### layers

February 11, 2020

#### 1 Modular neural nets

In the previous exercise, we computed the loss and gradient for a two-layer neural network in a single monolithic function. This isn't very difficult for a small two-layer network, but would be tedious and error-prone for larger networks. Ideally we want to build networks using a more modular design so that we can snap together different types of layers and loss functions in order to quickly experiment with different architectures.

In this exercise we will implement this approach, and develop a number of different layer types in isolation that can then be easily plugged together. For each layer we will implement forward and backward functions. The forward function will receive data, weights, and other parameters, and will return both an output and a cache object that stores data needed for the backward pass. The backward function will receive upstream derivatives and the cache object, and will return gradients with respect to the data and all of the weights. This will allow us to write code that looks like this:

```
def two_layer_net(X, W1, b1, W2, b2, reg):
        # Forward pass; compute scores
        s1, fc1_cache = affine_forward(X, W1, b1)
        a1, relu_cache = relu_forward(s1)
        scores, fc2_cache = affine_forward(a1, W2, b2)
        # Loss functions return data loss and gradients on scores
        data_loss, dscores = svm_loss(scores, y)
        # Compute backward pass
        da1, dW2, db2 = affine backward(dscores, fc2 cache)
        ds1 = relu_backward(da1, relu_cache)
        dX, dW1, db1 = affine backward(ds1, fc1 cache)
        # A real network would add regularization here
        # Return loss and gradients
        return loss, dW1, db1, dW2, db2
[1]: # As usual, a bit of setup
     import numpy as np
     import matplotlib.pyplot as plt
```

### 2 Affine layer: forward

Open the file cs231n/layers.py and implement the affine\_forward function.

Once you are done we will test your can test your implementation by running the following:

```
[2]: # Test the affine_forward function
     num_inputs = 2
     input\_shape = (4, 5, 6)
     output_dim = 3
     input_size = num_inputs * np.prod(input_shape)
     weight_size = output_dim * np.prod(input_shape)
     x = np.linspace(-0.1, 0.5, num=input_size).reshape(num_inputs, *input_shape)
     w = np.linspace(-0.2, 0.3, num=weight_size).reshape(np.prod(input_shape),_
     →output_dim)
     b = np.linspace(-0.3, 0.1, num=output dim)
     out, _ = affine_forward(x, w, b)
     correct_out = np.array([[ 1.49834967,  1.70660132,  1.91485297],
                             [ 3.25553199, 3.5141327, 3.77273342]])
     # Compare your output with ours. The error should be around 1e-9.
     print('Testing affine_forward function:')
     print('difference: ', rel_error(out, correct_out))
```

```
Testing affine_forward function: difference: 9.769847728806635e-10
```

#### 3 Affine layer: backward

Now implement the affine\_backward function. You can test your implementation using numeric gradient checking.

```
[10]: # Test the affine_backward function
      x = np.random.randn(10, 2, 3)
      w = np.random.randn(6, 5)
      b = np.random.randn(5)
      dout = np.random.randn(10, 5)
      dx num = eval numerical gradient array(lambda x: affine forward(x, w, b)[0], x, u
      dw_num = eval_numerical_gradient_array(lambda w: affine_forward(x, w, b)[0], w,__
      -dout)
      db num = eval numerical gradient array(lambda b: affine forward(x, w, b)[0], b, |
      -dout)
      _, cache = affine_forward(x, w, b)
      dx, dw, db = affine backward(dout, cache)
      # The error should be less than 1e-10
      print('Testing affine_backward function:')
      print('dx error: ', rel_error(dx_num, dx))
      print('dw error: ', rel_error(dw_num, dw))
      print('db error: ', rel_error(db_num, db))
```

Testing affine\_backward function: dx error: 6.229740884137868e-10 dw error: 8.216698307090757e-11 db error: 1.537238419837524e-10

# 4 ReLU layer: forward

Implement the relu\_forward function and test your implementation by running the following:

```
[ 0.22727273, 0.31818182, 0.40909091, 0.5, ]])
# Compare your output with ours. The error should be around 1e-8
print('Testing relu_forward function:')
print('difference: ', rel_error(out, correct_out))
```

Testing relu\_forward function: difference: 4.999999798022158e-08

#### 5 ReLU layer: backward

Implement the relu\_backward function and test your implementation using numeric gradient checking:

```
[13]: x = np.random.randn(10, 10)
dout = np.random.randn(*x.shape)

dx_num = eval_numerical_gradient_array(lambda x: relu_forward(x)[0], x, dout)

_, cache = relu_forward(x)
dx = relu_backward(dout, cache)

# The error should be around 1e-12
print('Testing relu_backward function:')
print('dx error: ', rel_error(dx_num, dx))
```

Testing relu\_backward function: dx error: 3.275612814312212e-12

## 6 Loss layers: Softmax and SVM

You implemented these loss functions in the last assignment, so we'll give them to you for free here. It's still a good idea to test them to make sure they work correctly.

```
[14]: num_classes, num_inputs = 10, 50
    x = 0.001 * np.random.randn(num_inputs, num_classes)
    y = np.random.randint(num_classes, size=num_inputs)

dx_num = eval_numerical_gradient(lambda x: svm_loss(x, y)[0], x, verbose=False)
    loss, dx = svm_loss(x, y)

# Test svm_loss function. Loss should be around 9 and dx error should be 1e-9
    print('Testing svm_loss:')
    print('loss: ', loss)
    print('dx error: ', rel_error(dx_num, dx))
```

Testing svm\_loss:

loss: 8.999901006028317

dx error: 3.6226026387582733e-09

Testing softmax\_loss:

loss: 2.3025756811914233

dx error: 1.0773563075545908e-08

#### 7 Convolution layer: forward naive

We are now ready to implement the forward pass for a convolutional layer. Implement the function conv\_forward\_naive in the file cs231n/layers.py.

You don't have to worry too much about efficiency at this point; just write the code in whatever way you find most clear.

You can test your implementation by running the following:

```
[62]: x_{shape} = (2, 3, 4, 4)
      w_{shape} = (3, 3, 4, 4)
      x = np.linspace(-0.1, 0.5, num=np.prod(x_shape)).reshape(x_shape)
      w = np.linspace(-0.2, 0.3, num=np.prod(w_shape)).reshape(w_shape)
      b = np.linspace(-0.1, 0.2, num=3)
      conv_param = {'stride': 2, 'pad': 1}
      out, _ = conv_forward_naive(x, w, b, conv_param)
      print(out)
      correct_out = np.array([[[[[-0.08759809, -0.10987781],
                                 [-0.18387192, -0.2109216]],
                                [[ 0.21027089, 0.21661097],
                                 [ 0.22847626, 0.23004637]],
                                [[ 0.50813986, 0.54309974],
                                 [ 0.64082444, 0.67101435]]],
                               [[[-0.98053589, -1.03143541],
                                 [-1.19128892, -1.24695841]],
                                [[ 0.69108355, 0.66880383],
                                 [ 0.59480972, 0.56776003]],
                                [[ 2.36270298, 2.36904306],
                                 [ 2.38090835, 2.38247847]]]])
```

```
# Compare your output to ours; difference should be around 1e-8
print('Testing conv_forward_naive')
print('difference: ', rel_error(out, correct_out))
[[[[-0.08759809 -0.10987781]
   [-0.18387192 -0.2109216 ]]
  [[ 0.21027089  0.21661097]
  [ 0.22847626  0.23004637]]
 [[ 0.50813986  0.54309974]
  [ 0.64082444   0.67101435]]]
 [[[-0.98053589 -1.03143541]
   [-1.19128892 -1.24695841]]
 [ 0.59480972  0.56776003]]
 [[ 2.36270298  2.36904306]
  Testing conv_forward_naive
difference: 2.2121476417505994e-08
```

## 8 Aside: Image processing via convolutions

As fun way to both check your implementation and gain a better understanding of the type of operation that convolutional layers can perform, we will set up an input containing two images and manually set up filters that perform common image processing operations (grayscale conversion and edge detection). The convolution forward pass will apply these operations to each of the input images. We can then visualize the results as a sanity check.

```
[63]: from scipy.misc import imread, imresize

kitten, puppy = imread('kitten.jpg'), imread('puppy.jpg')
# kitten is wide, and puppy is already square
d = kitten.shape[1] - kitten.shape[0]
kitten_cropped = kitten[:, d//2:-d//2, :]

img_size = 200  # Make this smaller if it runs too slow
x = np.zeros((2, 3, img_size, img_size))
x[0, :, :, :] = imresize(puppy, (img_size, img_size)).transpose((2, 0, 1))
x[1, :, :, :] = imresize(kitten_cropped, (img_size, img_size)).transpose((2, 0, ⊥
→1))
```

```
# Set up a convolutional weights holding 2 filters, each 3x3
w = np.zeros((2, 3, 3, 3))
# The first filter converts the image to grayscale.
# Set up the red, green, and blue channels of the filter.
w[0, 0, :, :] = [[0, 0, 0], [0, 0.3, 0], [0, 0, 0]]
w[0, 1, :, :] = [[0, 0, 0], [0, 0.6, 0], [0, 0, 0]]
w[0, 2, :, :] = [[0, 0, 0], [0, 0.1, 0], [0, 0, 0]]
# Second filter detects horizontal edges in the blue channel.
w[1, 2, :, :] = [[1, 2, 1], [0, 0, 0], [-1, -2, -1]]
# Vector of biases. We don't need any bias for the grayscale
# filter, but for the edge detection filter we want to add 128
# to each output so that nothing is negative.
b = np.array([0, 128])
# Compute the result of convolving each input in x with each filter in w,
# offsetting by b, and storing the results in out.
out, _ = conv_forward_naive(x, w, b, {'stride': 1, 'pad': 1})
def imshow noax(img, normalize=True):
    """ Tiny helper to show images as uint8 and remove axis labels """
    if normalize:
        img_max, img_min = np.max(img), np.min(img)
        img = 255.0 * (img - img min) / (img max - img min)
   plt.imshow(img.astype('uint8'))
   plt.gca().axis('off')
# Show the original images and the results of the conv operation
plt.subplot(2, 3, 1)
imshow_noax(puppy, normalize=False)
plt.title('Original image')
plt.subplot(2, 3, 2)
imshow_noax(out[0, 0])
plt.title('Grayscale')
plt.subplot(2, 3, 3)
imshow_noax(out[0, 1])
plt.title('Edges')
plt.subplot(2, 3, 4)
imshow noax(kitten cropped, normalize=False)
plt.subplot(2, 3, 5)
imshow noax(out[1, 0])
plt.subplot(2, 3, 6)
imshow_noax(out[1, 1])
plt.show()
```

/Users/ssipani/miniconda3/envs/cs7643/lib/python3.7/site-packages/ipykernel\_launcher.py:3: DeprecationWarning: `imread` is deprecated! `imread` is deprecated in SciPy 1.0.0, and will be removed in 1.2.0. Use ``imageio.imread`` instead.

This is separate from the ipykernel package so we can avoid doing imports until

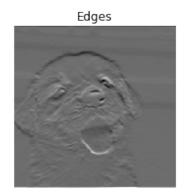
/Users/ssipani/miniconda3/envs/cs7643/lib/python3.7/site-packages/ipykernel\_launcher.py:10: DeprecationWarning: `imresize` is deprecated! `imresize` is deprecated in SciPy 1.0.0, and will be removed in 1.2.0. Use ``skimage.transform.resize`` instead.

# Remove the CWD from sys.path while we load stuff.
/Users/ssipani/miniconda3/envs/cs7643/lib/python3.7/sitepackages/ipykernel\_launcher.py:11: DeprecationWarning: `imresize` is deprecated!
`imresize` is deprecated in SciPy 1.0.0, and will be removed in 1.2.0.
Use ``skimage.transform.resize`` instead.

# This is added back by InteractiveShellApp.init\_path()

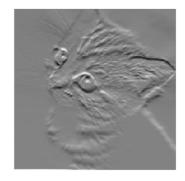












## 9 Convolution layer: backward naive

Next you need to implement the function conv\_backward\_naive in the file cs231n/layers.py. As usual, we will check your implementation with numeric gradient checking.

```
[70]: x = np.random.randn(4, 3, 5, 5)
     w = np.random.randn(2, 3, 3, 3)
     b = np.random.randn(2,)
     dout = np.random.randn(4, 2, 5, 5)
     conv param = {'stride': 1, 'pad': 1}
     dx num = eval_numerical_gradient_array(lambda x: conv_forward_naive(x, w, b,_
      dw_num = eval_numerical_gradient_array(lambda w: conv_forward_naive(x, w, b,_
      →conv_param)[0], w, dout)
     db num = eval numerical gradient array(lambda b: conv forward naive(x, w, b, ...
      out, cache = conv_forward_naive(x, w, b, conv_param)
     dx, dw, db = conv backward naive(dout, cache)
     # print(dx)
     # print(dw)
     # Your errors should be around 1e-9'
     print('Testing conv_backward_naive function')
     print('dx error: ', rel_error(dx, dx_num))
     print('dw error: ', rel_error(dw, dw_num))
     print('db error: ', rel_error(db, db_num))
```

Testing conv\_backward\_naive function dx error: 1.562511213608746e-09 dw error: 1.3934435161566938e-08 db error: 8.164009949533664e-12

# 10 Max pooling layer: forward naive

The last layer we need for a basic convolutional neural network is the max pooling layer. First implement the forward pass in the function max\_pool\_forward\_naive in the file cs231n/layers.py.

```
[72]: x_shape = (2, 3, 4, 4)
x = np.linspace(-0.3, 0.4, num=np.prod(x_shape)).reshape(x_shape)
pool_param = {'pool_width': 2, 'pool_height': 2, 'stride': 2}

out, _ = max_pool_forward_naive(x, pool_param)

correct_out = np.array([[[[-0.26315789, -0.24842105],
```

```
[-0.20421053, -0.18947368]],
[[-0.14526316, -0.13052632],
[-0.08631579, -0.07157895]],
[[-0.02736842, -0.01263158],
[0.03157895, 0.04631579]]],
[[[0.09052632, 0.10526316],
[0.14947368, 0.16421053]],
[[0.20842105, 0.22315789],
[0.26736842, 0.28210526]],
[[0.32631579, 0.34105263],
[[0.38526316, 0.4]]]]])

# Compare your output with ours. Difference should be around 1e-8.
print('Testing max_pool_forward_naive function:')
print('difference: ', rel_error(out, correct_out))
```

Testing max\_pool\_forward\_naive function: difference: 4.1666665157267834e-08

#### 11 Max pooling layer: backward naive

Implement the backward pass for a max pooling layer in the function max\_pool\_backward\_naive in the file cs231n/layers.py. As always we check the correctness of the backward pass using numerical gradient checking.

Testing max\_pool\_backward\_naive function: dx error: 3.275630137940067e-12

## 12 Fast layers

Making convolution and pooling layers fast can be challenging. To spare you the pain, we've provided fast implementations of the forward and backward passes for convolution and pooling layers in the file cs231n/fast layers.py.

The fast convolution implementation depends on a Cython extension; to compile it you need to run the following from the cs231n directory:

```
python setup.py build_ext --inplace
```

The API for the fast versions of the convolution and pooling layers is exactly the same as the naive versions that you implemented above: the forward pass receives data, weights, and parameters and produces outputs and a cache object; the backward pass receives upstream derivatives and the cache object and produces gradients with respect to the data and weights.

**NOTE:** The fast implementation for pooling will only perform optimally if the pooling regions are non-overlapping and tile the input. If these conditions are not met then the fast pooling implementation will not be much faster than the naive implementation.

You can compare the performance of the naive and fast versions of these layers by running the following:

```
[74]: from cs231n.fast_layers import conv_forward_fast, conv_backward_fast
      from time import time
      x = np.random.randn(100, 3, 31, 31)
      w = np.random.randn(25, 3, 3, 3)
      b = np.random.randn(25,)
      dout = np.random.randn(100, 25, 16, 16)
      conv_param = {'stride': 2, 'pad': 1}
      t0 = time()
      out_naive, cache_naive = conv_forward_naive(x, w, b, conv_param)
      t1 = time()
      out_fast, cache_fast = conv_forward_fast(x, w, b, conv_param)
      t2 = time()
      print('Testing conv_forward_fast:')
      print('Naive: %fs' % (t1 - t0))
      print('Fast: %fs' % (t2 - t1))
      print('Speedup: %fx' % ((t1 - t0) / (t2 - t1)))
      print('Difference: ', rel_error(out_naive, out_fast))
      t0 = time()
      dx naive, dw_naive, db_naive = conv_backward_naive(dout, cache_naive)
      dx_fast, dw_fast, db_fast = conv_backward_fast(dout, cache_fast)
      t2 = time()
      print('\nTesting conv_backward_fast:')
      print('Naive: %fs' % (t1 - t0))
      print('Fast: %fs' % (t2 - t1))
      print('Speedup: %fx' % ((t1 - t0) / (t2 - t1)))
      print('dx difference: ', rel_error(dx_naive, dx_fast))
```

```
print('dw difference: ', rel_error(dw_naive, dw_fast))
      print('db difference: ', rel_error(db_naive, db_fast))
     Testing conv_forward_fast:
     Naive: 10.193678s
     Fast: 0.027950s
     Speedup: 364.713683x
     Difference: 1.1552716289304518e-11
     Testing conv_backward_fast:
     Naive: 17.836557s
     Fast: 0.025802s
     Speedup: 691.288595x
     dx difference: 1.6750554469420988e-11
     dw difference: 8.182503010119549e-13
     db difference: 0.0
[75]: from cs231n.fast_layers import max_pool_forward_fast, max_pool_backward_fast
      x = np.random.randn(100, 3, 32, 32)
      dout = np.random.randn(100, 3, 16, 16)
      pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}
      t0 = time()
      out_naive, cache_naive = max_pool_forward_naive(x, pool_param)
      t1 = time()
      out_fast, cache_fast = max_pool_forward_fast(x, pool_param)
      t2 = time()
      print('Testing pool_forward_fast:')
      print('Naive: %fs' % (t1 - t0))
      print('fast: %fs' % (t2 - t1))
      print('speedup: %fx' % ((t1 - t0) / (t2 - t1)))
      print('difference: ', rel_error(out_naive, out_fast))
      t0 = time()
      dx_naive = max_pool_backward_naive(dout, cache_naive)
      t1 = time()
      dx_fast = max_pool_backward_fast(dout, cache_fast)
      t2 = time()
      print('\nTesting pool_backward_fast:')
      print('Naive: %fs' % (t1 - t0))
      print('speedup: %fx' % ((t1 - t0) / (t2 - t1)))
      print('dx difference: ', rel_error(dx_naive, dx_fast))
     Testing pool_forward_fast:
```

Naive: 0.792710s

fast: 0.005289s speedup: 149.883605x difference: 0.0 Testing pool\_backward\_fast: Naive: 0.754867s speedup: 31.264054x dx difference: 0.0

#### 13 Sandwich layers

There are a couple common layer "sandwiches" that frequently appear in ConvNets. For example convolutional layers are frequently followed by ReLU and pooling, and affine layers are frequently followed by ReLU. To make it more convenient to use these common patterns, we have defined several convenience layers in the file cs231n/layer\_utils.py. Lets grad-check them to make sure that they work correctly:

```
[76]: from cs231n.layer_utils import conv_relu_pool_forward, conv_relu_pool_backward
      x = np.random.randn(2, 3, 16, 16)
      w = np.random.randn(3, 3, 3, 3)
      b = np.random.randn(3,)
      dout = np.random.randn(2, 3, 8, 8)
      conv_param = {'stride': 1, 'pad': 1}
      pool param = {'pool height': 2, 'pool width': 2, 'stride': 2}
      out, cache = conv_relu_pool_forward(x, w, b, conv_param, pool_param)
      dx, dw, db = conv_relu_pool_backward(dout, cache)
      dx_num = eval_numerical_gradient_array(lambda x: conv_relu_pool_forward(x, w,_
       →b, conv_param, pool_param)[0], x, dout)
      dw_num = eval_numerical_gradient_array(lambda w: conv_relu_pool_forward(x, w,_
      →b, conv_param, pool_param)[0], w, dout)
      db_num = eval_numerical_gradient_array(lambda b: conv_relu_pool_forward(x, w,_
       →b, conv_param, pool_param)[0], b, dout)
      print('Testing conv_relu_pool_forward:')
      print('dx error: ', rel_error(dx_num, dx))
      print('dw error: ', rel_error(dw_num, dw))
      print('db error: ', rel_error(db_num, db))
```

Testing conv\_relu\_pool\_forward: dx error: 1.4306251613522052e-08 dw error: 5.693903828917013e-10 db error: 3.699192285144267e-11

```
[77]: from cs231n.layer_utils import conv_relu_forward, conv_relu_backward
      x = np.random.randn(2, 3, 8, 8)
      w = np.random.randn(3, 3, 3, 3)
      b = np.random.randn(3,)
      dout = np.random.randn(2, 3, 8, 8)
      conv param = {'stride': 1, 'pad': 1}
      out, cache = conv relu forward(x, w, b, conv param)
      dx, dw, db = conv_relu_backward(dout, cache)
      dx_num = eval_numerical_gradient_array(lambda x: conv_relu_forward(x, w, b,_
       →conv_param)[0], x, dout)
      dw num = eval numerical gradient array(lambda w: conv_relu_forward(x, w, b,__
      →conv_param)[0], w, dout)
      db_num = eval_numerical_gradient_array(lambda b: conv_relu_forward(x, w, b,_
      →conv param)[0], b, dout)
      print('Testing conv_relu_forward:')
      print('dx error: ', rel_error(dx_num, dx))
      print('dw error: ', rel_error(dw_num, dw))
      print('db error: ', rel_error(db_num, db))
     Testing conv_relu_forward:
     dx error: 4.518922954146391e-09
     dw error: 9.909016154816753e-09
     db error: 4.308314302730245e-11
[78]: from cs231n.layer_utils import affine_relu_forward, affine_relu_backward
      x = np.random.randn(2, 3, 4)
      w = np.random.randn(12, 10)
      b = np.random.randn(10)
      dout = np.random.randn(2, 10)
      out, cache = affine_relu_forward(x, w, b)
      dx, dw, db = affine_relu_backward(dout, cache)
      dx num = eval_numerical_gradient_array(lambda x: affine relu_forward(x, w, u
      \rightarrowb)[0], x, dout)
      dw_num = eval_numerical_gradient_array(lambda w: affine_relu_forward(x, w,_
      \rightarrowb)[0], w, dout)
      db num = eval_numerical_gradient_array(lambda b: affine relu_forward(x, w, u
      \rightarrowb)[0], b, dout)
      print('Testing affine_relu_forward:')
      print('dx error: ', rel_error(dx_num, dx))
```

```
print('dw error: ', rel_error(dw_num, dw))
print('db error: ', rel_error(db_num, db))

Testing affine_relu_forward:
   dx error: 1.4000743724844122e-10
   dw error: 2.004992303924378e-10
   db error: 7.434255696947127e-10
```