

CS-89.31: Deep Learning Generalization and Robustness

Amittai Siavava

04/27/2023

Problem 1.

Report on the results of the models.

As we can see in table 2, the first neural network, having 1024 hidden units, learns about 4 times more parameters than the second one which has 256 hidden units. This is expected since the number of hidden units in the first neural network is exactly four times as many as the number of hidden units in the second.

However, let's look at how that affects the

Final Result	Model 1	Model 2
Training Loss	1.945	1.959
Training Margin	−0.983	−0.987
Training Error	0.468	0.484
Validation Error	0.505	0.515

TABLE 1. Results.

Metric / Hyperparameter	Model 1	Model 2
Learning Rate	0.001	
Momentum	0.9	
Batch Size	64	
Epochs	25	
Stop Condition	0.01	
Dataset	CIFAR10	
Channels	3	
Classes	10	
Hidden Units	1024	256
Frobenius ₁	19.0	10.1
Frobenius ₂	4.53	4.36
Distance ₁	4.02	4.04
Distance ₂	3.15	2.88
Spectral ₁	1.5	1.5
Spectral ₂	1.93	1.84
Fro ²	86.0	44.0
$L1^2_{max}$	1300	725
Spec Dist	15.0	13.6
Dist Spec	248	119
Spec Dist sum	263	133
Spec L1max	14.8	13.4
L1max Spec	240	115
Spec L1max sum	255	129
Dist Fro	18.2	17.6
Parameters Learned	3.16×10^6	7.89×10^5
VC bound	9.19×10^9	2.03×10^9
L1max bound	3.93×10^{10}	1.21×10^9
Computed Bound	1.88×10^{10}	4.88×10^9

TABLE 2. Result comparisons for the neural network models.