

ECE-471 Selected Topics in Machine Learning

Prof. Curro

Assignment 4

Evan Bubniak

October 6, 2019

1 Results

On CIFAR-10, the CNN specified below achieves % accuracy on the training set, % accuracy on the validation set (a random selection of 10000 samples from the training set), and % accuracy on the test set. On CIFAR-100, it respectively achieves %, %, and %. Both CIFAR datasets are loaded using `tensorflow.keras.datasets` and preprocessed using Keras's `ImageDataGenerator` class, adding horizontal flips, rotations, crops, and ZCA whitening to improve regularization and expand the dataset size.

For this assignment I defined *state of the art* to be at least 90% accuracy on the CIFAR-10 dataset; I aimed to avoid deep models with more than 20 layers to optimize for training time and computing resources, as well as to allow more rapid model prototyping. This necessarily meant looking past models which had performed with high (> 95%) accuracy, but required well over a hundred epochs to train with relatively small batch sizes. I chose a batch size of 100 after initially trying a batch size of 1000, finding large batch sizes, even with batch normalization, caused drastic overfitting.

Several different techniques, including residual networks, a simple reuse of my MNIST CNN (which used only a single convolutional layer), and a deeper version of

the MNIST CNN which stacked multiple convolutional layers together, were tested. Even with normalization techniques, I found the residual network, which pooled at every layer, to drastically underperform on the training set (being no better than random guessing). Initially I suspected this to be a preprocessing mismatch between datasets, but the simpler model generalized better; the deep CNN came close to the target accuracy, achieving a top-5 categorical accuracy of 77.5% on the test set.

The final model I settled on was wide residual networks, which promised similar performance to ResNet-1001 with significantly fewer parameters and less train time.

Implementation bugs were a major headache, whereby the model not only underperformed on the training set, capping out at low accuracy even while using techniques from papers reporting high accuracy, but also achieved sub-guess accuracy on the validation set.

CIFAR-10:

Model	Training accuracy	Training top-k	Validation accuracy	validation top-k	(stop)
wide resnet (v1)	.41	.89	.099	.5089	(stop)

CIFAR-100:

Model	Training accuracy	Training top-k	Validation accuracy	validation top-k	(stop)
wide resnet (v1)	.12	.31	.021	.0868	(stop)

2 Code