Lab 4: Residual Neural Networks

Data Set

Now we will create a series of networks to classify the MNIST and CIFAR-10 datasets from keras (https://keras.io/datasets/). There are 10 classes in the both datasets. The CIFAR-10 dataset is a variety of small images with classes ranging from cars to dinosaurs.

Problem

You will need to submit your code and report. The report should include results. Also, you need to discuss your ideas and conclusions about the results. (e.g. You can say why a certain network architecture might be better than another.)

1. Use Keras like in lab 3 to load the mnist and cifar 10 database into train and test sets.
2. Prepare both datasets for use in deep learning models. Think about what shape the data needs to be in for a neural network. The default shape for a MNIST image is 28 x 28 pixels. The default shape for CIFAR-10 data is 32 x 32. Normalize the pixel data to be between 0 and 1.
3. We will use a similar baseline as in Lab 3. A sequential network with shape input, 64, 64, 64, 10. It will have a batch rate of 64 and we will use 20 epochs again. It will use the SGD optimizer with the following hyperparameters: learning rate=0.01, decay=1e-6, momentum=0.9, Nesterov=True. We didn't get a change to talk about Nesterov momentum in class but here is a link to a good article on it. https://dominikschmidt.xyz/nesterov-momentum/
4. Now create the same network but use the functional API from keras to turn it into a functional model. The documentation is here: https://keras.io/getting-started/functional-api-guide/
   The models should get around the same accuracy and training speeds and the model summary should be the same.
5. Now we will compare these models with a shallow toy ResNet that we looked at in class. The code for it can be found on blackboard. Run it on the MNIST dataset and compare its accuracy and training speed with the baseline model. Why did it affect training? Why do you think this is?
6. Now create a deeper ResNet with 10 Residual Blocks. Compare it on the MNIST data as well. How did that compare with the shallow ResNet? Should you use deeper or shallower networks? Why?
7. We will now make two shallow networks with different skip connection lengths. The number of trainable layers in a residual block as called 'skips' We will now compare a one and a three skip connection to our original two skip model. Examples of which are in the notebook on Blackboard. How do these models compare? Is there a difference? If so, why do you think that is?
8. Okay now let's do all the same on a much more complicated dataset, CIFAR-10. Create a regular function model to use as our baseline like we did for the MNIST data. How does it compare to MNIST in terms of accuracy? Why do you think that is?
9. Now let's make a shallow ResNet for CIFAR-10. How does it compare to the shallow net for MNIST and the baseline for CIFAR-10? Why do you think that is?

10. Now create a custom ResNet to solve the CIFAR-10 data. Try different learning rates, batch sizes, residual block depths, skip connections, neuron layer widths, and dropout weights. You can also try to apply a DenseNet or some other combination of layers that are novel. What combination of methods did you find that works best to improve accuracy? Why do you think this is? Justify your choices.
11. Think about how gradient descent passes through Neural Networks. Why do you think Residual networks work better than our baseline models? What could we do to generally improve our network architecture in future models?
12. Would ResNets be useful for your class group project? Discuss why or why not based on the data you are using.

Grading

Each section is worth equal points.