Overview

Andrew Showers

Fully Connected Neural Network Experiments

	Image Size	Learning Rate	Batch Size	Epochs	Loss Function	Training Loss	Training Accuracy	Testing Loss	Testing Accuracy	Runtime (seconds)
1	32x32	$1e^{-4}$	32	100	Cross Ent.	0.56994	0.7945	0.78156	0.729	19.22
2	32x32	$1e^{-4}$	32	200	Cross Ent.	0.47305	0.828	1.02617	0.701	25.40
3	32x32	$1e^{-4}$	32	500	Cross Ent.	0.30826	0.89	1.32393	0.704	42.68
4	32x32	$1e^{-4}$	16	100	Cross Ent.	0.83195	0.691	0.92442	0.66	16.21
5	32x32	$1e^{-4}$	16	500	Cross Ent.	0.55330	0.7975	0.99478	0.71	26.73
6	32x32	$1e^{-4}$	16	1000	Cross Ent.	0.33097	0.8535	1.10919	0.694	40.67
7	32x32	$5e^{-5}$	16	500	Cross Ent.	0.48625	0.82325	0.85533	0.714	27.18
8	32x32	$5e^{-5}$	16	1000	Cross Ent.	0.29790	0.8975	0.93337	0.739	40.16
9	32x32	$5e^{-5}$	16	2000	Cross Ent.	0.17407	0.9375	1.31197	0.698	66.07
10	32x32	$1e^{-4}$	8	5000	Cross Ent.	0.26871	0.89625	1.29394	0.728	82.17

Convolutional Neural Network Experiments

	Image Size	Learning Rate	Batch Size	Epochs	Loss Function	Training Loss	Training Accuracy	Testing Loss	Testing Accuracy	Runtime (seconds)
11	32x32	$1e^{-4}$	32	100	Cross Ent.	0.99127	0.602	1.12252	0.558	67.22
12	32x32	$1e^{-4}$	32	200	Cross Ent.	0.74052	0.726	0.93981	0.651	206.34
13	32x32	$1e^{-4}$	32	500	Cross Ent.	0.65623	0.75825	0.86704	0.687	295.02
14	32x32	$1e^{-4}$	32	1000	Cross Ent.	0.43259	0.8485	0.79208	0.732	572.06
15	32x32	$3e^{-4}$	32	100	Cross Ent.	0.84011	0.68525	0.98737	0.627	68.99
16	32x32	$3e^{-4}$	32	200	Cross Ent.	0.66858	0.7485	0.87709	0.66	122.83
17	32x32	$3e^{-4}$	32	500	Cross Ent.	0.39467	0.853	0.89863	0.722	274.77

A significant problem was avoiding overfitting. By allowing the model to train for too long, or by having an overly complex model, the model would learn to fit the training data and would no longer generalize well to testing data that has never been introduced to the model. This can be seen when the training and testing accuracy begin to diverge. For example, experiments 1 and 3 differed by the number of epochs from 100 to 500. While training accuracy increased as a result, the training accuracy decreased. Since the end-goal is to have a model usable to new, untested data, the testing accuracy is the metric that should be maximized.

For the experiments above, the best test accuracy came from the fully connected neural network in experiment 8 with a test accuracy of 73.9%. This implemented a smaller batch size and learning rate, but a significant number of epochs. However, doubling the epochs as seen in experiment 9 showed significant overfitting.