CS 231N
Convolutional Neural Networks for Visual Recognition
Winter 2016 Sample Midterm Exam

February 10, 2016

Full Name: _____

Question	Score
Multiple Choice (20 pts)	
True/False (20 pts)	
Short Answer (60 pts)	
Total (100 pts)	

Welcome to the CS231N Midterm Exam!

- The exam is 1 hour 15 minutes.
- No notes or electronic devices are allowed.

I understand and agree to uphold the Stanford Honor Code during this exam.

Signature: _____ Date: _____

Good luck!

1 Multiple Choice (20 points)

Circle the letters of your choice.

Each question is worth 2 points. Each one of the four individual choices is 0.5 points for a correct answer, or 0 points otherwise.

1. You start training your Neural Network but the loss is almost completely flat. What could be the cause?
   (a) The learning rate could be too low
   (b) The regularization strength could be too high
   (c) The class distribution could be very uneven in the dataset
   (d) The weight initialization scale could be incorrectly set

2. A VGGNet only uses a sequence of 3x3 CONV with stride 1 pad 1 and 2x2 POOL stride 2 pad 0 layers. It eventually transitions to Fully Connected layers and the classifier. There are 5 POOL layers in total. On ImageNet, the VGGNet takes 224x224 images. If we tried to run the VGGNet on a 32x32 input (e.g. CIFAR-10 image):
   (a) The code would crash on the very first CONV layer because 3x3 filters with stride 1 pad 1 wouldn't "fit" across 32x32 input
   (b) The amount of memory needed to store the forward activations in the first CONV layer would be reduced by a factor of 7 (since $224/32 = 7$)
   (c) The network would run fine until the very first Fully Connected layer, where it would crash
   (d) The network would run forward just fine but its predictions would, of course, be ImageNet class predictions

2 True / False (20 points)

For these problems, no explanation is required; simply circle True, False, or do not give an answer.

Scoring: There are 20 questions, each worth 1 point. To discourage guessing, incorrect answers are worth -1 point. Leaving a question blank will give 0 points.

1. **True or False:** Your Neural Network is not gradient checking. It could be that it's because you're using the Adagrad update instead of Vanilla SGD.

2. **True or False:** If the input to a ConvNet is a zero image (all zeros), then the class probabilities will come out uniform.

3. **True or False:** Turning off L2 weight regularization will likely lead to higher accuracy on the training set.

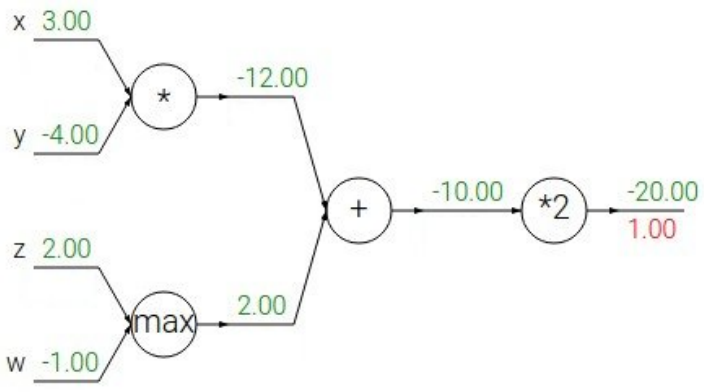
4. **True or False:** It's sufficient for symmetry breaking in a Neural Network to initialize all weights to 0, provided that the biases are random

3 Short Answer (60 points)

Answer each question in provided space.

3.1 Backpropagation

Fill in the missing gradients underneath the forward pass activations in each circuit diagram. The gradient of the output with respect to the loss is one (1.00) for every circut, and has already been filled in.



3.2 Convolutional Architectures

Consider the convolutional network defined by the layers in the left column below. Fill in the size of the activation volumes at each layer, and the number of parameters at each layer. You can write your answer as a multiplication (e.g. 128x128x3).

- CONV5-N denotes a convolutional layer with N neurons, each having 5x5xD filters, where D is the depth of the activation volume at the previous layer. Padding is 2, and stride is 1.
- POOL2 denotes a 2x2 max-pooling layer with stride 2 (pad 0)
- FC-N denotes a fully-connected layer with N neurons.

Layer	Activation Volume Dimensions (memory)	Number of parameters
INPUT	32x32x1	0
CONV5-10		
POOL2		
CONV5-10		
POOL2		
FC-10		