Convolutional neural networks

In this notebook, we'll put together our convolutional layers to implement a 3-layer CNN. Then, we'll ask you to implement a CNN that can achieve > 65% validation error on CIFAR-10.

CS231n has built a solid API for building these modular frameworks and training them, and we will use their very well implemented framework as opposed to "reinventing the wheel." This includes using their Solver, various utility functions, their layer structure, and their implementation of fast CNN layers. This also includes nndl.fc_net, nndl.layers, and nndl.layer_utils. As in prior assignments, we thank Serena Yeung & Justin Johnson for permission to use code written for the CS 231n class (cs231n.stanford.edu).

If you have not completed the Spatial BatchNorm Notebook, please see the following description from that notebook:

Please copy and paste your prior implemented code from HW #4 to start this assignment. If you did not correctly implement the layers in HW #4, you may collaborate with a classmate to use their layer implementations from HW #4. You may also visit TA or Prof OH to correct your implementation.

You'll want to copy and paste from HW #4:

```
- layers.py for your FC network layers, as well as batchnorm and dropout.
- layer_utils.py for your combined FC network layers.
- optim.py for your optimizers.
```

Be sure to place these in the `nndl/` directory so they're imported correctly. Note, as announced in class, we will not be releasing our solutions.

In [1]:
```
# As usual, a bit of setup

import numpy as np
import matplotlib.pyplot as plt
from nndl.cnn import *
from cs231n.data_utils import get_CIFAR10_data
from cs231n.gradient_check import eval_numerical_gradient_array, eval_numerica
l_gradient
from nndl.layers import *
from nndl.conv_layers import *
from cs231n.fast_layers import *
from cs231n.solver import Solver

%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0) # set default size of plots
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

# for auto-reloading external modules
# see http://stackoverflow.com/questions/1907993/autoreload-of-modules-in-ipyt
hon
%load_ext autoreload
%autoreload 2

def rel_error(x, y):
  """ returns relative error """
  return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y))))
```

```
<cyfunction col2im_6d_cython at 0x000002321257CC88>
```

In [2]:
```
# Load the (preprocessed) CIFAR10 data.

data = get_CIFAR10_data()
for k in data.keys():
  print('{}: {} '.format(k, data[k].shape))
```

```
X_train: (49000, 3, 32, 32)
y_train: (49000,)
X_val: (1000, 3, 32, 32)
y_val: (1000,)
X_test: (1000, 3, 32, 32)
y_test: (1000,)
```

Three layer CNN

In this notebook, you will implement a three layer CNN. The `ThreeLayerConvNet` class is in `nndl/cnn.py`. You'll need to modify that code for this section, including the initialization, as well as the calculation of the loss and gradients. You should be able to use the building blocks you have either earlier coded or that we have provided. Be sure to use the fast layers.

The architecture of this CNN will be:

conv - relu - 2x2 max pool - affine - relu - affine - softmax

We won't use batchnorm yet. You've also done enough of these to know how to debug; use the cells below.

Note: As we are implementing several layers CNN networks. The gradient error can be expected for the `eval_numerical_gradient()` function. If your `W1 max relative error` and `W2 max relative error` are around or below 0.01, they should be acceptable. Other errors should be less than 1e-5.

In [6]:
```
num_inputs = 2
input_dim = (3, 16, 16)
reg = 0.0
num_classes = 10
X = np.random.randn(num_inputs, *input_dim)
y = np.random.randint(num_classes, size=num_inputs)

model = ThreeLayerConvNet(num_filters=3, filter_size=3,
                          input_dim=input_dim, hidden_dim=7,
                          dtype=np.float64)
loss, grads = model.loss(X, y)
for param_name in sorted(grads):
    f = lambda _: model.loss(X, y)[0]
    param_grad_num = eval_numerical_gradient(f, model.params[param_name], verb
ose=False, h=1e-6)
    e = rel_error(param_grad_num, grads[param_name])
    print('{} max relative error: {}'.format(param_name, rel_error(param_grad_
num, grads[param_name])))
```

```
W1 max relative error: 0.00022102674273321275
W2 max relative error: 0.023755654587671298
W3 max relative error: 4.1922369320789225e-05
b1 max relative error: 1.1632528272375571e-05
b2 max relative error: 1.435885554342043e-06
b3 max relative error: 8.709922064435798e-10
```

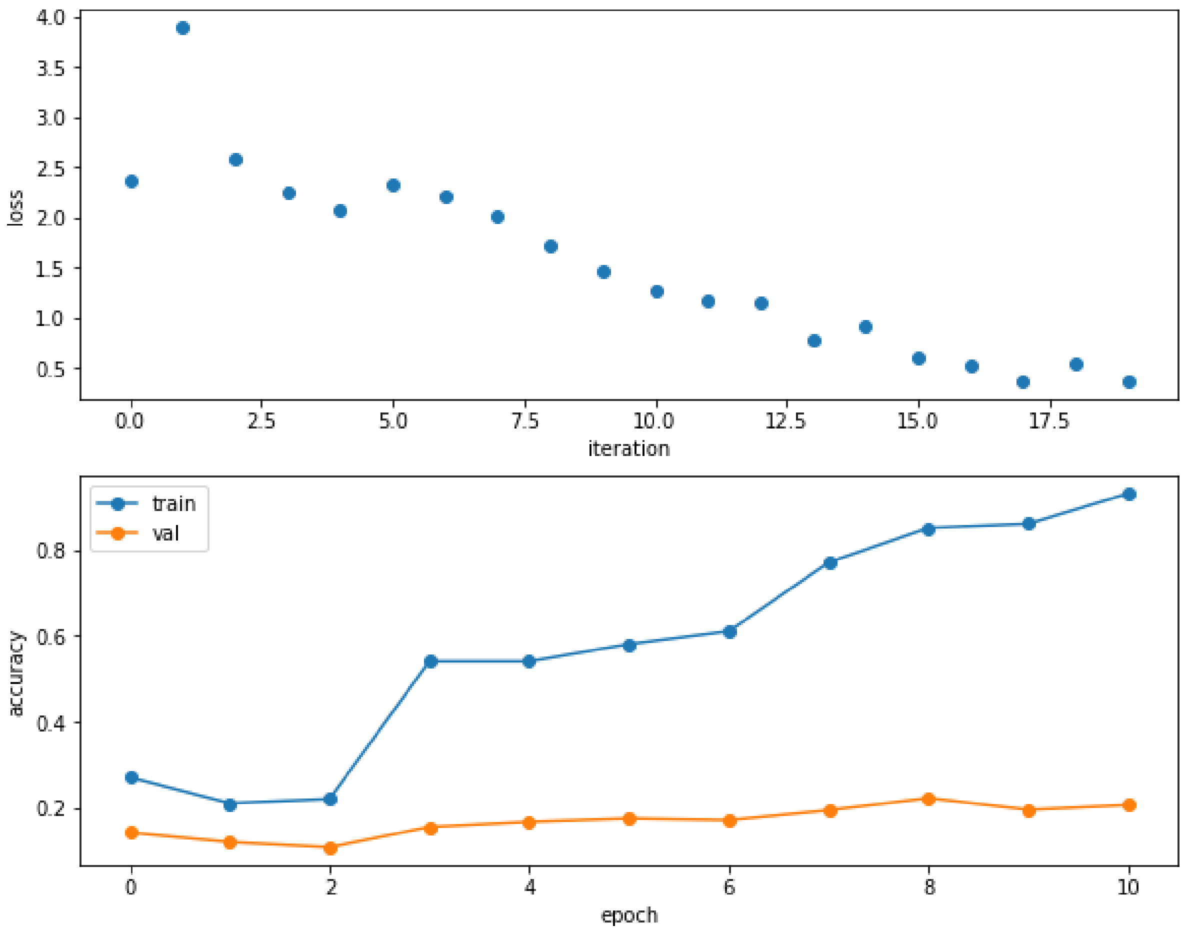
Overfit small dataset

To check your CNN implementation, let's overfit a small dataset.

In [7]:
```
num_train = 100
small_data = {
  'X_train': data['X_train'][:num_train],
  'y_train': data['y_train'][:num_train],
  'X_val': data['X_val'],
  'y_val': data['y_val'],
}

model = ThreeLayerConvNet(weight_scale=1e-2)

solver = Solver(model, small_data,
                num_epochs=10, batch_size=50,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                verbose=True, print_every=1)
solver.train()
```

```
(Iteration 1 / 20) loss: 2.364586
(Epoch 0 / 10) train acc: 0.270000; val_acc: 0.142000
(Iteration 2 / 20) loss: 3.897309
(Epoch 1 / 10) train acc: 0.210000; val_acc: 0.120000
(Iteration 3 / 20) loss: 2.587546
(Iteration 4 / 20) loss: 2.250113
(Epoch 2 / 10) train acc: 0.220000; val_acc: 0.108000
(Iteration 5 / 20) loss: 2.080509
(Iteration 6 / 20) loss: 2.329377
(Epoch 3 / 10) train acc: 0.540000; val_acc: 0.155000
(Iteration 7 / 20) loss: 2.213046
(Iteration 8 / 20) loss: 2.013081
(Epoch 4 / 10) train acc: 0.540000; val_acc: 0.167000
(Iteration 9 / 20) loss: 1.720941
(Iteration 10 / 20) loss: 1.472394
(Epoch 5 / 10) train acc: 0.580000; val_acc: 0.175000
(Iteration 11 / 20) loss: 1.269525
(Iteration 12 / 20) loss: 1.174309
(Epoch 6 / 10) train acc: 0.610000; val_acc: 0.171000
(Iteration 13 / 20) loss: 1.155775
(Iteration 14 / 20) loss: 0.777094
(Epoch 7 / 10) train acc: 0.770000; val_acc: 0.194000
(Iteration 15 / 20) loss: 0.910162
(Iteration 16 / 20) loss: 0.608730
(Epoch 8 / 10) train acc: 0.850000; val_acc: 0.222000
(Iteration 17 / 20) loss: 0.512748
(Iteration 18 / 20) loss: 0.372021
(Epoch 9 / 10) train acc: 0.860000; val_acc: 0.195000
(Iteration 19 / 20) loss: 0.549441
(Iteration 20 / 20) loss: 0.372587
(Epoch 10 / 10) train acc: 0.930000; val_acc: 0.206000
```

In [8]:
```
plt.subplot(2, 1, 1)
plt.plot(solver.loss_history, 'o')
plt.xlabel('iteration')
plt.ylabel('loss')

plt.subplot(2, 1, 2)
plt.plot(solver.train_acc_history, '-o')
plt.plot(solver.val_acc_history, '-o')
plt.legend(['train', 'val'], loc='upper left')
plt.xlabel('epoch')
plt.ylabel('accuracy')
plt.show()
```



Train the network

Now we train the 3 layer CNN on CIFAR-10 and assess its accuracy.

In [9]:
```
model = ThreeLayerConvNet(weight_scale=0.001, hidden_dim=500, reg=0.001)

solver = Solver(model, data,
                num_epochs=1, batch_size=50,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                verbose=True, print_every=20)
solver.train()
```

```
(Iteration 1 / 980) loss: 2.304630
(Epoch 0 / 1) train acc: 0.117000; val_acc: 0.097000
(Iteration 21 / 980) loss: 2.107427
(Iteration 41 / 980) loss: 2.299490
(Iteration 61 / 980) loss: 2.037402
(Iteration 81 / 980) loss: 1.824624
(Iteration 101 / 980) loss: 1.956848
(Iteration 121 / 980) loss: 1.776579
(Iteration 141 / 980) loss: 1.600286
(Iteration 161 / 980) loss: 1.813623
(Iteration 181 / 980) loss: 1.583842
(Iteration 201 / 980) loss: 2.033325
(Iteration 221 / 980) loss: 1.765937
(Iteration 241 / 980) loss: 1.862508
(Iteration 261 / 980) loss: 1.589407
(Iteration 281 / 980) loss: 1.725169
(Iteration 301 / 980) loss: 1.547960
(Iteration 321 / 980) loss: 1.591477
(Iteration 341 / 980) loss: 1.591618
(Iteration 361 / 980) loss: 1.822010
(Iteration 381 / 980) loss: 1.709555
(Iteration 401 / 980) loss: 1.909300
(Iteration 421 / 980) loss: 1.558605
(Iteration 441 / 980) loss: 1.511913
(Iteration 461 / 980) loss: 1.424503
(Iteration 481 / 980) loss: 1.598751
(Iteration 501 / 980) loss: 1.647390
(Iteration 521 / 980) loss: 1.509191
(Iteration 541 / 980) loss: 1.554988
(Iteration 561 / 980) loss: 1.480685
(Iteration 581 / 980) loss: 1.563901
(Iteration 601 / 980) loss: 1.657477
(Iteration 621 / 980) loss: 1.744823
(Iteration 641 / 980) loss: 1.395340
(Iteration 661 / 980) loss: 1.693074
(Iteration 681 / 980) loss: 1.778899
(Iteration 701 / 980) loss: 1.417933
(Iteration 721 / 980) loss: 1.754427
(Iteration 741 / 980) loss: 1.535386
(Iteration 761 / 980) loss: 1.620074
(Iteration 781 / 980) loss: 1.245937
(Iteration 801 / 980) loss: 1.304316
(Iteration 821 / 980) loss: 1.728228
(Iteration 841 / 980) loss: 1.761504
(Iteration 861 / 980) loss: 1.438008
(Iteration 881 / 980) loss: 1.894736
(Iteration 901 / 980) loss: 1.272871
(Iteration 921 / 980) loss: 1.439679
(Iteration 941 / 980) loss: 1.825545
(Iteration 961 / 980) loss: 1.360799
(Epoch 1 / 1) train acc: 0.441000; val_acc: 0.466000
```

Get > 65% validation accuracy on CIFAR-10.

In the last part of the assignment, we'll now ask you to train a CNN to get better than 65% validation accuracy on CIFAR-10.

Things you should try:

- Filter size: Above we used 7x7; but VGGNet and onwards showed stacks of 3x3 filters are good.
- Number of filters: Above we used 32 filters. Do more or fewer do better?
- Batch normalization: Try adding spatial batch normalization after convolution layers and vanilla batch normalization aafter affine layers. Do your networks train faster?
- Network architecture: Can a deeper CNN do better? Consider these architectures:
  - [conv-relu-pool]xN - conv - relu - [affine]xM - [softmax or SVM]
  - [conv-relu-pool]XN - [affine]XM - [softmax or SVM]
  - [conv-relu-conv-relu-pool]xN - [affine]xM - [softmax or SVM]

Tips for training

For each network architecture that you try, you should tune the learning rate and regularization strength. When doing this there are a couple important things to keep in mind:

- If the parameters are working well, you should see improvement within a few hundred iterations
- Remember the coarse-to-fine approach for hyperparameter tuning: start by testing a large range of hyperparameters for just a few training iterations to find the combinations of parameters that are working at all.
- Once you have found some sets of parameters that seem to work, search more finely around these parameters. You may need to train for more epochs.

In [12]:
```
# ================================================================ #
# YOUR CODE HERE:
#   Implement a CNN to achieve greater than 65% validation accuracy
#   on CIFAR-10.
# ================================================================ #
model = ThreeLayerConvNet(weight_scale=0.001, hidden_dim=500, reg=0.001)

solver = Solver(model, data,
                num_epochs=10, batch_size=1000,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                verbose=True, print_every=20)
solver.train()

# ================================================================ #
# END YOUR CODE HERE
# ================================================================ #
```

```
(Iteration 1 / 490) loss: 2.304645
(Epoch 0 / 10) train acc: 0.098000; val_acc: 0.121000
(Iteration 21 / 490) loss: 1.812656
(Iteration 41 / 490) loss: 1.545035
(Epoch 1 / 10) train acc: 0.493000; val_acc: 0.492000
(Iteration 61 / 490) loss: 1.468290
(Iteration 81 / 490) loss: 1.344017
(Epoch 2 / 10) train acc: 0.567000; val_acc: 0.553000
(Iteration 101 / 490) loss: 1.289569
(Iteration 121 / 490) loss: 1.179220
(Iteration 141 / 490) loss: 1.082809
(Epoch 3 / 10) train acc: 0.593000; val_acc: 0.592000
(Iteration 161 / 490) loss: 1.056155
(Iteration 181 / 490) loss: 1.037196
(Epoch 4 / 10) train acc: 0.688000; val_acc: 0.623000
(Iteration 201 / 490) loss: 1.016461
(Iteration 221 / 490) loss: 0.972576
(Iteration 241 / 490) loss: 0.954136
(Epoch 5 / 10) train acc: 0.703000; val_acc: 0.633000
(Iteration 261 / 490) loss: 0.845294
(Iteration 281 / 490) loss: 0.928544
(Epoch 6 / 10) train acc: 0.705000; val_acc: 0.623000
(Iteration 301 / 490) loss: 0.866793
(Iteration 321 / 490) loss: 0.864300
(Iteration 341 / 490) loss: 0.875562
(Epoch 7 / 10) train acc: 0.744000; val_acc: 0.655000
(Iteration 361 / 490) loss: 0.816512
(Iteration 381 / 490) loss: 0.738668
(Epoch 8 / 10) train acc: 0.780000; val_acc: 0.652000
(Iteration 401 / 490) loss: 0.691402
(Iteration 421 / 490) loss: 0.708379
(Iteration 441 / 490) loss: 0.685015
(Epoch 9 / 10) train acc: 0.763000; val_acc: 0.632000
(Iteration 461 / 490) loss: 0.768887
(Iteration 481 / 490) loss: 0.698631
(Epoch 10 / 10) train acc: 0.798000; val_acc: 0.652000
```

In [13]:
```
model.params = solver.best_params
```

In [14]:
```
y_test_pred = np.argmax(model.loss(data['X_test']), axis=1)
y_val_pred = np.argmax(model.loss(data['X_val']), axis=1)
print('Validation set accuracy: {}'.format(np.mean(y_val_pred == data['y_val'
])))
print('Test set accuracy: {}'.format(np.mean(y_test_pred == data['y_test'])))
```

```
Validation set accuracy: 0.655
Test set accuracy: 0.649
```

cnn.py

In [ ]:
```
import numpy as np

from nndl.layers import *
from nndl.conv_layers import *
from cs231n.fast_layers import *
from nndl.layer_utils import *
from nndl.conv_layer_utils import *

import pdb

"""
This code was originally written for CS 231n at Stanford University
(cs231n.stanford.edu).  It has been modified in various areas for use in the
ECE 239AS class at UCLA.  This includes the descriptions of what code to
implement as well as some slight potential changes in variable names to be
consistent with class nomenclature.  We thank Justin Johnson & Serena Yeung fo
r
permission to use this code.  To see the original version, please visit
cs231n.stanford.edu.
"""

class ThreeLayerConvNet(object):
  """
  A three-layer convolutional network with the following architecture:

  conv - relu - 2x2 max pool - affine - relu - affine - softmax

  The network operates on minibatches of data that have shape (N, C, H, W)
  consisting of N images, each with height H and width W and with C input
  channels.
  """

  def __init__(self, input_dim=(3, 32, 32), num_filters=32, filter_size=7,
               hidden_dim=100, num_classes=10, weight_scale=1e-3, reg=0.0,
               dtype=np.float32, use_batchnorm=False):
    """
    Initialize a new network.

    Inputs:
    - input_dim: Tuple (C, H, W) giving size of input data
    - num_filters: Number of filters to use in the convolutional layer
    - filter_size: Size of filters to use in the convolutional layer
    - hidden_dim: Number of units to use in the fully-connected hidden layer
    - num_classes: Number of scores to produce from the final affine layer.
    - weight_scale: Scalar giving standard deviation for random initialization
      of weights.
    - reg: Scalar giving L2 regularization strength
    - dtype: numpy datatype to use for computation.
    """
    self.use_batchnorm = use_batchnorm
    self.params = {}
    self.reg = reg
    self.dtype = dtype


    # ================================================================ #
```

```
    # YOUR CODE HERE:
    #   Initialize the weights and biases of a three layer CNN. To initialize:
    #     - the biases should be initialized to zeros.
    #     - the weights should be initialized to a matrix with entries
    #         drawn from a Gaussian distribution with zero mean and
    #         standard deviation given by weight_scale.
    # ================================================================ #
    self.params["W1"] = weight_scale * np.random.randn(num_filters,3,filter_si
ze,filter_size)
    self.params["b1"] = np.zeros(num_filters,)


    flattened_size = (1 + (input_dim[1]-2)/2)

    self.params["W2"] = weight_scale * np.random.randn(int(flattened_size**2 *
num_filters),hidden_dim)
    self.params["b2"] = np.zeros(hidden_dim,)

    self.params["W3"] =  weight_scale * np.random.randn(hidden_dim,num_classes
)
    self.params["b3"] = np.zeros(num_classes,)

    # ================================================================ #
    # END YOUR CODE HERE
    # ================================================================ #

    for k, v in self.params.items():
      self.params[k] = v.astype(dtype)


  def loss(self, X, y=None):
    """
    Evaluate loss and gradient for the three-layer convolutional network.

    Input / output: Same API as TwoLayerNet in fc_net.py.
    """
    W1, b1 = self.params['W1'], self.params['b1']
    W2, b2 = self.params['W2'], self.params['b2']
    W3, b3 = self.params['W3'], self.params['b3']

    # pass conv_param to the forward pass for the convolutional layer
    filter_size = W1.shape[2]
    conv_param = {'stride': 1, 'pad': (filter_size - 1) / 2}

    # pass pool_param to the forward pass for the max-pooling layer
    pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}

    scores = None

    # ================================================================ #
    # YOUR CODE HERE:
    #   Implement the forward pass of the three layer CNN.  Store the output
    #   scores as the variable "scores".
    # ================================================================ #
    scores, cache_l1 = conv_relu_pool_forward(X, W1, b1, conv_param, pool_para
m)
    output_shape = scores.shape
```

```
    scores, cache_l2 = affine_relu_forward(scores, W2, b2)
    scores, cache_l3 = affine_forward(scores, W3, b3)

    #scores = np.exp(scores-np.max(scores,1)[:,np.newaxis])/np.sum(np.exp(scor
es-np.max(scores,1)[:,np.newaxis]),1)[:,np.newaxis]

    # ================================================================ #
    # END YOUR CODE HERE
    # ================================================================ #

    if y is None:
      return scores

    loss, grads = 0, {}
    # ================================================================ #
    # YOUR CODE HERE:
    #   Implement the backward pass of the three layer CNN.  Store the grads
    #   in the grads dictionary, exactly as before (i.e., the gradient of
    #   self.params[k] will be grads[k]).  Store the loss as "loss", and
    #   don't forget to add regularization on ALL weight matrices.
    # ================================================================ #
    soft_loss, dx = softmax_loss(scores, y)
    reg_loss = 0.5 * (np.linalg.norm(W1)**2 + np.linalg.norm(W2)**2 + np.linal
g.norm(W3)**2)
    loss = soft_loss + self.reg*reg_loss

    dx, dw, db = affine_backward(dx, cache_l3)
    grads['W3'] = dw + 0.5*self.reg*2*W3
    grads['b3'] = db

    dx, dw, db = affine_relu_backward(dx, cache_l2)
    grads['W2'] = dw + 0.5*self.reg*2*W2
    grads['b2']= db

    dx = dx.reshape(*output_shape)

    _, dw, db = conv_relu_pool_backward(dx, cache_l1)
    grads['W1'] = dw + 0.5*self.reg*2*W1
    grads['b1']= db
    # ================================================================ #
    # END YOUR CODE HERE
    # ================================================================ #

    return loss, grads
```