

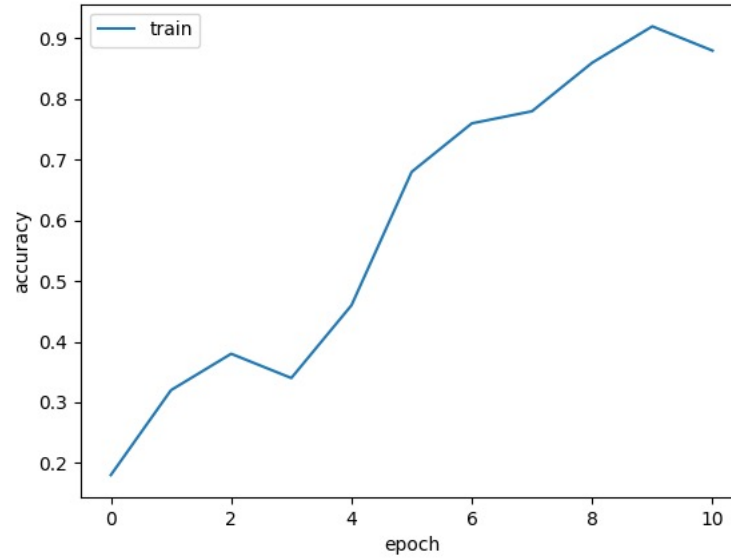
Assignment 2 Writeup

Name: Willie Zeng

GT Email: wzeng40@gatech.edu

Part-1 ConvNet

Put your learning curve here:



My CNN Model

Describe your model design in plain text here:

3 Convolution layers (ReLU activated) were used to extract High-Level features. 1 Conv. Layer would only extract low-level features such as lines/edges, which is not enough to get high accuracy predictions. Furthermore, 3 Maxpool layers (after each Conv) were used to discard noisy activations and decrease computation power. Classification is handled by the same fully-connected layer used in CNN and the twolayer model. I drew inspiration from VGG16 and used a 3x3 kernel to decrease the training time.

Describe your choice of hyper-parameters:

I used 15 epochs to give the model enough training to converge. A LR greater than 0.009 would result to spotty performances because the model would learn “too fast/much” from certain portions of data, only to perform poorly in other portions. A LR less than 0.006 would take too long to converge. I decided on 0.0088. Regularization is 0.002 to avoid overfitting by reducing variance. No warmup required and steps (6,8) and momentum were kept at default for training speed and accuracy.

What's your final accuracy on validation set?

0.725

Data Wrangling

What's your result of training with regular CE loss on imbalanced CIFAR-10?

Fill in your per-class accuracy in the table

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
CE Loss	0.929 0	0.883 0	0.459 0	0.3890	0.1270	0.0330	0.0890	0.0040	0.0000	0.0000

What's your result of training with CB-Focal loss on imbalanced CIFAR-10?

Tune the hyper-parameter beta and fill in your per-class accuracy in the table

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
beta=0.9 999	0.610 0	0.659 0	0.239 0	0.1660	0.2180	0.2750	0.3490	0.3700	0.3430	0.3870
beta=0.7 7	0.890 0	0.803 0	0.490	0.293	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000

Put your results of CE loss and CB-Focal Loss(best) together:

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
CE Loss	0.929 0	0.883 0	0.459 0	0.3890	0.1270	0.0330	0.0890	0.0040	0.0000	0.0000
CB-Focal	0.610 0	0.659 0	0.239 0	0.1660	0.2180	0.2750	0.3490	0.3700	0.3430	0.3870

Describe and explain your observation on the result:

The default (softmax) CE loss does not take into account the effective number of samples (shown below) and will result in lower accuracies. CB Focal loss re-weights the effective number of samples for every class (10 classes in our example). This is important because our data is a long-tail dataset (imbalanced). The easy examples (ie. already close to perfect probability) are down-weighted so that their contribution to the total loss is small. The obtained results are supported by the paper: a decrease in majority class (0-4) accuracy as well as an improvement in minority class accuracy (5-9) as beta increases from 0.77 to 0.9999.

$$E_n = (1 - \beta^n)/(1 - \beta)$$

Effective number of samples