

MNIST: (alpha=0.001)

Train_sample_count: 5,000

Validation_sample_count: 500

epoch_no	Validation loss	Accuracy	macro-f1
0	2.302208623582239	0.078	0.01447124304267
1	2.29950094848435	0.144	0.025174825174825
2	2.299456292505841	0.144	0.025174825174825
3	2.299488020903975	0.144	0.025174825174825
4	2.299500760329674	0.144	0.025174825174825

CIFAR-10: (alpha=0.001)

Train_sample_count: 5,000

Validation_sample_count: 500

epoch_no	Validation loss	Accuracy	macro-f1
0	2.302573841418047	0.1	0.018181818181818
1	2.302539142833419	0.1	0.018181818181818
2	2.302505152075309	0.108	0.019494584837545
3	2.302471852835557	0.108	0.020466786355476
4	2.302439229224588	0.114	0.017850637522769

Toy Dataset: (alpha=0.1)

epoch_no	Validation loss	Accuracy	macro-f1
0	1.381084839755680	0.248	0.099358974358974
10	1.031112601814972	0.496	0.417525773195876
20	0.88694462188274	0.744	0.667525773195876
30	0.797755662346022	0.992	0.992243867243867
31	0.790511982457563	1.0	1.0