



Progress Report Savaiz

Training

Completed
Deep Learning
Videos

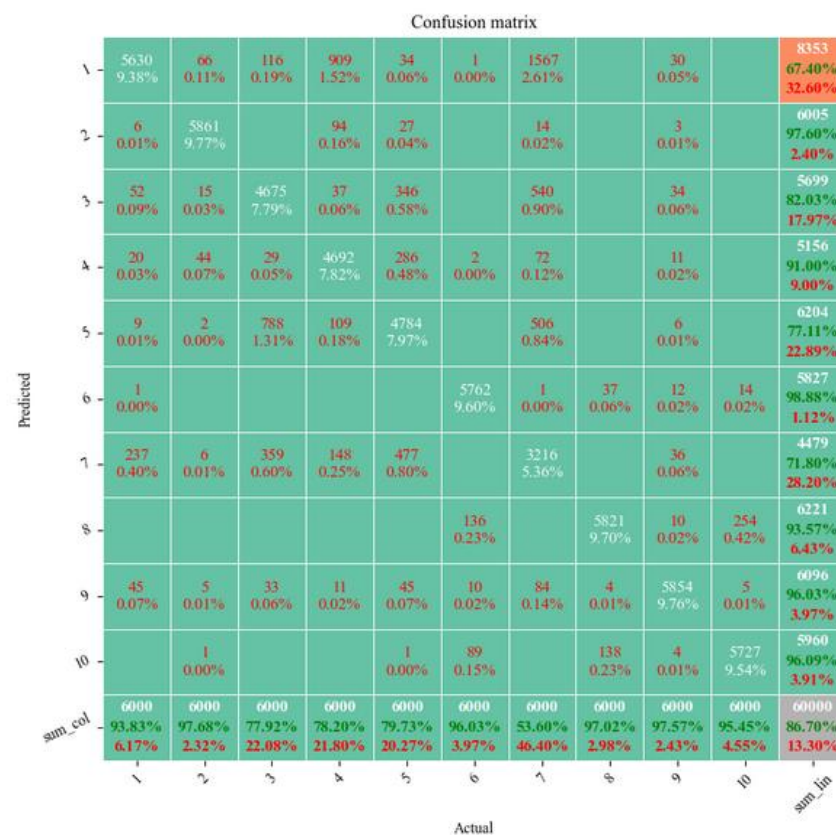
Completed
PyTorch
Tutorials

Attempted all
tasks

Clean

	1	2	3	4	5	6	7	8	9	10	sum_row
1	5896 9.83%		6 0.01%	2 0.00%	2 0.00%	1 0.00%	6 0.01%		3 0.01%	10 0.02%	5926 99.49% 0.51%
2		6673 11.12%	5 0.01%	7 0.01%	6 0.01%	5 0.01%	2 0.00%	1 0.00%	18 0.03%	1 0.00%	6718 99.33% 0.67%
3	1 0.00%	11 0.02%	5886 9.81%	29 0.05%	3 0.01%			8 0.01%	8 0.01%		5946 98.99% 1.01%
4	1 0.00%	1 0.00%	2 0.00%	6017 10.03%		18 0.03%		2 0.00%	4 0.01%	4 0.01%	6049 99.47% 0.53%
5		5 0.01%	2 0.00%		5785 9.64%		4 0.01%	7 0.01%	4 0.01%	28 0.05%	5835 99.14% 0.86%
6				32 0.05%		5371 8.95%	16 0.03%		19 0.03%	11 0.02%	5449 98.57% 1.43%
7	14 0.02%	8 0.01%	7 0.01%	1 0.00%	11 0.02%	13 0.02%	5871 9.79%				5925 99.09% 0.91%
8		41 0.07%	33 0.06%	21 0.03%	9 0.01%			6234 10.39%	1 0.00%	25 0.04%	6364 97.96% 2.04%
9	9 0.01%	1 0.00%	16 0.03%	11 0.02%	2 0.00%	11 0.02%	19 0.03%	2 0.00%	5761 9.60%	13 0.02%	5845 98.56% 1.44%
10	2 0.00%	2 0.00%	1 0.00%	11 0.02%	24 0.04%	2 0.00%		11 0.02%	33 0.06%	5857 9.76%	5943 98.55% 1.45%
col	5923 99.54% 0.46%	6742 98.98% 1.02%	5958 98.79% 1.21%	6131 98.14% 1.86%	5842 99.02% 0.98%	5421 99.08% 0.92%	5918 99.21% 0.79%	6265 99.51% 0.49%	5851 98.46% 1.54%	5949 98.45% 1.55%	60000 98.92% 1.08%
	1	2	3	4	5	6	7	8	9	10	
	Actual										

Accuracy: 0.8763833333333333



Bringing Errors % down in poisoned sampling

- Changed function based poisoning to manual.
- Accuracy brought up to 96% for poisoned dataset and 98% for clean dataset
- BadNets had 0.154% difference in errors. My results have 0.666% difference.
- Limited number of Epochs



Other Observations

50% backdoored
samples: 97%
accuracy

single target (1,9): 98%
accuracy

2 targets: 96% later 98%



Current hurdles

LIRA

- Understanding the algorithm
- Slow performance

Narcissus

- CIFAR
- Remote access