Data Set

In the last two labs, we have worked with standard research data used for baselines and testing models. We will do the same in this lab with the most complex baseline data CIFAR-100. We will also work with CIFAR-10 data like last time. The difference between CIFAR-10 and CIFAR-100 is there are 100 classes with fewer training examples in CIFAR-100. CIFAR-10 has 50,000 training examples per class and CIFAR-100 has 5,000 training examples. The combination of having a lot of classes with only a few training examples makes this data incredibly hard to train on.

Problem

The goal of this lab will be to understand how CNNs work and how much better they work than MLP (multi-layered perceptron) networks. You will need to submit your report. You should post your code or at least part of your code if possible, but you are not required to do so. The report should include results and can be a python notebook. 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. We are going to import CIFAR-10 and 100 with keras like we did in the last 2 labs.
2. Prepare both datasets for use in 2D Convolutional Networks. This will be different than the previous labs. Please see the released code.
3. Prepare both datasets for use in MLP Networks. This will be the same as before. We are going to create a MLP, CNN, and a CNN ResNet to compare how the different approaches work.
4. We will create a baseline model for both datasets with an MLP of the same size, shape and hyperparameters as lab 4. 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.
   How well does the MLP perform for each dataset? Is it performing above random chance? Why do you think there is a difference between the two datasets with this model? Is the model overfitting(look at the relation with training and validation accuracies)?
5. Now let's create a baseline for both of the datasets using a simple CNN with the functional api from keras. We will have this architecture as our baseline (When I give a number for a Convolutional layer it will always refer to the number of filters).
   Input, 32 conv, 32 conv, max pooling, 32 conv, 32 conv, max pooling, flatten, 512 dense, 0.5 dropout, 10/100 dense(output). We will use 10 epochs for all of the CNNs. This is because of the time it takes to train and our current situation with lack of access to the computer lab.
   How does the model compare to the MLP? Why do you think this is?
6. Now let's mess with the architecture some. Change all of the convolutional layers to have 64 filters for the CIFAR-10 data. How does using more filters affect training speed and accuracy?
7. Now let's do the opposite and change them all to 16. How does using less filters affect training speed and accuracy?

8. Now let's remove our pooling layers so the network is: input, 32 conv, 32 conv, 32 conv, 32 conv, flatten, 512 dense, 10/100 dense. How does removing the pooling layers affect training? Based off of your knowledge of what pooling does, why do you think this is?
9. Now compare the baseline model with data augmentation (Data augmentation can take a very long time, so I have pre-ran the model in the sample code. You can choose to use those results or rerun it yourself). How does augmenting the data change results? What is the augmentation doing? Why do you think the results changed so much? Is it closer in training and validation scores? Why do you think that is?
10. Now let's compare our models with a CNN resnet. For simplicity, you will be provided the architecture for CIFAR-10. Now create a CIFAR-100 CNN ResNet. How do these compare to the baselines? Why do you think that is based off of what you know about CNNs and ResNets. How does the CIFAR-10 CNN ResNet compare to the MLP ResNet from lab 4?
11. Now create a custom CNN that rivals the CNN ResNet in performance for both speed and accuracy (if you don't know enough, reference the Keras docs and ask questions in the discussion board) for CIFAR-100. Use any number of filters and layers you wish. Can you? If not, why not? If you did, what decisions did you make to increase the accuracy and speed of regular CNNs to the level of CNN ResNets?

You are graded on your responses to the questions in parts 4 to 11, with each part having equal points.