This is the 2-layer neural network notebook for ECE C147/C247 Homework #3

Please follow the notebook linearly to implement a two layer neural network.

Please print out the notebook entirely when completed.

The goal of this notebook is to give you experience with training a two layer neural network.

In [134…
```python
import random
import numpy as np
from utils.data_utils import load_CIFAR10
import matplotlib.pyplot as plt

%matplotlib inline
%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
```

Toy example

Before loading CIFAR-10, there will be a toy example to test your implementation of the forward and backward pass

In [135…
```python
from nndl.neural_net import TwoLayerNet
```

In [136…
```python
# Create a small net and some toy data to check your implementations.
# Note that we set the random seed for repeatable experiments.

input_size = 4
hidden_size = 10
num_classes = 3
num_inputs = 5

def init_toy_model():
    np.random.seed(0)
    return TwoLayerNet(input_size, hidden_size, num_classes, std=1e-1)

def init_toy_data():
    np.random.seed(1)
    X = 10 * np.random.randn(num_inputs, input_size)
    y = np.array([0, 1, 2, 2, 1])
    return X, y

net = init_toy_model()
X, y = init_toy_data()
```

Compute forward pass scores

In [137…
```python
## Implement the forward pass of the neural network.

# Note, there is a statement if y is None: return scores, which is why
# the following call will calculate the scores.
scores = net.loss(X)
print('Your scores:')
print(scores)
print()
print('correct scores:')
correct_scores = np.asarray([
    [-1.07260209,  0.05083871, -0.87253915],
    [-2.02778743, -0.10832494, -1.52641362],
    [-0.74225908,  0.15259725, -0.39578548],
    [-0.38172726,  0.10835902, -0.17328274],
    [-0.64417314, -0.18886813, -0.41106892]])
print(correct_scores)
print()

# The difference should be very small. We get < 1e-7
print('Difference between your scores and correct scores:')
print(np.sum(np.abs(scores - correct_scores)))
```

```
Your scores:
[[-1.07260209  0.05083871 -0.87253915]
 [-2.02778743 -0.10832494 -1.52641362]
 [-0.74225908  0.15259725 -0.39578548]
 [-0.38172726  0.10835902 -0.17328274]
 [-0.64417314 -0.18886813 -0.41106892]]

correct scores:
[[-1.07260209  0.05083871 -0.87253915]
 [-2.02778743 -0.10832494 -1.52641362]
 [-0.74225908  0.15259725 -0.39578548]
 [-0.38172726  0.10835902 -0.17328274]
 [-0.64417314 -0.18886813 -0.41106892]]

Difference between your scores and correct scores:
3.3812311957259755e-08
```

Forward pass loss

In [139…
```python
loss, _ = net.loss(X, y, reg=0.05)
correct_loss = 1.071696123862817

# should be very small, we get < 1e-12
print("Loss:",loss)
print('Difference between your loss and correct loss:')
print(np.sum(np.abs(loss - correct_loss)))
```

```
Loss: 1.071696123862817
Difference between your loss and correct loss:
0.0
```

Backward pass

Implements the backwards pass of the neural network. Check your gradients with the gradient check utilities provided.

In [140…
```python
from utils.gradient_check import eval_numerical_gradient

# Use numeric gradient checking to check your implementation of the backward pass.
# If your implementation is correct, the difference between the numeric and
# analytic gradients should be less than 1e-8 for each of W1, W2, b1, and b2.

loss, grads = net.loss(X, y, reg=0.05)

# these should all be less than 1e-8 or so
for param_name in grads:
    f = lambda W: net.loss(X, y, reg=0.05)[0]
    param_grad_num = eval_numerical_gradient(f, net.params[param_name], verbose=False)
    print('{} max relative error: {}'.format(param_name, rel_error(param_grad_num, grads[param_name])))
```

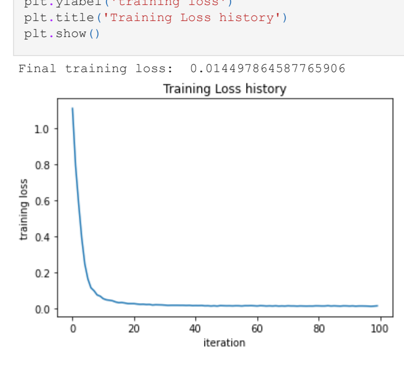
```
W2 max relative error: 2.9632233460136427e-10
b2 max relative error: 1.2482633693659668e-09
W1 max relative error: 1.28328951808708e-09
b1 max relative error: 3.172680285697327e-09
```

Training the network

Implement neural_net.train() to train the network via stochastic gradient descent, much like the softmax.

In [141…
```python
net = init_toy_model()
stats = net.train(X, y, X, y,
            learning_rate=1e-1, reg=5e-6,
            num_iters=100, verbose=False)

print('Final training loss: ', stats['loss_history'][-1])

# plot the loss history
plt.plot(stats['loss_history'])
plt.xlabel('iteration')
plt.ylabel('training loss')
plt.title('Training Loss history')
plt.show()
```

```
Final training loss:  0.014497864587765906
```



Classify CIFAR-10

Do classification on the CIFAR-10 dataset.

In [143…
```python
from utils.data_utils import load_CIFAR10

def get_CIFAR10_data(num_training=49000, num_validation=1000, num_test=1000):
    """
    Load the CIFAR-10 dataset from disk and perform preprocessing to prepare
    it for the two-layer neural net classifier.
    """
    # Load the raw CIFAR-10 data
    cifar10_dir = 'cifar-10-batches-py'
    X_train, y_train, X_test, y_test = load_CIFAR10(cifar10_dir)

    # Subsample the data
    mask = list(range(num_training, num_training + num_validation))
    X_val = X_train[mask]
    y_val = y_train[mask]
    mask = list(range(num_training))
    X_train = X_train[mask]
    y_train = y_train[mask]
    mask = list(range(num_test))
    X_test = X_test[mask]
    y_test = y_test[mask]

    # Normalize the data: subtract the mean image
    mean_image = np.mean(X_train, axis=0)
    X_train -= mean_image
    X_val -= mean_image
    X_test -= mean_image

    # Reshape data to rows
    X_train = X_train.reshape(num_training, -1)
    X_val = X_val.reshape(num_validation, -1)
    X_test = X_test.reshape(num_test, -1)

    return X_train, y_train, X_val, y_val, X_test, y_test


# Invoke the above function to get our data.
X_train, y_train, X_val, y_val, X_test, y_test = get_CIFAR10_data()
print('Train data shape: ', X_train.shape)
print('Train labels shape: ', y_train.shape)
print('Validation data shape: ', X_val.shape)
print('Validation labels shape: ', y_val.shape)
print('Test data shape: ', X_test.shape)
print('Test labels shape: ', y_test.shape)
```

```
Train data shape:  (49000, 3072)
Train labels shape:  (49000,)
Validation data shape:  (1000, 3072)
Validation labels shape:  (1000,)
Test data shape:  (1000, 3072)
Test labels shape:  (1000,)
```

Running SGD

If your implementation is correct, you should see a validation accuracy of around 28-29%.

In [144…
```python
input_size = 32 * 32 * 3
hidden_size = 50
num_classes = 10
net = TwoLayerNet(input_size, hidden_size, num_classes)

# Train the network
stats = net.train(X_train, y_train, X_val, y_val,
            num_iters=1000, batch_size=200,
            learning_rate=1e-4, learning_rate_decay=0.95,
            reg=0.25, verbose=True)

# Predict on the validation set
val_acc = (net.predict(X_val) == y_val).mean()
print('Validation accuracy: ', val_acc)

# Save this net as the variable subopt_net for later comparison.
subopt_net = net
```

```
iteration 0 / 1000: loss 2.302757518613176
iteration 100 / 1000: loss 2.302120159207236
iteration 200 / 1000: loss 2.2956136007408703
iteration 300 / 1000: loss 2.2518259043164135
iteration 400 / 1000: loss 2.188995235046776
iteration 500 / 1000: loss 2.1162527791897747
iteration 600 / 1000: loss 2.064670827698217
iteration 700 / 1000: loss 1.9901688623083942
iteration 800 / 1000: loss 2.002827640124685
iteration 900 / 1000: loss 1.9465176817856495
Validation accuracy:  0.283
```

Questions:

The training accuracy isn't great.

(1) What are some of the reasons why this is the case? Take the following cell to do some analyses and then report your answers in the cell following the one below.

(2) How should you fix the problems you identified in (1)?

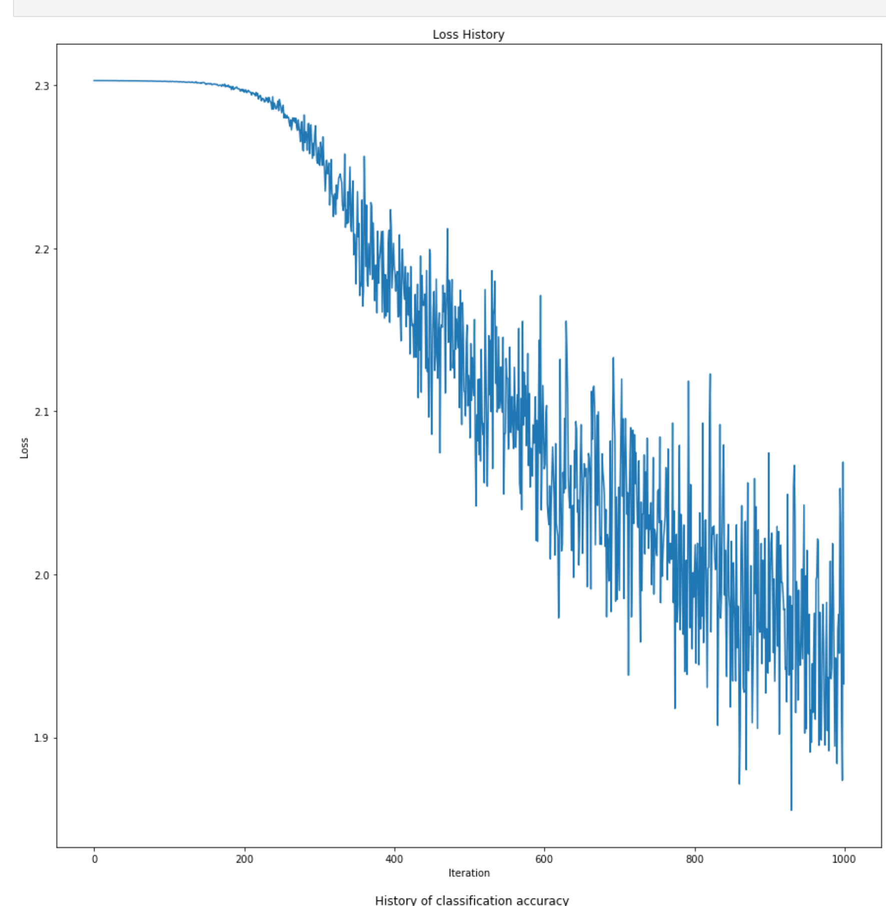
In [145…
```python
stats['train_acc_history']
```

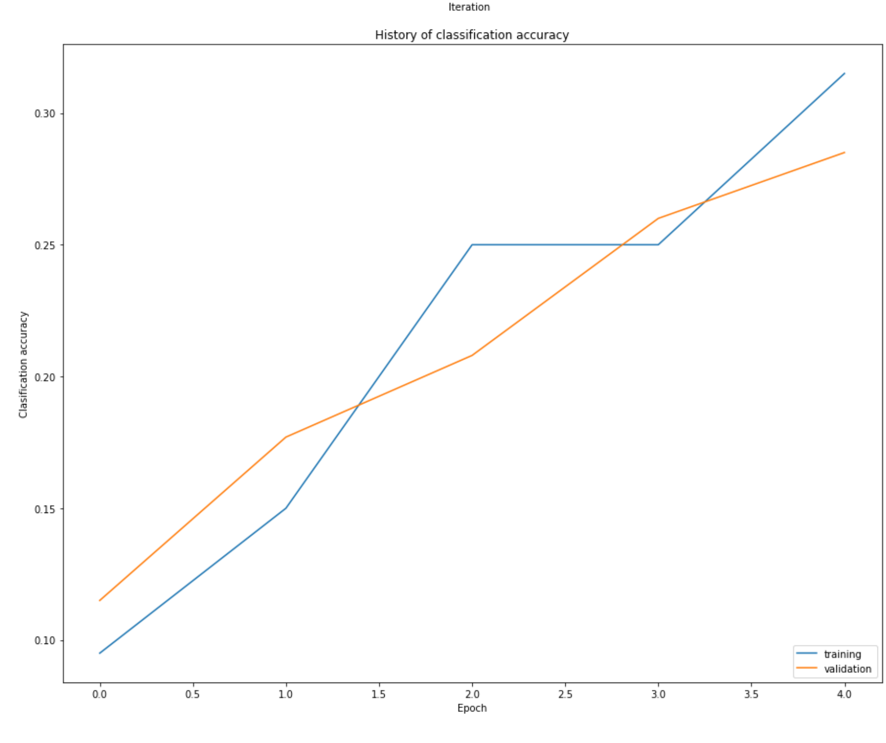
Out[145…
```
[0.095, 0.15, 0.25, 0.25, 0.315]
```

In [148…
```python
stats['val_acc_history']
```

Out[148…
```
[0.115, 0.177, 0.208, 0.26, 0.285]
```

In [149…
```python
# ================================================================ #
# YOUR CODE HERE:
#   Do some debugging to gain some insight into why the optimization
#   isn't great.
# ================================================================ #

# Plot the loss function and train / validation accuracies

plt.plot(stats['loss_history'])
plt.title('Loss History')
plt.xlabel('Iteration')
plt.ylabel('Loss')
plt.gcf().set_size_inches(15, 15)
plt.show()

fig, ax = plt.subplots()
plt.plot(stats['train_acc_history'], label='training')
plt.plot(stats['val_acc_history'], label='validation')

legend = ax.legend(loc='upper center', shadow=True)
frame = legend.get_frame()
frame.set_facecolor('0.80')

plt.title('History of classification accuracy')
plt.xlabel('Epoch')
plt.ylabel('Clasification accuracy')
plt.legend(loc='lower right')
plt.gcf().set_size_inches(15, 12)
plt.show()

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





Answers:

(1) From the first plot, we can say that the Loss for the first 200 iterations is the same and does not decrease. Post that, we can observe that the Loss decreases linearly which suggests that the Learning Rate used is not large enough. Secondly, from the Accuracy plot, we observe that for the first 1000 iterations, both the training and validation accuracy seems to increase and do not converge. This might suggest that we need to increase the number of iterations to make the model converge.

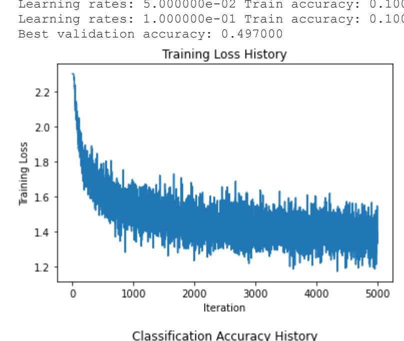
(2) The Learning Rate must be increased to ensure the Loss drops exponentially rather than Linearly. Also, the number of iterations must be increased to ensure model covergence.

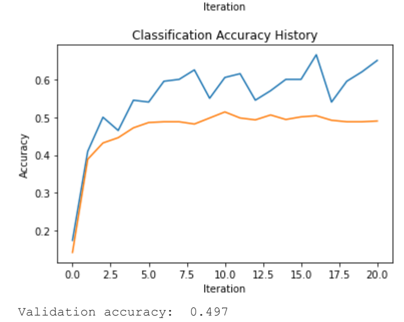
Optimize the neural network

Use the following part of the Jupyter notebook to optimize your hyperparameters on the validation set. Store your nets as best_net.

In [150…
```python
best_net = None # store the best model into this

# ================================================================ #
# YOUR CODE HERE:
#   Optimize over your hyperparameters to arrive at the best neural
#   network.  You should be able to get over 50% validation accuracy.
#   For this part of the notebook, we will give credit based on the
#   accuracy you get.  Your score on this question will be multiplied by:
#      min(floor((X - 28%)) / %22, 1)
#   where if you get 50% or higher validation accuracy, you get full
#   points.
#
#   Note, you need to use the same network structure (keep hidden_size = 50)!
# ================================================================ #

best_acc = -1
learning_rates = [1e-1, 5e-2, 3e-2, 1e-2, 5e-3, 3e-3, 1e-3]
results = {}
for lr in learning_rates:
    net = TwoLayerNet(input_size, hidden_size, num_classes)
    stats = net.train(X_train, y_train, X_val, y_val,
                num_iters=5000, batch_size=200,
                learning_rate=lr, learning_rate_decay=0.95,
                reg=0.55)

    y_train_pred = net.predict(X_train)
    acc_train = np.mean(y_train == y_train_pred)
    y_val_pred = net.predict(X_val)
    acc_val = np.mean(y_val == y_val_pred)

    results[lr] = (acc_train, acc_val)

    if best_acc < acc_val:
        best_acc = acc_val
        best_net = net

# Print out results
for lr in sorted(results):
    train_acc, val_acc = results[lr]
    print ('Learning rates: %e Train accuracy: %f Validation accuracy: %f' % (lr, train_acc, val_acc))

print ('Best validation accuracy: %f' % best_acc)

# Plot the loss function and train / validation accuracies
plt.plot(stats['loss_history'])
plt.title('Training Loss History')
plt.xlabel('Iteration')
plt.ylabel('Training Loss')
plt.show()

plt.plot(stats['train_acc_history'])
plt.plot(stats['val_acc_history'])
plt.title('Classification Accuracy History')
plt.xlabel('Iteration')
plt.ylabel('Accuracy')
plt.show()

# ================================================================ #
# END YOUR CODE HERE
# ================================================================ #
val_acc = (best_net.predict(X_val) == y_val).mean()
print('Validation accuracy: ', val_acc)
```

```
/Users/siddie/Desktop/Winter-2022/NN & DL/Neural Networks & Deep Learning/HW3/hw3-code/nndl/neural_net.py:110: RuntimeWarning: overflow encountered in exp
  class_probabilities = np.exp(scores)/np.sum(np.exp(scores), axis=1, keepdims=True)
/Users/siddie/Desktop/Winter-2022/NN & DL/Neural Networks & Deep Learning/HW3/hw3-code/nndl/neural_net.py:110: RuntimeWarning: invalid value encountered in true_divide
  class_probabilities = np.exp(scores)/np.sum(np.exp(scores), axis=1, keepdims=True)
/Users/siddie/Desktop/Winter-2022/NN & DL/Neural Networks & Deep Learning/HW3/hw3-code/nndl/neural_net.py:112: RuntimeWarning: divide by zero encountered in log
  log_loss = -np.log(prob_of_correct_y)
Learning rates: 1.000000e-03 Train accuracy: 0.564673 Validation accuracy: 0.497000
Learning rates: 3.000000e-03 Train accuracy: 0.544633 Validation accuracy: 0.477000
Learning rates: 5.000000e-03 Train accuracy: 0.100265 Validation accuracy: 0.087000
Learning rates: 1.000000e-02 Train accuracy: 0.100265 Validation accuracy: 0.087000
Learning rates: 3.000000e-02 Train accuracy: 0.100265 Validation accuracy: 0.087000
Learning rates: 5.000000e-02 Train accuracy: 0.100265 Validation accuracy: 0.087000
Learning rates: 1.000000e-01 Train accuracy: 0.100265 Validation accuracy: 0.087000
Best validation accuracy: 0.497000
```





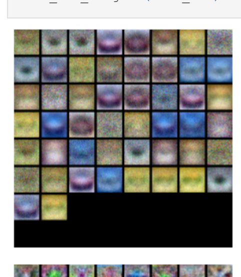

```
Validation accuracy:  0.497
```

In [156…
```python
best_net = None

Target_accuracy = 0.50


batch = 200
Learning_rate_min = -3.5
Learning_rate_max = -3

Reg_min = 0.1
Reg_max = 0.25


learning_rates = list(10**np.arange(Learning_rate_min,Learning_rate_max,0.2))
reg_coeffs = list(np.arange(Reg_min,Reg_max,0.05))

d1 = False
d2 = False


for rate in learning_rates:
    for regu in reg_coeffs:


        NeuralNetwork = TwoLayerNet(input_size, hidden_size, num_classes)


        NeuralNetwork.train(X_train, y_train, X_val, y_val,
            num_iters=3000, batch_size=batch,
            learning_rate=rate, learning_rate_decay=0.95,
            reg=regu, verbose=False)

        val_acc = (NeuralNetwork.predict(X_val) == y_val).mean()
        print('Validation accuracy:', val_acc,'\n')
        print('Learning Rate:', rate, '\n')
        print('Regression coefficent:',regu,'\n')

        if val_acc >= Target_accuracy:
            net = NeuralNetwork
            d1 = True
            print('Validation accuracy:', val_acc,'\n')
            print('Learning Rate:', rate, '\n')
            print('Regression coefficent:',regu,'\n')
            print('\n')
            break

    if d1:
        d2 = True
        break


best_net = net
# ================================================================ #
# END YOUR CODE HERE
# ================================================================ #
val_acc = (best_net.predict(X_val) == y_val).mean()
print('Validation accuracy: ', val_acc)
```

```
Validation accuracy: 0.468

Learning Rate: 0.00031622776601683794

Regression coefficent: 0.1

Validation accuracy: 0.458

Learning Rate: 0.00031622776601683794

Regression coefficent: 0.15000000000000002

Validation accuracy: 0.483

Learning Rate: 0.00031622776601683794

Regression coefficent: 0.20000000000000004

Validation accuracy: 0.506

Learning Rate: 0.0005011872336272725

Regression coefficent: 0.1

Validation accuracy: 0.506

Learning Rate: 0.0005011872336272725

Regression coefficent: 0.1


Validation accuracy:  0.506
```

In [ ]:

In [152…
```python
from utils.vis_utils import visualize_grid

# Visualize the weights of the network

def show_net_weights(net):
    W1 = net.params['W1']
    W1 = W1.T.reshape(32, 32, 3, -1).transpose(3, 0, 1, 2)
    plt.imshow(visualize_grid(W1, padding=3).astype('uint8'))
    plt.gca().axis('off')
    plt.show()

show_net_weights(subopt_net)
show_net_weights(best_net)
```



Question:

(1) What differences do you see in the weights between the suboptimal net and the best net you arrived at?

Answer:

(1) The weights of the subopt_net looks to be less pronounced in terms of visual features than that of the best_net. Specific shapes are easily discernable in the best_net, whereas the subopt_net net's weights appear to be smoothed or averaged. Moreover, we can also observe that the suboptimal net contains much more noises and hence we can hardly distinguis the difference between the images. But in the case of the, we have more features to distinguish images.

Evaluate on test set

In [153…
```python
test_acc = (best_net.predict(X_test) == y_test).mean()
print('Test accuracy: ', test_acc)
```

```
Test accuracy:  0.512
```

In [ ]:

neural_net.py

In [159…

```python
import numpy as np
import matplotlib.pyplot as plt


class TwoLayerNet(object):
  """
  A two-layer fully-connected neural network. The net has an input dimension of
  N, a hidden layer dimension of H, and performs classification over C classes.
  We train the network with a softmax loss function and L2 regularization on the
  weight matrices. The network uses a ReLU nonlinearity after the first fully
  connected layer.

  In other words, the network has the following architecture:

  input - fully connected layer - ReLU - fully connected layer - softmax

  The outputs of the second fully-connected layer are the scores for each class.
  """

  def __init__(self, input_size, hidden_size, output_size, std=1e-4):
    """
    Initialize the model. Weights are initialized to small random values and
    biases are initialized to zero. Weights and biases are stored in the
    variable self.params, which is a dictionary with the following keys:

    W1: First layer weights; has shape (H, D)
    b1: First layer biases; has shape (H,)
    W2: Second layer weights; has shape (C, H)
    b2: Second layer biases; has shape (C,)

    Inputs:
    - input_size: The dimension D of the input data.
    - hidden_size: The number of neurons H in the hidden layer.
    - output_size: The number of classes C.
    """
    self.params = {}
    self.params['W1'] = std * np.random.randn(hidden_size, input_size)
    self.params['b1'] = np.zeros(hidden_size)
    self.params['W2'] = std * np.random.randn(output_size, hidden_size)
    self.params['b2'] = np.zeros(output_size)


  def loss(self, X, y=None, reg=0.0):
    """
    Compute the loss and gradients for a two layer fully connected neural
    network.

    Inputs:
    - X: Input data of shape (N, D). Each X[i] is a training sample.
    - y: Vector of training labels. y[i] is the label for X[i], and each y[i] is
      an integer in the range 0 <= y[i] < C. This parameter is optional; if it
      is not passed then we only return scores, and if it is passed then we
      instead return the loss and gradients.
    - reg: Regularization strength.

    Returns:
    If y is None, return a matrix scores of shape (N, C) where scores[i, c] is
    the score for class c on input X[i].

    If y is not None, instead return a tuple of:
    - loss: Loss (data loss and regularization loss) for this batch of training
      samples.
    - grads: Dictionary mapping parameter names to gradients of those parameters
      with respect to the loss function; has the same keys as self.params.
    """
    # Unpack variables from the params dictionary
    W1, b1 = self.params['W1'], self.params['b1']
    W2, b2 = self.params['W2'], self.params['b2']
    N, D = X.shape

    # Compute the forward pass
    scores = None

    # ================================================================ #
    # YOUR CODE HERE:
    #   Calculate the output scores of the neural network.  The result
    #   should be (N, C). As stated in the description for this class,
    #   there should not be a ReLU layer after the second FC layer.
    #   The output of the second FC layer is the output scores. Do not
    #   use a for loop in your implementation.
    # ================================================================ #

    relu = lambda x: np.maximum(0, x)
    H1 = relu(X@(W1.T) + b1) # X is (5,4) W1.T (4,10) -> (5,10)
    scores = H1@W2.T + b2 # H1 is (5,10), W2 is (10,10) -> (5,10)


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


    # If the targets are not given then jump out, we're done
    if y is None:
      return scores

    # Compute the loss
    loss = None

    # ================================================================ #
    # YOUR CODE HERE:
    #   Calculate the loss of the neural network.  This includes the
    #   softmax loss and the L2 regularization for W1 and W2. Store the
    #   total loss in teh variable loss.  Multiply the regularization
    #   loss by 0.5 (in addition to the factor reg).
    # ================================================================ #

    # scores is num_examples by num_classes

    class_prob = np.exp(scores)/np.sum(np.exp(scores), axis=1, keepdims=True)
    prob_correct_y = class_prob[np.arange(N), y]
    log_loss = -np.log(prob_correct_y)
    sum_log_loss = np.sum(log_loss)
    loss = sum_log_loss/N

    frob_norm_w1 = np.sum(W1**2)
    frob_norm_w2 = np.sum(W2**2)
    reg_w1 = 0.5*reg*frob_norm_w1
    reg_w2 = 0.5*reg*frob_norm_w2
    regularized_loss = reg_w1 + reg_w2

    loss += regularized_loss


    #reg_loss = 0.5 * reg * (np.linalg.norm(W1, 'fro')**2 + np.linalg.norm(W2, 'fro')**2)
    #loss = (np.sum(-np.log(np.exp(scores[np.arange(N), y]) / np.sum(np.exp(scores), axis = 1))))/ N + reg_los

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

    grads = {}

    # ================================================================ #
    # YOUR CODE HERE:
    #   Implement the backward pass.  Compute the derivatives of the
    #   weights and the biases.  Store the results in the grads
    #   dictionary.  e.g., grads['W1'] should store the gradient for
    #   W1, and be of the same size as W1.
    # ================================================================ #

    update_scores = class_probabilities
    update_scores[np.arange(N), y] -=1
    update_scores /= N

    grads['W2'] = np.dot(H1.T, update_scores).T
    grads['b2'] = np.sum(update_scores, axis=0)
    dH2 = np.dot(update_scores, W2)

    dLdA = dH2
    dLdA[H1 <= 0] = 0

    grads['W1'] = np.dot(dLdA.T, X)
    grads['b1'] = np.sum(dLdA, axis=0)

    grads['W2'] += reg * W2
    grads['W1'] += reg * W1


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

    return loss, grads

  def train(self, X, y, X_val, y_val,
            learning_rate=1e-3, learning_rate_decay=0.95,
            reg=1e-5, num_iters=100,
            batch_size=200, verbose=False):
    """
    Train this neural network using stochastic gradient descent.

    Inputs:
    - X: A numpy array of shape (N, D) giving training data.
    - y: A numpy array f shape (N,) giving training labels; y[i] = c means that
      X[i] has label c, where 0 <= c < C.
    - X_val: A numpy array of shape (N_val, D) giving validation data.
    - y_val: A numpy array of shape (N_val,) giving validation labels.
    - learning_rate: Scalar giving learning rate for optimization.
    - learning_rate_decay: Scalar giving factor used to decay the learning rate
      after each epoch.
    - reg: Scalar giving regularization strength.
    - num_iters: Number of steps to take when optimizing.
    - batch_size: Number of training examples to use per step.
    - verbose: boolean; if true print progress during optimization.
    """
    num_train = X.shape[0]
    iterations_per_epoch = max(num_train / batch_size, 1)

    # Use SGD to optimize the parameters in self.model
    loss_history = []
    train_acc_history = []
    val_acc_history = []

    for it in np.arange(num_iters):
      X_batch = None
      y_batch = None

      # ================================================================ #
      # YOUR CODE HERE:
      #   Create a minibatch by sampling batch_size samples randomly.
      # ================================================================ #

      random_indices = np.random.choice(np.arange(num_train), batch_size)
      X_batch = X[random_indices]
      y_batch = y[random_indices]

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

       # Compute loss and gradients using the current minibatch
      loss, grads = self.loss(X_batch, y=y_batch, reg=reg)
      loss_history.append(loss)

      # ================================================================ #
      # YOUR CODE HERE:
      #   Perform a gradient descent step using the minibatch to update
      #   all parameters (i.e., W1, W2, b1, and b2).
      # ================================================================ #

      self.params['W2'] += -learning_rate * grads['W2']
      self.params['W1'] += -learning_rate * grads['W1']

      self.params['b2'] += -learning_rate * grads['b2']
      self.params['b1'] += -learning_rate * grads['b1']

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

      if verbose and it % 100 == 0:
        print('iteration {} / {}: loss {}'.format(it, num_iters, loss))

      # Every epoch, check train and val accuracy and decay learning rate.
      if it % iterations_per_epoch == 0:
        # Check accuracy
        train_acc = (self.predict(X_batch) == y_batch).mean()
        val_acc = (self.predict(X_val) == y_val).mean()
        train_acc_history.append(train_acc)
        val_acc_history.append(val_acc)

        # Decay learning rate
        learning_rate *= learning_rate_decay

    return {
      'loss_history': loss_history,
      'train_acc_history': train_acc_history,
      'val_acc_history': val_acc_history,
    }

  def predict(self, X):
    """
    Use the trained weights of this two-layer network to predict labels for
    data points. For each data point we predict scores for each of the C
    classes, and assign each data point to the class with the highest score.

    Inputs:
    - X: A numpy array of shape (N, D) giving N D-dimensional data points to
      classify.

    Returns:
    - y_pred: A numpy array of shape (N,) giving predicted labels for each of
      the elements of X. For all i, y_pred[i] = c means that X[i] is predicted
      to have class c, where 0 <= c < C.
    """
    y_pred = None

    # ================================================================ #
    # YOUR CODE HERE:
    #   Predict the class given the input data.
    # ================================================================ #

    num_examples = X.shape[0]
    y_pred = np.empty((num_examples,), dtype=int)
    H1_preact = np.dot(X, self.params['W1'].T) + self.params['b1']

    #apply RELU
    H1  = np.maximum(0, H1_preact)

    #second layer
    H2  = np.dot(H1, self.params['W2'].T) + self.params['b2']

    #apply softmax
    softmax = np.exp(H2)/np.sum(np.exp(H2), axis=1, keepdims=True)

    for i in range(num_examples):
        max_index = np.argmax(softmax[i])
        y_pred[i] = max_index

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

    return y_pred
```

In [ ]: