In [ ]:
```python
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 for
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'] = np.random.normal(0, weight_scale, [num_filters, input_dim
[0], filter_size, filter_size])
    self.params['b1'] = np.zeros(num_filters)
    W1_lenx = int((input_dim[1] - 2) / 2) + 1 # b/c pad is set such that conv lay
er doesn't shrink it, but there is a pool
    W1_leny = int((input_dim[2] - 2) / 2) + 1
    self.params['W2'] = np.random.normal(0, weight_scale, [W1_lenx * W1_leny * nu
m_filters, hidden_dim])
    self.params['b2'] = np.zeros(hidden_dim)
    self.params['W3'] = np.random.normal(0, weight_scale, [hidden_dim, num_classe
s])
    self.params['b3'] = np.zeros(num_classes)

    if self.use_batchnorm:
        self.bn_params = [{'mode': 'train', 'eps': 1e-5, 'momentum': 0.9} for i i
n np.arange(self.num_layers - 1)]
        self.params['gamma1'] = np.ones(input_dim[0])
        self.params['beta1'] = np.zeros(input_dim[0])
        self.params['gamma2'] = np.ones()
        self.params['beta2'] = np.zeros()
    # ================================================================ #
    # 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".
    # ================================================================ #



    if not self.use_batchnorm:
        # conv - relu - 2x2 max pool - affine - relu - affine - softmax
        p1, p1_cache = conv_relu_pool_forward(X, W1, b1, conv_param, pool_param)
# a1 -> h1 -> p1
        h2, h2_cache = affine_relu_forward(p1, W2, b2) # p1 -> a2 -> h2
        scores, a3_cache = affine_forward(h2, W3, b3)
```

```
    else:
        # conv - sbn - relu - 2x2 max pool - affine - bn - relu - affine - softma
x
        a, conv_cache = conv_forward_fast(x, w, b, conv_param)
        out, cache = spatial_batchnorm_forward(x, gamma, beta, bn_param)
        s, relu_cache = relu_forward(a)
        out, pool_cache = max_pool_forward_fast(s, pool_param)
        out, cache = affine_forward(x, w, b)
        out, cache = batchnorm_forward(x, gamma, beta, bn_param)
        s, relu_cache = relu_forward(a)
        out, cache = affine_forward(x, w, b)
    # ================================================================ #
    # 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.
    # ================================================================ #
    loss, dx = softmax_loss(scores, y)
    loss += 0.5 * self.reg * np.sum([np.sum(self.params['W{}'.format(layer + 1)]
** 2) for layer in range(3)])

    if not self.use_batchnorm:
        dl_dh2, grads['W3'], grads['b3'] = affine_backward(dx, a3_cache)
        grads['W3'] += self.reg * self.params['W3']
        dl_dp1, grads['W2'], grads['b2'] = affine_relu_backward(dl_dh2, h2_cache)
        grads['W2'] += self.reg * self.params['W2']
        dl_dx, grads['W1'], grads['b1'] = conv_relu_pool_backward(dl_dp1, p1_cach
e)
        grads['W1'] += self.reg * self.params['W1']
    else:
        pass
    # ================================================================ #
    # END YOUR CODE HERE
    # ================================================================ #

    return loss, grads


pass
```