.8				0.884		0.890			
29	0.899												

28	0.899												
27	0.899												
26	0.899												
25	0.899												
24	0.899						0.86						
23	0.899												
22	0.899												
21	0.899												
20	0.899	0.838		0.84									
19	0.899												0.8
18	0.899						0.86						
17	0.899												
16	0.899												
15	0.899												
14	0.899					0.85	0.85						0.8
13	0.899										0.8		
12	0.899												
11	0.899												0.8
10	0.899	0.770		0.77				0.873		0.879			
9	0.899						0.85						
8	0.899					0.84		0.870		0.874		0.8	
7	0.899												0.8
6	0.899	0.693		0.69			0.84						
5	0.899					0.84	0.85	0.858		0.861		0.8	0.8
4	0.899						0.84						0.8
3	0.899	0.652		0.65		0.83	0.84	0.833		0.838		0.8	0.8
2	0.899					0.82	0.83			0.821		0.8	
1	0.899									0.788		0.8	