Assignment 2 Theory Problem Set

DO NOT TAG

Name: Ken Chen
GT Email: kchen351@gatech.edu

Theory PS Q1. Feel free to add extra slides if needed.

1.

$$X' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & X(0,0) & X(0,1) & X(0,2) & 0 & 0 \\ 0 & 0 & X(1,0) & X(1,1) & X(1,2) & 0 & 0 \\ 0 & 0 & X(2,0) & X(2,1) & X(2,2) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

kernel_size:

$\frac{3-3+1+4}{4}+1=2$

$$X' * W = \begin{bmatrix} X(0,0)\cdot W(2,2) & X(0,2)\cdot W(2,0) \\ X(2,0)\cdot W(0,2) & X(2,2)\cdot W(0,0) \end{bmatrix}$$

$$Y = A X$$

$$\begin{bmatrix} X(0,0)\cdot W(2,2) \\ X(0,2)\cdot W(2,0) \\ X(2,0)\cdot W(0,2) \\ X(2,2)\cdot W(0,0) \end{bmatrix} = \begin{bmatrix} W(2,2) & W(2,0) & W(0,2) & W(0,0) \end{bmatrix} \begin{bmatrix} X(0,0) \\ X(0,1) \\ X(0,2) \\ \vdots \\ X(2,1) \\ X(2,2) \end{bmatrix}$$

1 3 6 9

Y = A X

Theory PS Q2. Feel free to add extra slides if needed.

2.

$$h_1 = W^{(1)}x + b^{(1)} = \begin{bmatrix} 1.5x \\ 0.5x+1 \end{bmatrix}$$

$$r_1 = \max(0, h_1) = \begin{cases} \begin{bmatrix} 1.5x \\ 0.5x+1 \end{bmatrix}, & x > 0 \\ \begin{bmatrix} 0 \\ 0.5x+1 \end{bmatrix}, & -2 < x \le 0 \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x \le -2 \end{cases}$$

$$h_2 = W^{(2)} \cdot r_1 + b^{(2)} = \begin{cases} \begin{bmatrix} 2.5x+2 \\ 3.5x+1 \end{bmatrix}, & x > 0 \\ \begin{bmatrix} 0.5x+1 \\ x+2 \end{bmatrix}, & -2 < x \le 0 \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x \le -2 \end{cases}$$

$$r_2 = \max(0, h_2) = \begin{cases} \begin{bmatrix} 2.5x+2 \\ 3.5x+1 \end{bmatrix}, & x > 0 \\ \begin{bmatrix} 0.5x+1 \\ x+2 \end{bmatrix}, & -2 < x \le 0 \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x \le -2 \end{cases}$$

$$h(x) = W^{(3)} \cdot r_2 + b^{(3)} = \begin{cases} 6x+3, & x > 0 \\ 1.5x+3, & -2 < x \le 0 \\ 0, & x \le -2 \end{cases}$$

$x_0 = 1$ or 2, $w = 6$, $b = 3$

$x_0 = -1$, $w = 1.5$, $b = 3$

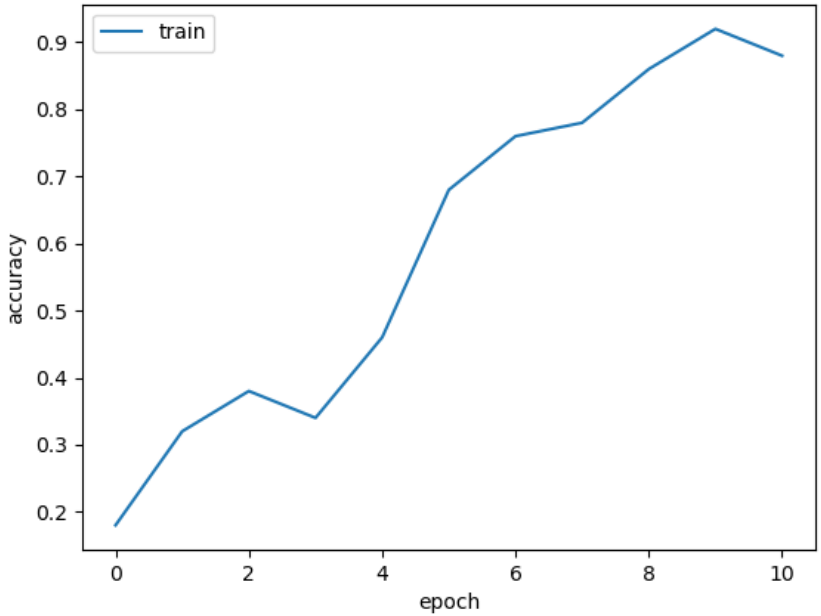
Assignment 2 Writeup

DO NOT TAG

Part-1 ConvNet

DO NOT TAG

Put your learning curve here:



My CNN Model

DO NOT TAG

Describe and justify your model design in plain text here:

I build my model based on the idea of ResNet and create shortcuts among layers.

- The network has three convolution block. Each block consists of three convolution layer followed by relu activation layer and batch normalization layer. At the end, there are two fully connected layer.
- The channels for each block are 32, 64, 128 and remains the same within a block.
- There is a max pooling layer at the end of each block.
- Two shortcuts are created. The first one is between block 1 and block2 and the second one is between block2 and block3.
- Use same padding to keep the same dimensions in convolution layers.

Describe and justify your choice of hyper-parameters:

- batch_size: 128. Remains the same as the naïve convolution model.
- learning_rate: 0.01. Since we have created shortcuts in the network, it is easier to optimize the model.
- regularization: 0.0004. This parameter is not sensitive, so I just keep it.
- epochs: 10. Remains the same as the naïve convolution model.
- momentum: 0.95. Chosen arbitrarily.

What’s your final accuracy on validation set?

0.8577

Data Wrangling

DO NOT TAG

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	0.890	0.742	0.446	0.207	0.007	0.004	0.000	0.000	0.000	0.000

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=0.99	0.861	0.772	0.429	0.355	0.057	0.004	0.001	0.000	0.000	0.000
beta=0.9999	0.578	0.675	0.310	0.178	0.237	0.170	0.318	0.413	0.486	0.176

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	0.890	0.742	0.446	0.207	0.007	0.004	0.000	0.000	0.000	0.000
CB-Focal	0.578	0.675	0.310	0.178	0.237	0.170	0.318	0.413	0.486	0.176

Describe and explain your observation on the result:

*Explanation should go into **WHY** things work the way they do in the context of Machine Learning theory/intuition, along with justification for your experimentation methodology. **DO NOT** just describe the results, you should explain the reasoning behind your choices and what behavior you expected. Also, be cognizant of the best way to mindful and show the results that best emphasizes your key observations. If you need more than one slide to answer the question, you are free to create new slides.*

Findings:

- For regular CE Loss in the imbalanced dataset, the accuracy is below 10% for class 4-9 and above 0.7 for class 1-2.
- When CB-Focal is applied with beta=0.99, the accuracy increases a little bit for class 2, 3 but decreases for class 1, 4. All the changes are not significant.
- When CB-Focal is applied with beta=0.9999, the accuracy increases a lot for class 4-9, but decreases for class 1-3. The increases are phenomenon while the decreases are modest.

Explanations:

beta=0.99, weight=[0.78, 0.78, 0.78, 0.78, 0.78, 0.80, 0.86, 1.04, 1.39, 1.98]
beta=0.9999, weight=[0.05, 0.08, 0.12, 0.20, 0.32, 0.52, 0.87, 1.44, 2.40, 4.00]

- Focal loss get weighted based on the parameter beta. Larger the beta, higher the weight and the first term multiplied by the loss. The weight get higher for classes with lower samples. Therefore, we see that larger beta parameter gets more balanced accuracy among classes.
- As we are balancing the accuracy among classes, we sacrifice the accuracy for the whole model and the accuracy for classes with large sample size. Therefore, we see the decrease in accuracy in class 1-3.