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.

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 [106]:

```
# As usual, a bit of setup

import numpy as np
import matplotlib.pyplot as plt
from nndl.cnn import *
from utils.data_utils import get_CIFAR10_data
from utils.gradient_check import eval_numerical_gradient_array, eval_numerical_gradient
from nndl.layers import *
from nndl.conv_layers import *
from utils.fast_layers import *
from utils.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-ipython
%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))))
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

In [107]:

```
# 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 [109]:

```
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], verbose=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: 0.011821375430781197
W2 max relative error: 0.0016865050976615941
W3 max relative error: 1.1990575927951157e-05
b1 max relative error: 9.427363623190101e-06
b2 max relative error: 2.821574180080946e-08
b3 max relative error: 8.538894241102408e-10
```

Overfit small dataset

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

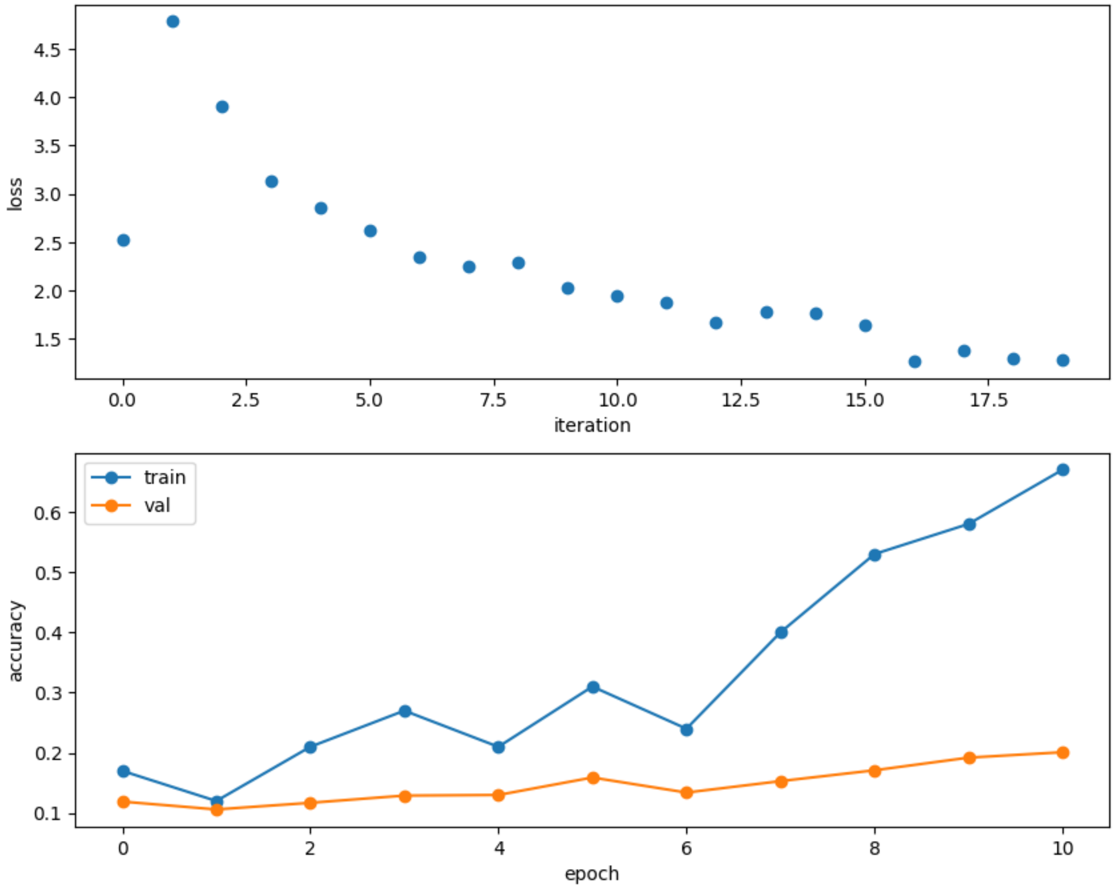
In [110]:

```
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: 2.524423
(Epoch 0 / 10) train acc: 0.170000; val_acc: 0.119000
(Iteration 2 / 20) loss: 4.781698
(Epoch 1 / 10) train acc: 0.120000; val_acc: 0.106000
(Iteration 3 / 20) loss: 3.909344
(Iteration 4 / 20) loss: 3.136751
(Epoch 2 / 10) train acc: 0.210000; val_acc: 0.117000
(Iteration 5 / 20) loss: 2.850169
(Iteration 6 / 20) loss: 2.625665
(Epoch 3 / 10) train acc: 0.270000; val_acc: 0.129000
(Iteration 7 / 20) loss: 2.344947
(Iteration 8 / 20) loss: 2.254606
(Epoch 4 / 10) train acc: 0.210000; val_acc: 0.130000
(Iteration 9 / 20) loss: 2.285453
(Iteration 10 / 20) loss: 2.030125
(Epoch 5 / 10) train acc: 0.310000; val_acc: 0.159000
(Iteration 11 / 20) loss: 1.947709
(Iteration 12 / 20) loss: 1.879866
(Epoch 6 / 10) train acc: 0.240000; val_acc: 0.134000
(Iteration 13 / 20) loss: 1.668075
(Iteration 14 / 20) loss: 1.783985
(Epoch 7 / 10) train acc: 0.400000; val_acc: 0.153000
(Iteration 15 / 20) loss: 1.767605
(Iteration 16 / 20) loss: 1.644005
(Epoch 8 / 10) train acc: 0.530000; val_acc: 0.171000
(Iteration 17 / 20) loss: 1.268453
(Iteration 18 / 20) loss: 1.385116
(Epoch 9 / 10) train acc: 0.580000; val_acc: 0.192000
(Iteration 19 / 20) loss: 1.297587
(Iteration 20 / 20) loss: 1.282454
(Epoch 10 / 10) train acc: 0.670000; val_acc: 0.201000
```

In [111]:

```
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 [112]:

```
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: 2.304379
(Epoch 0 / 1) train acc: 0.088000; val_acc: 0.119000
(Iteration 21 / 980) loss: 2.109635
(Iteration 41 / 980) loss: 2.076350
(Iteration 61 / 980) loss: 2.016401
(Iteration 81 / 980) loss: 2.114350
(Iteration 101 / 980) loss: 1.780723
(Iteration 121 / 980) loss: 1.679214
(Iteration 141 / 980) loss: 1.846902
(Iteration 161 / 980) loss: 1.686197
(Iteration 181 / 980) loss: 1.849539
(Iteration 201 / 980) loss: 1.848983
(Iteration 221 / 980) loss: 1.607722
(Iteration 241 / 980) loss: 1.408573
(Iteration 261 / 980) loss: 1.667033
(Iteration 281 / 980) loss: 1.774849
(Iteration 301 / 980) loss: 1.994323
(Iteration 321 / 980) loss: 1.878220
(Iteration 341 / 980) loss: 1.884071
(Iteration 361 / 980) loss: 1.707486
(Iteration 381 / 980) loss: 1.483961
(Iteration 401 / 980) loss: 1.907075
(Iteration 421 / 980) loss: 1.763213
(Iteration 441 / 980) loss: 1.760534
(Iteration 461 / 980) loss: 1.695799
(Iteration 481 / 980) loss: 1.714589
(Iteration 501 / 980) loss: 1.912646
(Iteration 521 / 980) loss: 1.412743
(Iteration 541 / 980) loss: 1.697497
(Iteration 561 / 980) loss: 1.799456
(Iteration 581 / 980) loss: 1.762910
(Iteration 601 / 980) loss: 1.662797
(Iteration 621 / 980) loss: 1.479918
(Iteration 641 / 980) loss: 1.925249
(Iteration 661 / 980) loss: 1.667685
(Iteration 681 / 980) loss: 1.878623
(Iteration 701 / 980) loss: 1.685491
(Iteration 721 / 980) loss: 1.682482
(Iteration 741 / 980) loss: 1.718161
(Iteration 761 / 980) loss: 1.618900
(Iteration 781 / 980) loss: 1.576266
(Iteration 801 / 980) loss: 1.567092
(Iteration 821 / 980) loss: 1.634341
(Iteration 841 / 980) loss: 1.870606
(Iteration 861 / 980) loss: 1.857803
(Iteration 881 / 980) loss: 1.689935
(Iteration 901 / 980) loss: 1.449273
(Iteration 921 / 980) loss: 1.558852
(Iteration 941 / 980) loss: 1.357863
(Iteration 961 / 980) loss: 1.314725
(Epoch 1 / 1) train acc: 0.474000; val_acc: 0.445000
```

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 [115]:

```
# ================================================================ #
# YOUR CODE HERE:
#   Implement a CNN to achieve greater than 65% validation accuracy
#   on CIFAR-10.
# ================================================================ #

#I use 64 3 * 3 filters, with hidden_dim = 400
model = ThreeLayerConvNet(num_filters=64, filter_size=3, weight_scale=0.001, hidden_dim=400, reg=0.001)

#increase the num_epochs and batch_size while adding lr_decay
solver = Solver(model, data,
                num_epochs=8, batch_size=600,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                lr_decay = 0.95,
                verbose=True, print_every=20)
solver.train()

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

```
(Iteration 1 / 648) loss: 2.305887
(Epoch 0 / 8) train acc: 0.096000; val_acc: 0.119000
(Iteration 21 / 648) loss: 1.776225
(Iteration 41 / 648) loss: 1.600341
(Iteration 61 / 648) loss: 1.443222
(Iteration 81 / 648) loss: 1.362421
(Epoch 1 / 8) train acc: 0.549000; val_acc: 0.557000
(Iteration 101 / 648) loss: 1.403169
(Iteration 121 / 648) loss: 1.212663
(Iteration 141 / 648) loss: 1.229137
(Iteration 161 / 648) loss: 1.167617
(Epoch 2 / 8) train acc: 0.626000; val_acc: 0.593000
(Iteration 181 / 648) loss: 1.214655
(Iteration 201 / 648) loss: 1.087428
(Iteration 221 / 648) loss: 1.097019
(Iteration 241 / 648) loss: 1.077913
(Epoch 3 / 8) train acc: 0.648000; val_acc: 0.606000
(Iteration 261 / 648) loss: 1.013566
(Iteration 281 / 648) loss: 0.987600
(Iteration 301 / 648) loss: 0.998555
(Iteration 321 / 648) loss: 1.009989
(Epoch 4 / 8) train acc: 0.700000; val_acc: 0.617000
(Iteration 341 / 648) loss: 0.916954
(Iteration 361 / 648) loss: 0.952466
(Iteration 381 / 648) loss: 0.823338
(Iteration 401 / 648) loss: 0.889315
(Epoch 5 / 8) train acc: 0.757000; val_acc: 0.635000
(Iteration 421 / 648) loss: 0.924732
(Iteration 441 / 648) loss: 0.762855
(Iteration 461 / 648) loss: 0.797278
(Iteration 481 / 648) loss: 0.810951
(Epoch 6 / 8) train acc: 0.775000; val_acc: 0.638000
(Iteration 501 / 648) loss: 0.771332
(Iteration 521 / 648) loss: 0.739867
(Iteration 541 / 648) loss: 0.750929
(Iteration 561 / 648) loss: 0.763980
(Epoch 7 / 8) train acc: 0.775000; val_acc: 0.651000
(Iteration 581 / 648) loss: 0.635425
(Iteration 601 / 648) loss: 0.714568
(Iteration 621 / 648) loss: 0.685466
(Iteration 641 / 648) loss: 0.784730
(Epoch 8 / 8) train acc: 0.818000; val_acc: 0.652000
```

In [ ]:

```
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.
    # ================================================================ #

    C, H, W = input_dim #get the input dimensions

    #initialize the weights and biases for three layers

    #conv - relu - 2x2 max pool
    self.params['W1'] = np.random.normal(0, scale = weight_scale, size = (num_filters, C, filter_size, filter_size))
    self.params['b1'] = np.zeros(num_filters)

    #affine - relu
    self.params['W2'] = np.random.normal(0, scale = weight_scale, size = (num_filters * (H // 2) * (W // 2), hidden_dim))
    self.params['b2'] = np.zeros(hidden_dim)

    #affine - softmax
    self.params['W3'] = np.random.normal(0, scale = weight_scale, size = (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".
```

```python
    # ============================================================= #

    #conv - relu - 2x2 max pool
    convout, convcache = conv_relu_pool_forward(X, W1, b1, conv_param, pool_param)

    #affine - relu
    arout, arcache = affine_relu_forward(convout, W2, b2)

    #affine - softmax
    scores, ascache = affine_forward(arout, W3, b3)
    # ============================================================= #
    # 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.
    # ============================================================= #

    #calculate the softmax loss
    loss, dscores = softmax_loss(scores, y)

    #update the loss based on the regularization
    loss += 0.5 * self.reg * (np.sum(W1 ** 2) + np.sum(W2 ** 2) + np.sum(W3 ** 2))

    #affine backward
    dab, grads['W3'], grads['b3'] = affine_backward(dscores, ascache)

    #affine-relu backward
    dconv, grads['W2'], grads['b2'] = affine_relu_backward(dab, arcache)

    #conv-relu-pool backward
    dx, grads['W1'], grads['b1'] = conv_relu_pool_backward(dconv, convcache)

    #update the gradients of weights
    grads['W1'] += grads['W1'] * self.reg
    grads['W2'] += grads['W2'] * self.reg
    grads['W3'] += grads['W3'] * self.reg

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

    return loss, grads
```

In [ ]: