

Classification error (%) on the first 1000 test samples

		original	mnist		original	fashion_mnist	
			scenario 1	scenario 2		scenario 1	scenario 2
UNENCRYPTED	CW I_2	1.49	100.00		8.30	100.00	
	CW I_0		100.00			100.00	
	CW I_∞		100.00			100.00	
	FGSM	2.10	82.94		9.50	94.25	
PERMUTATED	CW I_2	3.70	100.00	4.50	12.30	100.00	12.70
	CW I_0		100.00	7.30		100.00	12.50
	CW I_∞		100.00	5.40		100.00	12.90
	FGSM	4.19	89.14		12.00	91.82	
ECB	CW I_2	encrypt v1	16.58		55.66	irrelevant	irrelevant
		encrypt v2	18.11		41.97		
	FGSM	encrypt v1	20.88		59.23		
		encrypt v2	19.95		46.25		
CBC	CW I_2	encrypt v1	64.07	irrelevant	72.12	irrelevant	irrelevant
		encrypt v2	69.12		64.47		
	FGSM	encrypt v1	88.65		90.00		
		encrypt v2	88.65		90.00		
CTR	CW I_2	encrypt v1	88.65	irrelevant	90.00	irrelevant	irrelevant
		encrypt v2	88.65		90.00		
	FGSM	encrypt v1	88.65		90.00		
		encrypt v2	88.65		90.00		

scenario 1: the attacker gets an oracle to the attacked model which for given an image, performs the permutation, feeds through the model's layers and returns the logits

scenario 2: the attacker gets to know only the architecture, i.e. he gets an oracle to the unencrypted version

accuracies of permutation on different image sizes
(padding done with 0's around the original)

	image size	error rate	min/epoch
mnist	28x28	3.63	4
	40x40	2.65	5
	60x60	2.69	12
	100x100	2.3	14
fashion_mnist	28x28	12.4	
	40x40	12.07	13
	60x60		
	100x100		