Results

Model	LeNet	*All attacks are black-box*
Optimizer	SGD	
Learning rate	0.001	
Batch size	32	

Method	Baseline					
Epochs	51		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	747	88.11	78.69	65.26	40.5	23.12
1	732	88.24	78.44	64.78	39.62	21.96
2	769	88.27	78.82	65.01	39.28	21.05
3	727	88.57	79.46	65.37	39.21	21.29
4	715	88.55	79.24	65.55	41	23.75
Mean	**738**	**88.35**	**78.93**	**65.19**	**39.92**	**22.23**
Std	20.78	0.2	0.41	0.3	0.79	1.17

Method	Isoreg					
Epochs	51					
Lambda	0.1					
Warm-up	1 epoch					
Epsilon	3.5	*~16/255*				
Norm	Holder		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	3580	88.14	78.96	66.11	42.74	25.26
1	3650	88.12	78.22	65.07	40.88	24.08
2	3807	88.32	79.1	65.44	40.55	22.33
3	4029	88.6	79.47	66.07	41.14	22.95
4	3834	88.23	79.39	66.92	43.46	26.51
Mean	**3780**	**88.28**	**79.03**	**65.92**	**41.75**	**24.23**
Std	175.03	0.19	0.5	0.71	1.27	1.7

Method Isoreg

Epochs	51					
Lambda	3					
Warm-up	1 epoch					
Epsilon	3.5	*~16/255*				
Norm	Holder		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	3981	71.24	63.59	53.85	38.97	29.44
1	4078	87.54	78.04	64.77	41.88	26.58
2	3828	86.49	78.3	67.1	47.3	30.97
3	3389	83.1	75.34	64.25	46.21	32.6
4	3304	82.87	74.65	64.53	47.88	34.55
Mean	**3716**	**82.25**	**73.98**	**62.9**	**44.45**	**30.83**
Std	350.17	6.49	6.03	5.18	3.86	3.04

Method	Isoreg					
Epochs	51					
Lambda	4					
Warm-up	1 epoch					
Epsilon	3.5	*~16/255*				
Norm	Holder		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
5	3885	79.79	71.97	60.99	44.13	31.5
6	3479	81.49	72.19	61.51	43.85	30.99
7	3721	83.03	74.03	62.79	44.71	30.14
8	3919	69.68	62.03	52.54	37.73	28.51
9	3865	80.01	71.62	60.99	43.87	30.11
Mean	**3773.8**	**78.8**	**70.37**	**59.76**	**42.86**	**30.25**
Std	181.29	5.26	4.75	4.1	2.89	1.14

Method	Isoreg
Epochs	51
Lambda	2
Warm-up	1 epoch

Epsilon		3.5	*~16/255*				
Norm		Holder		**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	0	3544	87.89	78.19	64.83	40.8	23.68
	1	3939	87.99	78.01	64.43	40.05	23.12
	2	4002	87.96	78.23	64.74	39.86	21.92
	3	3950	88.27	78.58	64.93	39.82	22.15
	4	3825	88.21	78.78	65.39	41.62	25.31
Mean		**3852**	**88.06**	**78.36**	**64.86**	**40.43**	**23.24**
Std		183.89	0.17	0.31	0.35	0.77	1.36

Method		Isoreg					
Batch size		64					
Epochs		51					
Lambda		3					
Warm-up		1 epoch					
Epsilon		3.5	*~16/255*				
Norm		Holder		**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	10	2062	85.8	75.86	62.88	40.9	25.13
	11	2105	85.94	76.28	64.25	41.62	25.78
	12	2085	85.55	75.65	63.16	41.11	25.78
	13	2045	85.97	75.73	63.21	40.62	25.03
	14	2045	86.05	76	63.47	40.77	25.22
Mean		**2068.4**	**85.86**	**75.9**	**63.39**	**41**	**25.39**
Std		26.23	0.2	0.25	0.52	0.39	0.36

Method		Isoreg	
Batch size		64	
Epochs		51	
Lambda		12	
Warm-up		1 epoch	
Epsilon		3.5	*~16/255*

Norm		Holder		**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	15	1934	65.8	52.45	42.95	31.64	23.01
	16	1853	69.83	56.21	45.43	33.17	24.67
	17	1779	71.46	58.55	47.83	34.93	26.48
	18	1745	67.07	54.13	44.2	32.71	24.17
	19	1735	67.04	54.84	45.9	34.38	25.55
Mean		**1809.2**	**68.24**	**55.24**	**45.26**	**33.37**	**24.78**
Std		83.71	2.33	2.3	1.84	1.32	1.32

Method		Isoreg					
Batch size		128					
Epochs		51					
Lambda		4					
Warm-up		1 epoch					
Epsilon		3.5	*~16/255*				
Norm		Holder		**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	30	1123	84.18	75.91	65.43	46.24	30.01
	31	1454	84.01	75.92	65.34	46.39	30.39
	32	1734	83.59	75.4	65.32	46.58	30.76
	33	1733	83.73	74.99	64.26	45.32	30.16
	34	1718	83.5	75.49	65.2	46.36	30.86
Mean		**1552.4**	**83.8**	**75.54**	**65.11**	**46.18**	**30.44**
Std		267.9	0.29	0.39	0.48	0.49	0.37

Method		Isorandom					
Batch size		64					
Epochs		51					
Lambda		5					
Warm-up		1 epoch					
Epsilon		3.5	*~16/255*				
Norm		Holder		**PGD linf, 10, 0.01**			

Seed	Training time	Benign	4/255	8/255	16/255	32/255
0	1153	75.76	67.69	58.74	41.05	26.18
1	1131	74.97	65.44	55.41	37.28	24.9
2	1125	73.58	63.9	54.51	40.3	28.85
3	1108	75.14	65.96	55.53	38.64	27.3
4	1129	75.29	66.16	56.04	39.38	27.75
Mean	**1129.2**	**74.95**	**65.83**	**56.05**	**39.33**	**27**
Std	16.1	0.82	1.37	1.6	1.46	1.51

Method	Isorandom
Batch size	64
Epochs	51
Lambda	3
Warm-up	1 epoch
Epsilon	3.5 *~16/255*
Norm	Holder

PGD linf, 10, 0.01

Seed	Training time	Benign	4/255	8/255	16/255	32/255
0	1097	80.21	70.88	60.39	40.37	24.89
1	1102	77.96	67.91	56.51	38.7	26.98
2	1017	77.52	66.97	55.4	37.47	24.17
3	1034	77.8	68.67	58.47	40.15	25.5
4	1024	80.56	71.63	60.43	40.2	24.9
Mean	**1054.8**	**78.81**	**69.21**	**58.24**	**39.38**	**25.29**
Std	41.29	1.45	1.98	2.27	1.26	1.06

Method	Temperature
Batch size	64
Epochs	51
Warm-up	2 epochs
Epsilon	1.74 *~8/255*

PGD linf, 10, 0.01

Seed	Training time	Benign	4/255	8/255	16/255	32/255
0	2929	78.23	74.9	71.44	62.42	50.4
1	2880	79.13	75.51	71.3	60.45	47.9

2	1842	79.2	75.72	71.46	60.95	47.94
3	1739	78.91	75.16	71.04	60.91	47.52
4	1735	79.47	75.56	71.33	60.33	48.04
Mean	**2225**	**78.99**	**75.37**	**71.31**	**61.01**	**48.36**
Std	622.02	0.47	0.33	0.17	0.83	1.16

Method	Eigenbound					
Epochs	51					
Lambda	2					
Warm-up	1 epoch					
Epsilon	3.5	*~16/255*				
Norm	Holder		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	3471	62.01	51.19	42.45	31.14	22.84
1	3847	60.43	49.61	41.03	31.38	22.88
2	3917	60.26	49.02	40.12	30.03	21.33
3	3889	58.41	47.82	40.12	30.42	21.59
4	3753	58.2	48	40.75	31.72	24.05
Mean	**3775.4**	**59.86**	**49.13**	**40.89**	**30.94**	**22.54**
Std	181.14	1.58	1.37	0.96	0.7	1.1

Method	Teacher					
Epochs	51					
Temp	20		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
-5	832	81.36	75.02	66.66	49.77	34.01
-4	838	81.68	74.92	65.92	48.41	32.44
-3	831	81.3	75.2	68	53.05	37.63
-2	834	81.09	74.67	66.92	51.03	35.64
-1	850	81.23	75.22	67.12	51.45	35.82
Mean	**837**	**81.33**	**75.01**	**66.92**	**50.74**	**35.11**
Std	7.75	0.22	0.23	0.75	1.75	1.97

Method	Distillation					
Epochs	51					
Temp	20		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	1385	79.07	74.1	68.37	56.36	43.22
1	1391	79.69	74.22	67.63	53.91	39.31
2	1394	78.97	74.74	69.71	59.59	47.77
3	1385	78.98	74.02	68.27	56.21	42.46
4	1380	79.29	74.34	68.16	56.13	42.61
Mean	**1387**	**79.2**	**74.28**	**68.43**	**56.44**	**43.07**
Std	5.52	0.3	0.28	0.77	2.03	3.03

Method	Gn					
Epochs	51					
Std	16/255		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	814	87.77	79.95	68.83	45.85	25.69
1	837	87.68	80.17	68.5	44.65	24.67
2	831	87.87	80.09	69.04	44.76	24.04
3	833	88.12	80.55	69.77	45.54	24.13
4	846	88.19	80.54	68.83	46.28	26.72
Mean	**832.2**	**87.93**	**80.26**	**68.99**	**45.42**	**25.05**
Std	11.69	0.22	0.27	0.47	0.7	1.14

Method	FGSM					
Epochs	51					
Budget	16/255		**PGD linf, 10, 0.01**			
Seed	**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
0	1410	83.95	82.64	81.26	77.84	70.94
1	1425	84.11	82.92	81.6	77.89	70.16
2	1433	84.05	82.66	81.17	77.72	69.93
3	1410	84.23	83.1	81.68	78.28	70.6

	4	1403	84.33	83.04	81.56	78.54	72.33
Mean		**1416.2**	**84.13**	**82.87**	**81.45**	**78.05**	**70.79**
Std		12.36	0.15	0.21	0.22	0.34	0.94

Method		Jacreg	*Randomized*				
Epochs		51					
Lambda		5		**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	0	866	88.11	78.69	65.26	40.5	23.12
	1	863	88.24	78.44	64.78	39.62	21.96
	2	857	88.27	78.82	65.01	39.28	21.05
	3	849	88.57	79.46	65.37	39.21	21.29
	4	903	88.55	79.24	65.55	41	23.75
Mean		**867.6**	**88.35**	**78.93**	**65.19**	**39.92**	**22.23**
Std		20.83	0.2	0.41	0.3	0.79	1.17

Method		Jacreg	*Randomized*				
Epochs		51					
Lambda		0.9	*(1-lbd)*ce+lbd*reg*	**PGD linf, 10, 0.01**			
Seed		**Training time**	**Benign**	**4/255**	**8/255**	**16/255**	**32/255**
	5	1042	80.56	75.88	70.36	58.69	47.61
	6	1057	80.35	75.19	69.25	57.04	43.79
	7	1041	80.39	75.41	69.92	58.17	45.87
	8	1020	80.8	75.68	69.17	56.73	43.92
	9	1020	80.53	75.58	70.15	59.02	46.45
Mean		**1036**	**80.53**	**75.55**	**69.77**	**57.93**	**45.53**
Std		15.92	0.18	0.26	0.54	1.01	1.65