

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_numerical_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-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))))
```

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 [24]:

```
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: 0.00025200276523324475
W2 max relative error: 0.0020005501731676966
W3 max relative error: 5.299491830391451e-05
b1 max relative error: 3.0485099495645355e-05
b2 max relative error: 1.4887541758067226e-07
b3 max relative error: 1.4950389511421633e-09
```

Overfit small dataset

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

In [5]:

```
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)

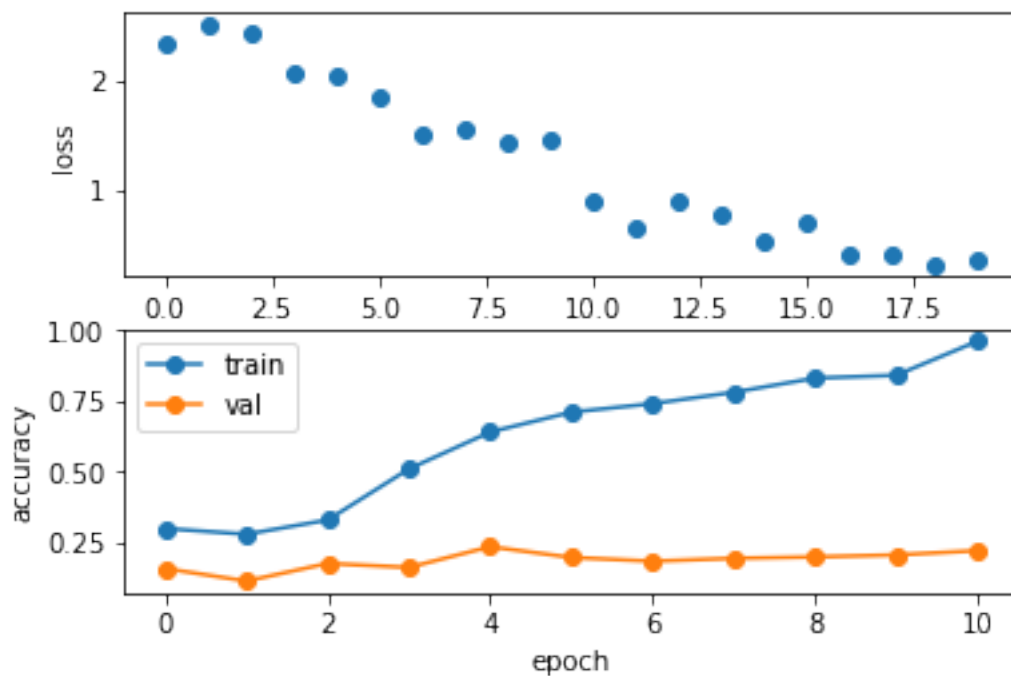
solver.train()

(Iteration 1 / 20) loss: 2.345829
(Epoch 0 / 10) train acc: 0.300000; val_acc: 0.159000
(Iteration 2 / 20) loss: 2.510076
(Epoch 1 / 10) train acc: 0.280000; val_acc: 0.115000
(Iteration 3 / 20) loss: 2.422183
(Iteration 4 / 20) loss: 2.057817
(Epoch 2 / 10) train acc: 0.330000; val_acc: 0.177000
(Iteration 5 / 20) loss: 2.038447
(Iteration 6 / 20) loss: 1.840685
(Epoch 3 / 10) train acc: 0.510000; val_acc: 0.163000
(Iteration 7 / 20) loss: 1.513158
(Iteration 8 / 20) loss: 1.543478
(Epoch 4 / 10) train acc: 0.640000; val_acc: 0.236000
(Iteration 9 / 20) loss: 1.435310
(Iteration 10 / 20) loss: 1.445056
(Epoch 5 / 10) train acc: 0.710000; val_acc: 0.198000
(Iteration 11 / 20) loss: 0.902785
(Iteration 12 / 20) loss: 0.646936
(Epoch 6 / 10) train acc: 0.740000; val_acc: 0.185000
(Iteration 13 / 20) loss: 0.900255
(Iteration 14 / 20) loss: 0.780840
(Epoch 7 / 10) train acc: 0.780000; val_acc: 0.195000
(Iteration 15 / 20) loss: 0.519394
(Iteration 16 / 20) loss: 0.693193
(Epoch 8 / 10) train acc: 0.830000; val_acc: 0.200000
(Iteration 17 / 20) loss: 0.412564
(Iteration 18 / 20) loss: 0.401652
(Epoch 9 / 10) train acc: 0.840000; val_acc: 0.207000
(Iteration 19 / 20) loss: 0.315550
(Iteration 20 / 20) loss: 0.353503
(Epoch 10 / 10) train acc: 0.960000; val_acc: 0.222000
```

In [6]:

```
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 [27]:

```
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: 2.304584
(Epoch 0 / 1) train acc: 0.101000; val_acc: 0.087000
(Iteration 21 / 980) loss: 2.283690
(Iteration 41 / 980) loss: 1.964456
(Iteration 61 / 980) loss: 1.912247
(Iteration 81 / 980) loss: 1.857748
(Iteration 101 / 980) loss: 2.094499
(Iteration 121 / 980) loss: 1.626825
(Iteration 141 / 980) loss: 1.968974
(Iteration 161 / 980) loss: 1.717889
(Iteration 181 / 980) loss: 1.592656
(Iteration 201 / 980) loss: 1.660863
(Iteration 221 / 980) loss: 1.673992
(Iteration 241 / 980) loss: 1.762689
(Iteration 261 / 980) loss: 1.609536
(Iteration 281 / 980) loss: 1.808322
(Iteration 301 / 980) loss: 1.658396
(Iteration 321 / 980) loss: 1.615102
(Iteration 341 / 980) loss: 1.731093
(Iteration 361 / 980) loss: 1.605881
(Iteration 381 / 980) loss: 1.702318
(Iteration 401 / 980) loss: 1.623992
(Iteration 421 / 980) loss: 1.541325
(Iteration 441 / 980) loss: 1.798779
(Iteration 461 / 980) loss: 1.721100
(Iteration 481 / 980) loss: 1.440733
(Iteration 501 / 980) loss: 1.622227
(Iteration 521 / 980) loss: 1.612905
(Iteration 541 / 980) loss: 1.359487
(Iteration 561 / 980) loss: 1.732130
(Iteration 581 / 980) loss: 1.599631
(Iteration 601 / 980) loss: 1.587611
(Iteration 621 / 980) loss: 1.494315
(Iteration 641 / 980) loss: 1.881654
(Iteration 661 / 980) loss: 1.749850
(Iteration 681 / 980) loss: 1.904053
(Iteration 701 / 980) loss: 1.514551
(Iteration 721 / 980) loss: 1.478754
(Iteration 741 / 980) loss: 1.563152
(Iteration 761 / 980) loss: 1.809675
(Iteration 781 / 980) loss: 1.599222
(Iteration 801 / 980) loss: 1.793895
(Iteration 821 / 980) loss: 1.656165
(Iteration 841 / 980) loss: 1.551278
(Iteration 861 / 980) loss: 1.408013
(Iteration 881 / 980) loss: 1.744363
(Iteration 901 / 980) loss: 1.588459
(Iteration 921 / 980) loss: 1.562788
(Iteration 941 / 980) loss: 1.347889
(Iteration 961 / 980) loss: 1.301060
(Epoch 1 / 1) train acc: 0.468000; val_acc: 0.471000

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 after 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 of 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 [22]:

```
# ===== #
# YOUR CODE HERE:
#   Implement a CNN to achieve greater than 65% validation accuracy
#   on CIFAR-10.
# ===== #
model = ThreeLayerConvNet(num_filters=64,
                           filter_size=3,
                           hidden_dim=500,
                           reg=0.01,
                           weight_scale=1e-3)

solver = Solver(model, data,
                 num_epochs=20, batch_size=500,
                 update_rule='adam',
                 optim_config={
                     'learning_rate': 1e-3,
                 },
                 lr_decay=0.95,
                 verbose=True, print_every=1)

solver.train()

plt.figure()
plt.plot(solver.val_acc_history, '-x')
plt.plot(0.65*np.ones(len(solver.val_acc_history)), '--')
plt.xlabel('epoch')
plt.ylabel('val_acc')
plt.grid()
plt.show()

plt.figure()
plt.plot(solver.train_acc_history, '-x')
plt.plot(0.65*np.ones(len(solver.val_acc_history)), '--')
plt.xlabel('epoch')
plt.ylabel('train_acc')
plt.grid()
plt.show()

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

```
(Iteration 1 / 1960) loss: 2.343565
(Epoch 0 / 20) train acc: 0.105000; val_acc: 0.112000
(Iteration 2 / 1960) loss: 2.385247
(Iteration 3 / 1960) loss: 2.296587
(Iteration 4 / 1960) loss: 2.237274
(Iteration 5 / 1960) loss: 2.186507
(Iteration 6 / 1960) loss: 2.567092
(Iteration 7 / 1960) loss: 2.753471
(Iteration 8 / 1960) loss: 2.297671
(Iteration 9 / 1960) loss: 2.213289
(Iteration 10 / 1960) loss: 2.225299
(Iteration 11 / 1960) loss: 2.239106
(Iteration 12 / 1960) loss: 2.188178
(Iteration 13 / 1960) loss: 2.173347
```


(Iteration 14 / 1960)	loss: 2.154017
(Iteration 15 / 1960)	loss: 2.061950
(Iteration 16 / 1960)	loss: 2.105759
(Iteration 17 / 1960)	loss: 2.161776
(Iteration 18 / 1960)	loss: 2.174631
(Iteration 19 / 1960)	loss: 2.079294
(Iteration 20 / 1960)	loss: 1.979490
(Iteration 21 / 1960)	loss: 2.054473
(Iteration 22 / 1960)	loss: 2.016585
(Iteration 23 / 1960)	loss: 2.032285
(Iteration 24 / 1960)	loss: 2.017034
(Iteration 25 / 1960)	loss: 1.957774
(Iteration 26 / 1960)	loss: 2.025098
(Iteration 27 / 1960)	loss: 1.956825
(Iteration 28 / 1960)	loss: 1.869516
(Iteration 29 / 1960)	loss: 1.966186
(Iteration 30 / 1960)	loss: 1.881188
(Iteration 31 / 1960)	loss: 1.921577
(Iteration 32 / 1960)	loss: 1.892150
(Iteration 33 / 1960)	loss: 1.876006
(Iteration 34 / 1960)	loss: 1.891655
(Iteration 35 / 1960)	loss: 1.954159
(Iteration 36 / 1960)	loss: 1.755801
(Iteration 37 / 1960)	loss: 1.805282
(Iteration 38 / 1960)	loss: 1.906301
(Iteration 39 / 1960)	loss: 1.856070
(Iteration 40 / 1960)	loss: 1.840667
(Iteration 41 / 1960)	loss: 1.808860
(Iteration 42 / 1960)	loss: 1.836913
(Iteration 43 / 1960)	loss: 1.833615
(Iteration 44 / 1960)	loss: 1.885141
(Iteration 45 / 1960)	loss: 1.875073
(Iteration 46 / 1960)	loss: 1.762202
(Iteration 47 / 1960)	loss: 1.812130
(Iteration 48 / 1960)	loss: 1.841729
(Iteration 49 / 1960)	loss: 1.794073
(Iteration 50 / 1960)	loss: 1.845120
(Iteration 51 / 1960)	loss: 1.759300
(Iteration 52 / 1960)	loss: 1.827716
(Iteration 53 / 1960)	loss: 1.752667
(Iteration 54 / 1960)	loss: 1.856389
(Iteration 55 / 1960)	loss: 1.731462
(Iteration 56 / 1960)	loss: 1.813152
(Iteration 57 / 1960)	loss: 1.755596
(Iteration 58 / 1960)	loss: 1.741325
(Iteration 59 / 1960)	loss: 1.725246
(Iteration 60 / 1960)	loss: 1.752294
(Iteration 61 / 1960)	loss: 1.662688
(Iteration 62 / 1960)	loss: 1.668639
(Iteration 63 / 1960)	loss: 1.651701
(Iteration 64 / 1960)	loss: 1.667660
(Iteration 65 / 1960)	loss: 1.669732
(Iteration 66 / 1960)	loss: 1.565830
(Iteration 67 / 1960)	loss: 1.617982
(Iteration 68 / 1960)	loss: 1.586608
(Iteration 69 / 1960)	loss: 1.636444
(Iteration 70 / 1960)	loss: 1.676938

(Iteration 71 / 1960) loss: 1.684794
(Iteration 72 / 1960) loss: 1.656593
(Iteration 73 / 1960) loss: 1.609120
(Iteration 74 / 1960) loss: 1.575116
(Iteration 75 / 1960) loss: 1.673141
(Iteration 76 / 1960) loss: 1.540742
(Iteration 77 / 1960) loss: 1.628490
(Iteration 78 / 1960) loss: 1.624494
(Iteration 79 / 1960) loss: 1.805095
(Iteration 80 / 1960) loss: 1.614804
(Iteration 81 / 1960) loss: 1.556009
(Iteration 82 / 1960) loss: 1.576091
(Iteration 83 / 1960) loss: 1.702229
(Iteration 84 / 1960) loss: 1.506828
(Iteration 85 / 1960) loss: 1.626026
(Iteration 86 / 1960) loss: 1.501582
(Iteration 87 / 1960) loss: 1.709705
(Iteration 88 / 1960) loss: 1.567630
(Iteration 89 / 1960) loss: 1.568545
(Iteration 90 / 1960) loss: 1.667193
(Iteration 91 / 1960) loss: 1.593867
(Iteration 92 / 1960) loss: 1.544554
(Iteration 93 / 1960) loss: 1.559168
(Iteration 94 / 1960) loss: 1.561963
(Iteration 95 / 1960) loss: 1.630886
(Iteration 96 / 1960) loss: 1.472907
(Iteration 97 / 1960) loss: 1.547311
(Iteration 98 / 1960) loss: 1.543203
(Epoch 1 / 20) train acc: 0.518000; val_acc: 0.501000
(Iteration 99 / 1960) loss: 1.656219
(Iteration 100 / 1960) loss: 1.527999
(Iteration 101 / 1960) loss: 1.557148
(Iteration 102 / 1960) loss: 1.535021
(Iteration 103 / 1960) loss: 1.527311
(Iteration 104 / 1960) loss: 1.544737
(Iteration 105 / 1960) loss: 1.528173
(Iteration 106 / 1960) loss: 1.445582
(Iteration 107 / 1960) loss: 1.560132
(Iteration 108 / 1960) loss: 1.494620
(Iteration 109 / 1960) loss: 1.468717
(Iteration 110 / 1960) loss: 1.537121
(Iteration 111 / 1960) loss: 1.457865
(Iteration 112 / 1960) loss: 1.450961
(Iteration 113 / 1960) loss: 1.624874
(Iteration 114 / 1960) loss: 1.437167
(Iteration 115 / 1960) loss: 1.582322
(Iteration 116 / 1960) loss: 1.483846
(Iteration 117 / 1960) loss: 1.514785
(Iteration 118 / 1960) loss: 1.472052
(Iteration 119 / 1960) loss: 1.422220
(Iteration 120 / 1960) loss: 1.412731
(Iteration 121 / 1960) loss: 1.382991
(Iteration 122 / 1960) loss: 1.457935
(Iteration 123 / 1960) loss: 1.449998
(Iteration 124 / 1960) loss: 1.573014
(Iteration 125 / 1960) loss: 1.464498
(Iteration 126 / 1960) loss: 1.417206

(Iteration 127 / 1960)	loss: 1.486660
(Iteration 128 / 1960)	loss: 1.465129
(Iteration 129 / 1960)	loss: 1.451660
(Iteration 130 / 1960)	loss: 1.373159
(Iteration 131 / 1960)	loss: 1.459731
(Iteration 132 / 1960)	loss: 1.399709
(Iteration 133 / 1960)	loss: 1.476719
(Iteration 134 / 1960)	loss: 1.447652
(Iteration 135 / 1960)	loss: 1.410237
(Iteration 136 / 1960)	loss: 1.441349
(Iteration 137 / 1960)	loss: 1.442403
(Iteration 138 / 1960)	loss: 1.424790
(Iteration 139 / 1960)	loss: 1.432346
(Iteration 140 / 1960)	loss: 1.435105
(Iteration 141 / 1960)	loss: 1.451795
(Iteration 142 / 1960)	loss: 1.512999
(Iteration 143 / 1960)	loss: 1.483192
(Iteration 144 / 1960)	loss: 1.426327
(Iteration 145 / 1960)	loss: 1.545199
(Iteration 146 / 1960)	loss: 1.452568
(Iteration 147 / 1960)	loss: 1.464190
(Iteration 148 / 1960)	loss: 1.461245
(Iteration 149 / 1960)	loss: 1.476561
(Iteration 150 / 1960)	loss: 1.478875
(Iteration 151 / 1960)	loss: 1.453696
(Iteration 152 / 1960)	loss: 1.463028
(Iteration 153 / 1960)	loss: 1.548183
(Iteration 154 / 1960)	loss: 1.437075
(Iteration 155 / 1960)	loss: 1.551315
(Iteration 156 / 1960)	loss: 1.452716
(Iteration 157 / 1960)	loss: 1.554678
(Iteration 158 / 1960)	loss: 1.495523
(Iteration 159 / 1960)	loss: 1.464560
(Iteration 160 / 1960)	loss: 1.421957
(Iteration 161 / 1960)	loss: 1.446223
(Iteration 162 / 1960)	loss: 1.500812
(Iteration 163 / 1960)	loss: 1.495957
(Iteration 164 / 1960)	loss: 1.503432
(Iteration 165 / 1960)	loss: 1.501002
(Iteration 166 / 1960)	loss: 1.555011
(Iteration 167 / 1960)	loss: 1.413926
(Iteration 168 / 1960)	loss: 1.412978
(Iteration 169 / 1960)	loss: 1.503217
(Iteration 170 / 1960)	loss: 1.539780
(Iteration 171 / 1960)	loss: 1.352872
(Iteration 172 / 1960)	loss: 1.410864
(Iteration 173 / 1960)	loss: 1.431207
(Iteration 174 / 1960)	loss: 1.424105
(Iteration 175 / 1960)	loss: 1.379849
(Iteration 176 / 1960)	loss: 1.342909
(Iteration 177 / 1960)	loss: 1.439173
(Iteration 178 / 1960)	loss: 1.328158
(Iteration 179 / 1960)	loss: 1.449530
(Iteration 180 / 1960)	loss: 1.481471
(Iteration 181 / 1960)	loss: 1.411855
(Iteration 182 / 1960)	loss: 1.431526
(Iteration 183 / 1960)	loss: 1.397030

(Iteration 184 / 1960) loss: 1.452568
(Iteration 185 / 1960) loss: 1.413170
(Iteration 186 / 1960) loss: 1.348119
(Iteration 187 / 1960) loss: 1.402097
(Iteration 188 / 1960) loss: 1.383570
(Iteration 189 / 1960) loss: 1.338503
(Iteration 190 / 1960) loss: 1.338460
(Iteration 191 / 1960) loss: 1.374681
(Iteration 192 / 1960) loss: 1.428324
(Iteration 193 / 1960) loss: 1.367912
(Iteration 194 / 1960) loss: 1.336644
(Iteration 195 / 1960) loss: 1.431452
(Iteration 196 / 1960) loss: 1.454461
(Epoch 2 / 20) train acc: 0.573000; val_acc: 0.562000
(Iteration 197 / 1960) loss: 1.392558
(Iteration 198 / 1960) loss: 1.301093
(Iteration 199 / 1960) loss: 1.426567
(Iteration 200 / 1960) loss: 1.389757
(Iteration 201 / 1960) loss: 1.391878
(Iteration 202 / 1960) loss: 1.452522
(Iteration 203 / 1960) loss: 1.344350
(Iteration 204 / 1960) loss: 1.386305
(Iteration 205 / 1960) loss: 1.454666
(Iteration 206 / 1960) loss: 1.384060
(Iteration 207 / 1960) loss: 1.420764
(Iteration 208 / 1960) loss: 1.411529
(Iteration 209 / 1960) loss: 1.432689
(Iteration 210 / 1960) loss: 1.381148
(Iteration 211 / 1960) loss: 1.344451
(Iteration 212 / 1960) loss: 1.359113
(Iteration 213 / 1960) loss: 1.400985
(Iteration 214 / 1960) loss: 1.298071
(Iteration 215 / 1960) loss: 1.340721
(Iteration 216 / 1960) loss: 1.366304
(Iteration 217 / 1960) loss: 1.307961
(Iteration 218 / 1960) loss: 1.424375
(Iteration 219 / 1960) loss: 1.340998
(Iteration 220 / 1960) loss: 1.420868
(Iteration 221 / 1960) loss: 1.300157
(Iteration 222 / 1960) loss: 1.384309
(Iteration 223 / 1960) loss: 1.434550
(Iteration 224 / 1960) loss: 1.359279
(Iteration 225 / 1960) loss: 1.420866
(Iteration 226 / 1960) loss: 1.374814
(Iteration 227 / 1960) loss: 1.335307
(Iteration 228 / 1960) loss: 1.398738
(Iteration 229 / 1960) loss: 1.452722
(Iteration 230 / 1960) loss: 1.352280
(Iteration 231 / 1960) loss: 1.394428
(Iteration 232 / 1960) loss: 1.436880
(Iteration 233 / 1960) loss: 1.358661
(Iteration 234 / 1960) loss: 1.391406
(Iteration 235 / 1960) loss: 1.216938
(Iteration 236 / 1960) loss: 1.323487
(Iteration 237 / 1960) loss: 1.447208
(Iteration 238 / 1960) loss: 1.357084
(Iteration 239 / 1960) loss: 1.446930

```
(Iteration 240 / 1960) loss: 1.370830
(Iteration 241 / 1960) loss: 1.317280
(Iteration 242 / 1960) loss: 1.348621
(Iteration 243 / 1960) loss: 1.296701
(Iteration 244 / 1960) loss: 1.316138
(Iteration 245 / 1960) loss: 1.281218
(Iteration 246 / 1960) loss: 1.284664
(Iteration 247 / 1960) loss: 1.312905
(Iteration 248 / 1960) loss: 1.411246
(Iteration 249 / 1960) loss: 1.390820
(Iteration 250 / 1960) loss: 1.323026
(Iteration 251 / 1960) loss: 1.307383
(Iteration 252 / 1960) loss: 1.437548
(Iteration 253 / 1960) loss: 1.320769
(Iteration 254 / 1960) loss: 1.246387
(Iteration 255 / 1960) loss: 1.341220
(Iteration 256 / 1960) loss: 1.383244
(Iteration 257 / 1960) loss: 1.317451
(Iteration 258 / 1960) loss: 1.352973
(Iteration 259 / 1960) loss: 1.428377
(Iteration 260 / 1960) loss: 1.275409
(Iteration 261 / 1960) loss: 1.251642
(Iteration 262 / 1960) loss: 1.368146
(Iteration 263 / 1960) loss: 1.314124
(Iteration 264 / 1960) loss: 1.299500
(Iteration 265 / 1960) loss: 1.292476
(Iteration 266 / 1960) loss: 1.294657
(Iteration 267 / 1960) loss: 1.317000
(Iteration 268 / 1960) loss: 1.389452
(Iteration 269 / 1960) loss: 1.286788
(Iteration 270 / 1960) loss: 1.202395
(Iteration 271 / 1960) loss: 1.263226
(Iteration 272 / 1960) loss: 1.416715
(Iteration 273 / 1960) loss: 1.363573
(Iteration 274 / 1960) loss: 1.311214
(Iteration 275 / 1960) loss: 1.410112
(Iteration 276 / 1960) loss: 1.347486
(Iteration 277 / 1960) loss: 1.259169
(Iteration 278 / 1960) loss: 1.344002
(Iteration 279 / 1960) loss: 1.300010
(Iteration 280 / 1960) loss: 1.383392
(Iteration 281 / 1960) loss: 1.237433
(Iteration 282 / 1960) loss: 1.473972
(Iteration 283 / 1960) loss: 1.390722
(Iteration 284 / 1960) loss: 1.409848
(Iteration 285 / 1960) loss: 1.365595
(Iteration 286 / 1960) loss: 1.348342
(Iteration 287 / 1960) loss: 1.320317
(Iteration 288 / 1960) loss: 1.299623
(Iteration 289 / 1960) loss: 1.399951
(Iteration 290 / 1960) loss: 1.303331
(Iteration 291 / 1960) loss: 1.320469
(Iteration 292 / 1960) loss: 1.402613
(Iteration 293 / 1960) loss: 1.339932
(Iteration 294 / 1960) loss: 1.279359
(Epoch 3 / 20) train acc: 0.641000; val_acc: 0.581000
(Iteration 295 / 1960) loss: 1.338577
```

(Iteration 296 / 1960)	loss: 1.312307
(Iteration 297 / 1960)	loss: 1.264532
(Iteration 298 / 1960)	loss: 1.223045
(Iteration 299 / 1960)	loss: 1.302771
(Iteration 300 / 1960)	loss: 1.397036
(Iteration 301 / 1960)	loss: 1.173235
(Iteration 302 / 1960)	loss: 1.284125
(Iteration 303 / 1960)	loss: 1.264594
(Iteration 304 / 1960)	loss: 1.375416
(Iteration 305 / 1960)	loss: 1.263190
(Iteration 306 / 1960)	loss: 1.240894
(Iteration 307 / 1960)	loss: 1.304236
(Iteration 308 / 1960)	loss: 1.245926
(Iteration 309 / 1960)	loss: 1.279383
(Iteration 310 / 1960)	loss: 1.366658
(Iteration 311 / 1960)	loss: 1.268112
(Iteration 312 / 1960)	loss: 1.213483
(Iteration 313 / 1960)	loss: 1.201949
(Iteration 314 / 1960)	loss: 1.228707
(Iteration 315 / 1960)	loss: 1.214473
(Iteration 316 / 1960)	loss: 1.244929
(Iteration 317 / 1960)	loss: 1.243458
(Iteration 318 / 1960)	loss: 1.335669
(Iteration 319 / 1960)	loss: 1.337207
(Iteration 320 / 1960)	loss: 1.309196
(Iteration 321 / 1960)	loss: 1.193133
(Iteration 322 / 1960)	loss: 1.248229
(Iteration 323 / 1960)	loss: 1.353841
(Iteration 324 / 1960)	loss: 1.255357
(Iteration 325 / 1960)	loss: 1.336337
(Iteration 326 / 1960)	loss: 1.278474
(Iteration 327 / 1960)	loss: 1.217753
(Iteration 328 / 1960)	loss: 1.229630
(Iteration 329 / 1960)	loss: 1.223097
(Iteration 330 / 1960)	loss: 1.201419
(Iteration 331 / 1960)	loss: 1.307785
(Iteration 332 / 1960)	loss: 1.308294
(Iteration 333 / 1960)	loss: 1.332053
(Iteration 334 / 1960)	loss: 1.199496
(Iteration 335 / 1960)	loss: 1.296040
(Iteration 336 / 1960)	loss: 1.200038
(Iteration 337 / 1960)	loss: 1.180671
(Iteration 338 / 1960)	loss: 1.252866
(Iteration 339 / 1960)	loss: 1.247616
(Iteration 340 / 1960)	loss: 1.209785
(Iteration 341 / 1960)	loss: 1.335454
(Iteration 342 / 1960)	loss: 1.255429
(Iteration 343 / 1960)	loss: 1.109344
(Iteration 344 / 1960)	loss: 1.213695
(Iteration 345 / 1960)	loss: 1.168615
(Iteration 346 / 1960)	loss: 1.303120
(Iteration 347 / 1960)	loss: 1.192766
(Iteration 348 / 1960)	loss: 1.333370
(Iteration 349 / 1960)	loss: 1.257825
(Iteration 350 / 1960)	loss: 1.277491
(Iteration 351 / 1960)	loss: 1.224563
(Iteration 352 / 1960)	loss: 1.338509

```
(Iteration 353 / 1960) loss: 1.241605
(Iteration 354 / 1960) loss: 1.369178
(Iteration 355 / 1960) loss: 1.300301
(Iteration 356 / 1960) loss: 1.177496
(Iteration 357 / 1960) loss: 1.258438
(Iteration 358 / 1960) loss: 1.236423
(Iteration 359 / 1960) loss: 1.234327
(Iteration 360 / 1960) loss: 1.275632
(Iteration 361 / 1960) loss: 1.193131
(Iteration 362 / 1960) loss: 1.279269
(Iteration 363 / 1960) loss: 1.256672
(Iteration 364 / 1960) loss: 1.191595
(Iteration 365 / 1960) loss: 1.189030
(Iteration 366 / 1960) loss: 1.208358
(Iteration 367 / 1960) loss: 1.160067
(Iteration 368 / 1960) loss: 1.209630
(Iteration 369 / 1960) loss: 1.284986
(Iteration 370 / 1960) loss: 1.254717
(Iteration 371 / 1960) loss: 1.225981
(Iteration 372 / 1960) loss: 1.135024
(Iteration 373 / 1960) loss: 1.301225
(Iteration 374 / 1960) loss: 1.228468
(Iteration 375 / 1960) loss: 1.268487
(Iteration 376 / 1960) loss: 1.192865
(Iteration 377 / 1960) loss: 1.287887
(Iteration 378 / 1960) loss: 1.246220
(Iteration 379 / 1960) loss: 1.210576
(Iteration 380 / 1960) loss: 1.323272
(Iteration 381 / 1960) loss: 1.305621
(Iteration 382 / 1960) loss: 1.208296
(Iteration 383 / 1960) loss: 1.221347
(Iteration 384 / 1960) loss: 1.205301
(Iteration 385 / 1960) loss: 1.175472
(Iteration 386 / 1960) loss: 1.195498
(Iteration 387 / 1960) loss: 1.165331
(Iteration 388 / 1960) loss: 1.199360
(Iteration 389 / 1960) loss: 1.184669
(Iteration 390 / 1960) loss: 1.223175
(Iteration 391 / 1960) loss: 1.252050
(Iteration 392 / 1960) loss: 1.249400
(Epoch 4 / 20) train acc: 0.641000; val_acc: 0.582000
(Iteration 393 / 1960) loss: 1.275563
(Iteration 394 / 1960) loss: 1.277089
(Iteration 395 / 1960) loss: 1.166340
(Iteration 396 / 1960) loss: 1.106129
(Iteration 397 / 1960) loss: 1.305783
(Iteration 398 / 1960) loss: 1.217910
(Iteration 399 / 1960) loss: 1.165010
(Iteration 400 / 1960) loss: 1.245411
(Iteration 401 / 1960) loss: 1.160872
(Iteration 402 / 1960) loss: 1.159988
(Iteration 403 / 1960) loss: 1.334059
(Iteration 404 / 1960) loss: 1.274649
(Iteration 405 / 1960) loss: 1.362636
(Iteration 406 / 1960) loss: 1.278265
(Iteration 407 / 1960) loss: 1.253624
(Iteration 408 / 1960) loss: 1.177305
```

(Iteration 409 / 1960)	loss: 1.154916
(Iteration 410 / 1960)	loss: 1.177574
(Iteration 411 / 1960)	loss: 1.190610
(Iteration 412 / 1960)	loss: 1.166158
(Iteration 413 / 1960)	loss: 1.131403
(Iteration 414 / 1960)	loss: 1.206104
(Iteration 415 / 1960)	loss: 1.233479
(Iteration 416 / 1960)	loss: 1.222900
(Iteration 417 / 1960)	loss: 1.143752
(Iteration 418 / 1960)	loss: 1.174865
(Iteration 419 / 1960)	loss: 1.157255
(Iteration 420 / 1960)	loss: 1.177431
(Iteration 421 / 1960)	loss: 1.172918
(Iteration 422 / 1960)	loss: 1.139912
(Iteration 423 / 1960)	loss: 1.218439
(Iteration 424 / 1960)	loss: 1.224836
(Iteration 425 / 1960)	loss: 1.214973
(Iteration 426 / 1960)	loss: 1.084563
(Iteration 427 / 1960)	loss: 1.237829
(Iteration 428 / 1960)	loss: 1.151373
(Iteration 429 / 1960)	loss: 1.233995
(Iteration 430 / 1960)	loss: 1.182605
(Iteration 431 / 1960)	loss: 1.204417
(Iteration 432 / 1960)	loss: 1.177185
(Iteration 433 / 1960)	loss: 1.314953
(Iteration 434 / 1960)	loss: 1.212040
(Iteration 435 / 1960)	loss: 1.108720
(Iteration 436 / 1960)	loss: 1.296088
(Iteration 437 / 1960)	loss: 1.186346
(Iteration 438 / 1960)	loss: 1.170571
(Iteration 439 / 1960)	loss: 1.195957
(Iteration 440 / 1960)	loss: 1.106785
(Iteration 441 / 1960)	loss: 1.170406
(Iteration 442 / 1960)	loss: 1.149499
(Iteration 443 / 1960)	loss: 1.131150
(Iteration 444 / 1960)	loss: 1.255600
(Iteration 445 / 1960)	loss: 1.187329
(Iteration 446 / 1960)	loss: 1.150247
(Iteration 447 / 1960)	loss: 1.212108
(Iteration 448 / 1960)	loss: 1.206954
(Iteration 449 / 1960)	loss: 1.284515
(Iteration 450 / 1960)	loss: 1.173900
(Iteration 451 / 1960)	loss: 1.182931
(Iteration 452 / 1960)	loss: 1.229125
(Iteration 453 / 1960)	loss: 1.218586
(Iteration 454 / 1960)	loss: 1.131264
(Iteration 455 / 1960)	loss: 1.146587
(Iteration 456 / 1960)	loss: 1.095432
(Iteration 457 / 1960)	loss: 1.170648
(Iteration 458 / 1960)	loss: 1.125780
(Iteration 459 / 1960)	loss: 1.139272
(Iteration 460 / 1960)	loss: 1.113035
(Iteration 461 / 1960)	loss: 1.209321
(Iteration 462 / 1960)	loss: 1.199324
(Iteration 463 / 1960)	loss: 1.188383
(Iteration 464 / 1960)	loss: 1.307411
(Iteration 465 / 1960)	loss: 1.166276


```
(Iteration 466 / 1960) loss: 1.268051
(Iteration 467 / 1960) loss: 1.118144
(Iteration 468 / 1960) loss: 1.146240
(Iteration 469 / 1960) loss: 1.090068
(Iteration 470 / 1960) loss: 1.160244
(Iteration 471 / 1960) loss: 1.171434
(Iteration 472 / 1960) loss: 1.233182
(Iteration 473 / 1960) loss: 1.176643
(Iteration 474 / 1960) loss: 1.210008
(Iteration 475 / 1960) loss: 1.175728
(Iteration 476 / 1960) loss: 1.156809
(Iteration 477 / 1960) loss: 1.155816
(Iteration 478 / 1960) loss: 1.139871
(Iteration 479 / 1960) loss: 1.140959
(Iteration 480 / 1960) loss: 1.113259
(Iteration 481 / 1960) loss: 1.132460
(Iteration 482 / 1960) loss: 1.068985
(Iteration 483 / 1960) loss: 1.143768
(Iteration 484 / 1960) loss: 1.162180
(Iteration 485 / 1960) loss: 1.117656
(Iteration 486 / 1960) loss: 1.205023
(Iteration 487 / 1960) loss: 1.101107
(Iteration 488 / 1960) loss: 1.093846
(Iteration 489 / 1960) loss: 1.152653
(Iteration 490 / 1960) loss: 1.072900
(Epoch 5 / 20) train acc: 0.711000; val_acc: 0.608000
(Iteration 491 / 1960) loss: 1.206960
(Iteration 492 / 1960) loss: 1.046964
(Iteration 493 / 1960) loss: 1.121059
(Iteration 494 / 1960) loss: 1.098119
(Iteration 495 / 1960) loss: 1.131381
(Iteration 496 / 1960) loss: 1.082017
(Iteration 497 / 1960) loss: 1.170042
(Iteration 498 / 1960) loss: 1.074863
(Iteration 499 / 1960) loss: 1.141657
(Iteration 500 / 1960) loss: 1.113048
(Iteration 501 / 1960) loss: 1.165978
(Iteration 502 / 1960) loss: 1.157064
(Iteration 503 / 1960) loss: 1.098214
(Iteration 504 / 1960) loss: 1.056568
(Iteration 505 / 1960) loss: 1.129871
(Iteration 506 / 1960) loss: 1.153898
(Iteration 507 / 1960) loss: 1.013504
(Iteration 508 / 1960) loss: 1.202377
(Iteration 509 / 1960) loss: 1.192519
(Iteration 510 / 1960) loss: 1.029274
(Iteration 511 / 1960) loss: 1.107948
(Iteration 512 / 1960) loss: 1.155326
(Iteration 513 / 1960) loss: 1.150360
(Iteration 514 / 1960) loss: 1.143704
(Iteration 515 / 1960) loss: 1.072230
(Iteration 516 / 1960) loss: 1.128015
(Iteration 517 / 1960) loss: 1.102275
(Iteration 518 / 1960) loss: 1.142337
(Iteration 519 / 1960) loss: 1.102406
(Iteration 520 / 1960) loss: 1.067618
(Iteration 521 / 1960) loss: 1.029053
```

(Iteration 522 / 1960)	loss: 1.055566
(Iteration 523 / 1960)	loss: 1.058885
(Iteration 524 / 1960)	loss: 1.149429
(Iteration 525 / 1960)	loss: 1.038422
(Iteration 526 / 1960)	loss: 1.090745
(Iteration 527 / 1960)	loss: 1.043819
(Iteration 528 / 1960)	loss: 1.156871
(Iteration 529 / 1960)	loss: 1.132671
(Iteration 530 / 1960)	loss: 1.073038
(Iteration 531 / 1960)	loss: 1.058899
(Iteration 532 / 1960)	loss: 1.176025
(Iteration 533 / 1960)	loss: 1.065304
(Iteration 534 / 1960)	loss: 1.181873
(Iteration 535 / 1960)	loss: 1.164481
(Iteration 536 / 1960)	loss: 1.096735
(Iteration 537 / 1960)	loss: 1.216968
(Iteration 538 / 1960)	loss: 1.071864
(Iteration 539 / 1960)	loss: 1.089276
(Iteration 540 / 1960)	loss: 1.248330
(Iteration 541 / 1960)	loss: 1.037796
(Iteration 542 / 1960)	loss: 1.108008
(Iteration 543 / 1960)	loss: 1.056439
(Iteration 544 / 1960)	loss: 1.105645
(Iteration 545 / 1960)	loss: 1.051639
(Iteration 546 / 1960)	loss: 1.076130
(Iteration 547 / 1960)	loss: 1.069958
(Iteration 548 / 1960)	loss: 1.060296
(Iteration 549 / 1960)	loss: 1.002601
(Iteration 550 / 1960)	loss: 1.020808
(Iteration 551 / 1960)	loss: 1.058107
(Iteration 552 / 1960)	loss: 1.009333
(Iteration 553 / 1960)	loss: 1.075036
(Iteration 554 / 1960)	loss: 1.030835
(Iteration 555 / 1960)	loss: 1.065214
(Iteration 556 / 1960)	loss: 1.136553
(Iteration 557 / 1960)	loss: 1.063940
(Iteration 558 / 1960)	loss: 1.023964
(Iteration 559 / 1960)	loss: 1.060176
(Iteration 560 / 1960)	loss: 1.109479
(Iteration 561 / 1960)	loss: 1.125190
(Iteration 562 / 1960)	loss: 1.088228
(Iteration 563 / 1960)	loss: 1.023498
(Iteration 564 / 1960)	loss: 1.091070
(Iteration 565 / 1960)	loss: 1.062764
(Iteration 566 / 1960)	loss: 1.003654
(Iteration 567 / 1960)	loss: 1.133328
(Iteration 568 / 1960)	loss: 1.010293
(Iteration 569 / 1960)	loss: 1.068877
(Iteration 570 / 1960)	loss: 1.113059
(Iteration 571 / 1960)	loss: 1.138209
(Iteration 572 / 1960)	loss: 1.016691
(Iteration 573 / 1960)	loss: 1.062738
(Iteration 574 / 1960)	loss: 1.177924
(Iteration 575 / 1960)	loss: 1.158074
(Iteration 576 / 1960)	loss: 1.195214
(Iteration 577 / 1960)	loss: 1.075904
(Iteration 578 / 1960)	loss: 1.100019

(Iteration 579 / 1960) loss: 1.102972
(Iteration 580 / 1960) loss: 1.061835
(Iteration 581 / 1960) loss: 1.095352
(Iteration 582 / 1960) loss: 1.070432
(Iteration 583 / 1960) loss: 1.044802
(Iteration 584 / 1960) loss: 1.182380
(Iteration 585 / 1960) loss: 1.171878
(Iteration 586 / 1960) loss: 0.980198
(Iteration 587 / 1960) loss: 0.997123
(Iteration 588 / 1960) loss: 1.083686
(Epoch 6 / 20) train acc: 0.731000; val_acc: 0.629000
(Iteration 589 / 1960) loss: 1.105487
(Iteration 590 / 1960) loss: 1.085008
(Iteration 591 / 1960) loss: 1.082960
(Iteration 592 / 1960) loss: 1.025160
(Iteration 593 / 1960) loss: 1.085742
(Iteration 594 / 1960) loss: 1.020690
(Iteration 595 / 1960) loss: 1.043723
(Iteration 596 / 1960) loss: 1.074316
(Iteration 597 / 1960) loss: 1.015401
(Iteration 598 / 1960) loss: 1.071931
(Iteration 599 / 1960) loss: 1.074104
(Iteration 600 / 1960) loss: 1.056239
(Iteration 601 / 1960) loss: 1.095043
(Iteration 602 / 1960) loss: 1.042159
(Iteration 603 / 1960) loss: 0.945953
(Iteration 604 / 1960) loss: 1.108429
(Iteration 605 / 1960) loss: 0.984939
(Iteration 606 / 1960) loss: 1.052311
(Iteration 607 / 1960) loss: 1.057996
(Iteration 608 / 1960) loss: 1.041506
(Iteration 609 / 1960) loss: 1.057919
(Iteration 610 / 1960) loss: 1.099802
(Iteration 611 / 1960) loss: 1.111453
(Iteration 612 / 1960) loss: 1.010703
(Iteration 613 / 1960) loss: 1.015927
(Iteration 614 / 1960) loss: 0.979765
(Iteration 615 / 1960) loss: 0.993874
(Iteration 616 / 1960) loss: 1.008439
(Iteration 617 / 1960) loss: 1.039568
(Iteration 618 / 1960) loss: 1.060415
(Iteration 619 / 1960) loss: 1.042920
(Iteration 620 / 1960) loss: 1.109177
(Iteration 621 / 1960) loss: 0.989688
(Iteration 622 / 1960) loss: 1.088570
(Iteration 623 / 1960) loss: 1.135076
(Iteration 624 / 1960) loss: 1.096167
(Iteration 625 / 1960) loss: 1.023673
(Iteration 626 / 1960) loss: 1.092261
(Iteration 627 / 1960) loss: 1.148437
(Iteration 628 / 1960) loss: 1.094455
(Iteration 629 / 1960) loss: 0.996288
(Iteration 630 / 1960) loss: 1.087909
(Iteration 631 / 1960) loss: 1.045902
(Iteration 632 / 1960) loss: 1.131463
(Iteration 633 / 1960) loss: 0.973617
(Iteration 634 / 1960) loss: 1.088032

```
(Iteration 635 / 1960) loss: 1.060118
(Iteration 636 / 1960) loss: 1.023245
(Iteration 637 / 1960) loss: 0.955042
(Iteration 638 / 1960) loss: 1.002959
(Iteration 639 / 1960) loss: 1.122478
(Iteration 640 / 1960) loss: 1.092968
(Iteration 641 / 1960) loss: 1.069761
(Iteration 642 / 1960) loss: 1.089094
(Iteration 643 / 1960) loss: 1.095090
(Iteration 644 / 1960) loss: 0.993867
(Iteration 645 / 1960) loss: 1.120848
(Iteration 646 / 1960) loss: 1.076250
(Iteration 647 / 1960) loss: 1.016783
(Iteration 648 / 1960) loss: 1.200020
(Iteration 649 / 1960) loss: 1.045813
(Iteration 650 / 1960) loss: 1.033173
(Iteration 651 / 1960) loss: 1.079458
(Iteration 652 / 1960) loss: 1.106126
(Iteration 653 / 1960) loss: 1.022450
(Iteration 654 / 1960) loss: 1.064883
(Iteration 655 / 1960) loss: 1.162483
(Iteration 656 / 1960) loss: 1.125261
(Iteration 657 / 1960) loss: 1.102560
(Iteration 658 / 1960) loss: 1.034036
(Iteration 659 / 1960) loss: 0.977655
(Iteration 660 / 1960) loss: 1.088136
(Iteration 661 / 1960) loss: 0.987187
(Iteration 662 / 1960) loss: 1.067578
(Iteration 663 / 1960) loss: 1.067584
(Iteration 664 / 1960) loss: 1.166006
(Iteration 665 / 1960) loss: 1.112140
(Iteration 666 / 1960) loss: 1.052624
(Iteration 667 / 1960) loss: 1.141982
(Iteration 668 / 1960) loss: 0.969895
(Iteration 669 / 1960) loss: 1.011788
(Iteration 670 / 1960) loss: 1.092740
(Iteration 671 / 1960) loss: 1.030817
(Iteration 672 / 1960) loss: 0.963717
(Iteration 673 / 1960) loss: 0.959759
(Iteration 674 / 1960) loss: 1.000428
(Iteration 675 / 1960) loss: 0.984000
(Iteration 676 / 1960) loss: 1.040215
(Iteration 677 / 1960) loss: 1.081308
(Iteration 678 / 1960) loss: 1.094185
(Iteration 679 / 1960) loss: 1.023256
(Iteration 680 / 1960) loss: 0.978403
(Iteration 681 / 1960) loss: 1.017512
(Iteration 682 / 1960) loss: 1.082726
(Iteration 683 / 1960) loss: 1.101599
(Iteration 684 / 1960) loss: 1.038600
(Iteration 685 / 1960) loss: 1.057108
(Iteration 686 / 1960) loss: 1.023659
(Epoch 7 / 20) train acc: 0.740000; val_acc: 0.661000
(Iteration 687 / 1960) loss: 0.980124
(Iteration 688 / 1960) loss: 0.945500
(Iteration 689 / 1960) loss: 1.078681
(Iteration 690 / 1960) loss: 0.954989
```

(Iteration 691 / 1960)	loss: 0.998909
(Iteration 692 / 1960)	loss: 0.999771
(Iteration 693 / 1960)	loss: 1.032277
(Iteration 694 / 1960)	loss: 1.008141
(Iteration 695 / 1960)	loss: 1.035668
(Iteration 696 / 1960)	loss: 1.069425
(Iteration 697 / 1960)	loss: 0.992656
(Iteration 698 / 1960)	loss: 1.066951
(Iteration 699 / 1960)	loss: 1.143246
(Iteration 700 / 1960)	loss: 1.029997
(Iteration 701 / 1960)	loss: 1.057053
(Iteration 702 / 1960)	loss: 1.067131
(Iteration 703 / 1960)	loss: 1.048024
(Iteration 704 / 1960)	loss: 0.974636
(Iteration 705 / 1960)	loss: 1.046118
(Iteration 706 / 1960)	loss: 1.003157
(Iteration 707 / 1960)	loss: 0.976072
(Iteration 708 / 1960)	loss: 1.061559
(Iteration 709 / 1960)	loss: 0.975450
(Iteration 710 / 1960)	loss: 1.013566
(Iteration 711 / 1960)	loss: 1.010462
(Iteration 712 / 1960)	loss: 0.871600
(Iteration 713 / 1960)	loss: 1.008118
(Iteration 714 / 1960)	loss: 1.044039
(Iteration 715 / 1960)	loss: 1.016194
(Iteration 716 / 1960)	loss: 0.898704
(Iteration 717 / 1960)	loss: 0.954149
(Iteration 718 / 1960)	loss: 1.067985
(Iteration 719 / 1960)	loss: 1.042098
(Iteration 720 / 1960)	loss: 0.971655
(Iteration 721 / 1960)	loss: 1.035870
(Iteration 722 / 1960)	loss: 1.000561
(Iteration 723 / 1960)	loss: 1.011613
(Iteration 724 / 1960)	loss: 0.965840
(Iteration 725 / 1960)	loss: 1.015596
(Iteration 726 / 1960)	loss: 0.983931
(Iteration 727 / 1960)	loss: 0.936562
(Iteration 728 / 1960)	loss: 1.017746
(Iteration 729 / 1960)	loss: 1.008738
(Iteration 730 / 1960)	loss: 0.980646
(Iteration 731 / 1960)	loss: 1.081706
(Iteration 732 / 1960)	loss: 0.959752
(Iteration 733 / 1960)	loss: 1.069284
(Iteration 734 / 1960)	loss: 0.969742
(Iteration 735 / 1960)	loss: 1.047777
(Iteration 736 / 1960)	loss: 0.934751
(Iteration 737 / 1960)	loss: 1.034189
(Iteration 738 / 1960)	loss: 1.021987
(Iteration 739 / 1960)	loss: 0.937978
(Iteration 740 / 1960)	loss: 0.918595
(Iteration 741 / 1960)	loss: 0.947706
(Iteration 742 / 1960)	loss: 1.060869
(Iteration 743 / 1960)	loss: 0.949495
(Iteration 744 / 1960)	loss: 0.954686
(Iteration 745 / 1960)	loss: 1.047198
(Iteration 746 / 1960)	loss: 1.039581
(Iteration 747 / 1960)	loss: 0.985952

```
(Iteration 748 / 1960) loss: 0.886510
(Iteration 749 / 1960) loss: 1.039093
(Iteration 750 / 1960) loss: 0.995172
(Iteration 751 / 1960) loss: 0.963759
(Iteration 752 / 1960) loss: 1.046218
(Iteration 753 / 1960) loss: 0.982973
(Iteration 754 / 1960) loss: 0.958262
(Iteration 755 / 1960) loss: 0.989272
(Iteration 756 / 1960) loss: 0.929278
(Iteration 757 / 1960) loss: 0.951263
(Iteration 758 / 1960) loss: 1.052806
(Iteration 759 / 1960) loss: 1.019299
(Iteration 760 / 1960) loss: 1.002960
(Iteration 761 / 1960) loss: 1.021357
(Iteration 762 / 1960) loss: 1.037942
(Iteration 763 / 1960) loss: 1.003835
(Iteration 764 / 1960) loss: 0.966497
(Iteration 765 / 1960) loss: 0.927913
(Iteration 766 / 1960) loss: 0.967427
(Iteration 767 / 1960) loss: 0.885969
(Iteration 768 / 1960) loss: 0.991151
(Iteration 769 / 1960) loss: 0.905246
(Iteration 770 / 1960) loss: 1.004756
(Iteration 771 / 1960) loss: 0.981178
(Iteration 772 / 1960) loss: 0.983467
(Iteration 773 / 1960) loss: 0.990601
(Iteration 774 / 1960) loss: 0.858445
(Iteration 775 / 1960) loss: 0.977288
(Iteration 776 / 1960) loss: 0.927716
(Iteration 777 / 1960) loss: 0.907865
(Iteration 778 / 1960) loss: 0.948854
(Iteration 779 / 1960) loss: 0.928119
(Iteration 780 / 1960) loss: 1.031947
(Iteration 781 / 1960) loss: 0.916854
(Iteration 782 / 1960) loss: 0.991048
(Iteration 783 / 1960) loss: 0.914151
(Iteration 784 / 1960) loss: 0.972864
(Epoch 8 / 20) train acc: 0.715000; val_acc: 0.626000
(Iteration 785 / 1960) loss: 0.994806
(Iteration 786 / 1960) loss: 0.926439
(Iteration 787 / 1960) loss: 0.955513
(Iteration 788 / 1960) loss: 1.043645
(Iteration 789 / 1960) loss: 0.950009
(Iteration 790 / 1960) loss: 0.922774
(Iteration 791 / 1960) loss: 1.010784
(Iteration 792 / 1960) loss: 0.946571
(Iteration 793 / 1960) loss: 0.942611
(Iteration 794 / 1960) loss: 0.945035
(Iteration 795 / 1960) loss: 0.957011
(Iteration 796 / 1960) loss: 0.938585
(Iteration 797 / 1960) loss: 0.950407
(Iteration 798 / 1960) loss: 0.990769
(Iteration 799 / 1960) loss: 0.891829
(Iteration 800 / 1960) loss: 0.901766
(Iteration 801 / 1960) loss: 0.976105
(Iteration 802 / 1960) loss: 1.008589
(Iteration 803 / 1960) loss: 1.023258
```

(Iteration 804 / 1960)	loss: 0.961734
(Iteration 805 / 1960)	loss: 1.022924
(Iteration 806 / 1960)	loss: 0.949404
(Iteration 807 / 1960)	loss: 0.937876
(Iteration 808 / 1960)	loss: 0.924420
(Iteration 809 / 1960)	loss: 1.001674
(Iteration 810 / 1960)	loss: 0.928206
(Iteration 811 / 1960)	loss: 1.041732
(Iteration 812 / 1960)	loss: 1.008256
(Iteration 813 / 1960)	loss: 0.879201
(Iteration 814 / 1960)	loss: 0.924865
(Iteration 815 / 1960)	loss: 0.971744
(Iteration 816 / 1960)	loss: 0.968948
(Iteration 817 / 1960)	loss: 1.036146
(Iteration 818 / 1960)	loss: 0.994697
(Iteration 819 / 1960)	loss: 0.830217
(Iteration 820 / 1960)	loss: 0.935518
(Iteration 821 / 1960)	loss: 1.086457
(Iteration 822 / 1960)	loss: 0.901337
(Iteration 823 / 1960)	loss: 0.971786
(Iteration 824 / 1960)	loss: 1.007351
(Iteration 825 / 1960)	loss: 0.960865
(Iteration 826 / 1960)	loss: 1.015437
(Iteration 827 / 1960)	loss: 0.925988
(Iteration 828 / 1960)	loss: 0.992310
(Iteration 829 / 1960)	loss: 0.919648
(Iteration 830 / 1960)	loss: 0.925868
(Iteration 831 / 1960)	loss: 0.847005
(Iteration 832 / 1960)	loss: 0.914660
(Iteration 833 / 1960)	loss: 1.009910
(Iteration 834 / 1960)	loss: 0.964191
(Iteration 835 / 1960)	loss: 0.923890
(Iteration 836 / 1960)	loss: 0.943962
(Iteration 837 / 1960)	loss: 0.990956
(Iteration 838 / 1960)	loss: 0.978520
(Iteration 839 / 1960)	loss: 0.908868
(Iteration 840 / 1960)	loss: 0.984400
(Iteration 841 / 1960)	loss: 0.931814
(Iteration 842 / 1960)	loss: 0.929026
(Iteration 843 / 1960)	loss: 0.925687
(Iteration 844 / 1960)	loss: 0.968651
(Iteration 845 / 1960)	loss: 0.990598
(Iteration 846 / 1960)	loss: 0.891473
(Iteration 847 / 1960)	loss: 1.003045
(Iteration 848 / 1960)	loss: 0.899779
(Iteration 849 / 1960)	loss: 1.076496
(Iteration 850 / 1960)	loss: 0.870864
(Iteration 851 / 1960)	loss: 0.969330
(Iteration 852 / 1960)	loss: 0.964397
(Iteration 853 / 1960)	loss: 0.968212
(Iteration 854 / 1960)	loss: 1.028611
(Iteration 855 / 1960)	loss: 0.969761
(Iteration 856 / 1960)	loss: 1.018058
(Iteration 857 / 1960)	loss: 1.082443
(Iteration 858 / 1960)	loss: 1.029729
(Iteration 859 / 1960)	loss: 0.898576
(Iteration 860 / 1960)	loss: 1.047539

```
(Iteration 861 / 1960) loss: 0.926091
(Iteration 862 / 1960) loss: 1.012229
(Iteration 863 / 1960) loss: 0.962936
(Iteration 864 / 1960) loss: 0.916402
(Iteration 865 / 1960) loss: 0.978747
(Iteration 866 / 1960) loss: 0.982574
(Iteration 867 / 1960) loss: 0.894701
(Iteration 868 / 1960) loss: 0.964914
(Iteration 869 / 1960) loss: 0.953197
(Iteration 870 / 1960) loss: 0.958114
(Iteration 871 / 1960) loss: 0.890796
(Iteration 872 / 1960) loss: 0.930689
(Iteration 873 / 1960) loss: 0.889322
(Iteration 874 / 1960) loss: 1.023274
(Iteration 875 / 1960) loss: 0.995970
(Iteration 876 / 1960) loss: 0.881085
(Iteration 877 / 1960) loss: 0.920005
(Iteration 878 / 1960) loss: 0.899613
(Iteration 879 / 1960) loss: 0.938083
(Iteration 880 / 1960) loss: 0.956628
(Iteration 881 / 1960) loss: 0.934709
(Iteration 882 / 1960) loss: 0.921354
(Epoch 9 / 20) train acc: 0.757000; val_acc: 0.663000
(Iteration 883 / 1960) loss: 0.840649
(Iteration 884 / 1960) loss: 0.976873
(Iteration 885 / 1960) loss: 0.892653
(Iteration 886 / 1960) loss: 0.883909
(Iteration 887 / 1960) loss: 0.838967
(Iteration 888 / 1960) loss: 0.924668
(Iteration 889 / 1960) loss: 0.919403
(Iteration 890 / 1960) loss: 0.887872
(Iteration 891 / 1960) loss: 0.860892
(Iteration 892 / 1960) loss: 0.869693
(Iteration 893 / 1960) loss: 0.843793
(Iteration 894 / 1960) loss: 0.910228
(Iteration 895 / 1960) loss: 0.964715
(Iteration 896 / 1960) loss: 0.869991
(Iteration 897 / 1960) loss: 0.933294
(Iteration 898 / 1960) loss: 0.930019
(Iteration 899 / 1960) loss: 0.873695
(Iteration 900 / 1960) loss: 0.984831
(Iteration 901 / 1960) loss: 0.923538
(Iteration 902 / 1960) loss: 0.976797
(Iteration 903 / 1960) loss: 0.922247
(Iteration 904 / 1960) loss: 0.923313
(Iteration 905 / 1960) loss: 0.918265
(Iteration 906 / 1960) loss: 0.920554
(Iteration 907 / 1960) loss: 0.854225
(Iteration 908 / 1960) loss: 0.880953
(Iteration 909 / 1960) loss: 0.832276
(Iteration 910 / 1960) loss: 0.823222
(Iteration 911 / 1960) loss: 0.972283
(Iteration 912 / 1960) loss: 0.922825
(Iteration 913 / 1960) loss: 0.880972
(Iteration 914 / 1960) loss: 0.898618
(Iteration 915 / 1960) loss: 0.900149
(Iteration 916 / 1960) loss: 0.907088
```


(Iteration 917 / 1960)	loss: 0.890361
(Iteration 918 / 1960)	loss: 0.991791
(Iteration 919 / 1960)	loss: 0.957492
(Iteration 920 / 1960)	loss: 0.850665
(Iteration 921 / 1960)	loss: 0.914964
(Iteration 922 / 1960)	loss: 0.999617
(Iteration 923 / 1960)	loss: 0.922603
(Iteration 924 / 1960)	loss: 0.846517
(Iteration 925 / 1960)	loss: 0.964169
(Iteration 926 / 1960)	loss: 0.932077
(Iteration 927 / 1960)	loss: 0.831494
(Iteration 928 / 1960)	loss: 0.857945
(Iteration 929 / 1960)	loss: 0.973495
(Iteration 930 / 1960)	loss: 0.910534
(Iteration 931 / 1960)	loss: 0.869657
(Iteration 932 / 1960)	loss: 0.846595
(Iteration 933 / 1960)	loss: 0.974776
(Iteration 934 / 1960)	loss: 0.874458
(Iteration 935 / 1960)	loss: 0.841683
(Iteration 936 / 1960)	loss: 0.869939
(Iteration 937 / 1960)	loss: 0.844365
(Iteration 938 / 1960)	loss: 0.936231
(Iteration 939 / 1960)	loss: 0.855485
(Iteration 940 / 1960)	loss: 1.021888
(Iteration 941 / 1960)	loss: 0.836070
(Iteration 942 / 1960)	loss: 0.927155
(Iteration 943 / 1960)	loss: 0.949799
(Iteration 944 / 1960)	loss: 0.866261
(Iteration 945 / 1960)	loss: 0.884948
(Iteration 946 / 1960)	loss: 0.922765
(Iteration 947 / 1960)	loss: 0.920997
(Iteration 948 / 1960)	loss: 0.839794
(Iteration 949 / 1960)	loss: 0.938477
(Iteration 950 / 1960)	loss: 0.921336
(Iteration 951 / 1960)	loss: 0.958452
(Iteration 952 / 1960)	loss: 0.911705
(Iteration 953 / 1960)	loss: 0.907667
(Iteration 954 / 1960)	loss: 0.928543
(Iteration 955 / 1960)	loss: 0.907044
(Iteration 956 / 1960)	loss: 0.894357
(Iteration 957 / 1960)	loss: 0.878338
(Iteration 958 / 1960)	loss: 0.928859
(Iteration 959 / 1960)	loss: 0.980847
(Iteration 960 / 1960)	loss: 0.914376
(Iteration 961 / 1960)	loss: 0.935544
(Iteration 962 / 1960)	loss: 0.882842
(Iteration 963 / 1960)	loss: 0.974036
(Iteration 964 / 1960)	loss: 0.835443
(Iteration 965 / 1960)	loss: 0.887539
(Iteration 966 / 1960)	loss: 0.988907
(Iteration 967 / 1960)	loss: 0.910143
(Iteration 968 / 1960)	loss: 0.895512
(Iteration 969 / 1960)	loss: 0.966832
(Iteration 970 / 1960)	loss: 0.872864
(Iteration 971 / 1960)	loss: 0.976849
(Iteration 972 / 1960)	loss: 0.904292
(Iteration 973 / 1960)	loss: 0.909927

(Iteration 974 / 1960) loss: 0.841184
(Iteration 975 / 1960) loss: 0.940086
(Iteration 976 / 1960) loss: 0.948542
(Iteration 977 / 1960) loss: 0.927738
(Iteration 978 / 1960) loss: 0.897737
(Iteration 979 / 1960) loss: 0.920774
(Iteration 980 / 1960) loss: 0.904144
(Epoch 10 / 20) train acc: 0.756000; val_acc: 0.637000
(Iteration 981 / 1960) loss: 0.948811
(Iteration 982 / 1960) loss: 0.988839
(Iteration 983 / 1960) loss: 0.934907
(Iteration 984 / 1960) loss: 0.938880
(Iteration 985 / 1960) loss: 0.852843
(Iteration 986 / 1960) loss: 0.941859
(Iteration 987 / 1960) loss: 0.908369
(Iteration 988 / 1960) loss: 0.873579
(Iteration 989 / 1960) loss: 0.985067
(Iteration 990 / 1960) loss: 0.912449
(Iteration 991 / 1960) loss: 0.891438
(Iteration 992 / 1960) loss: 0.943250
(Iteration 993 / 1960) loss: 0.923468
(Iteration 994 / 1960) loss: 0.828857
(Iteration 995 / 1960) loss: 0.846419
(Iteration 996 / 1960) loss: 0.968215
(Iteration 997 / 1960) loss: 0.890815
(Iteration 998 / 1960) loss: 0.918667
(Iteration 999 / 1960) loss: 0.929176
(Iteration 1000 / 1960) loss: 0.859195
(Iteration 1001 / 1960) loss: 0.947852
(Iteration 1002 / 1960) loss: 0.831059
(Iteration 1003 / 1960) loss: 0.864148
(Iteration 1004 / 1960) loss: 0.852625
(Iteration 1005 / 1960) loss: 0.864143
(Iteration 1006 / 1960) loss: 0.888255
(Iteration 1007 / 1960) loss: 0.954782
(Iteration 1008 / 1960) loss: 0.876854
(Iteration 1009 / 1960) loss: 0.834019
(Iteration 1010 / 1960) loss: 0.912594
(Iteration 1011 / 1960) loss: 0.806200
(Iteration 1012 / 1960) loss: 0.835862
(Iteration 1013 / 1960) loss: 0.879097
(Iteration 1014 / 1960) loss: 0.836343
(Iteration 1015 / 1960) loss: 0.902438
(Iteration 1016 / 1960) loss: 0.913302
(Iteration 1017 / 1960) loss: 0.938899
(Iteration 1018 / 1960) loss: 0.925228
(Iteration 1019 / 1960) loss: 0.873403
(Iteration 1020 / 1960) loss: 0.915067
(Iteration 1021 / 1960) loss: 0.948861
(Iteration 1022 / 1960) loss: 0.875682
(Iteration 1023 / 1960) loss: 0.862898
(Iteration 1024 / 1960) loss: 0.869146
(Iteration 1025 / 1960) loss: 0.846699
(Iteration 1026 / 1960) loss: 0.841958
(Iteration 1027 / 1960) loss: 0.902681
(Iteration 1028 / 1960) loss: 0.874448
(Iteration 1029 / 1960) loss: 0.798141

(Iteration 1030 / 1960) loss: 0.867458
(Iteration 1031 / 1960) loss: 0.891765
(Iteration 1032 / 1960) loss: 0.883317
(Iteration 1033 / 1960) loss: 0.897972
(Iteration 1034 / 1960) loss: 0.909787
(Iteration 1035 / 1960) loss: 0.821204
(Iteration 1036 / 1960) loss: 0.853015
(Iteration 1037 / 1960) loss: 0.862679
(Iteration 1038 / 1960) loss: 0.957125
(Iteration 1039 / 1960) loss: 0.898215
(Iteration 1040 / 1960) loss: 0.863521
(Iteration 1041 / 1960) loss: 0.831484
(Iteration 1042 / 1960) loss: 0.768342
(Iteration 1043 / 1960) loss: 0.869013
(Iteration 1044 / 1960) loss: 0.859801
(Iteration 1045 / 1960) loss: 0.842093
(Iteration 1046 / 1960) loss: 0.806257
(Iteration 1047 / 1960) loss: 0.927884
(Iteration 1048 / 1960) loss: 0.889262
(Iteration 1049 / 1960) loss: 0.915208
(Iteration 1050 / 1960) loss: 0.890457
(Iteration 1051 / 1960) loss: 0.915051
(Iteration 1052 / 1960) loss: 0.830852
(Iteration 1053 / 1960) loss: 0.878642
(Iteration 1054 / 1960) loss: 0.858194
(Iteration 1055 / 1960) loss: 0.902377
(Iteration 1056 / 1960) loss: 0.875298
(Iteration 1057 / 1960) loss: 0.892943
(Iteration 1058 / 1960) loss: 0.857874
(Iteration 1059 / 1960) loss: 0.927468
(Iteration 1060 / 1960) loss: 0.910976
(Iteration 1061 / 1960) loss: 0.988243
(Iteration 1062 / 1960) loss: 0.873762
(Iteration 1063 / 1960) loss: 0.877133
(Iteration 1064 / 1960) loss: 0.847426
(Iteration 1065 / 1960) loss: 0.983021
(Iteration 1066 / 1960) loss: 0.835803
(Iteration 1067 / 1960) loss: 0.879004
(Iteration 1068 / 1960) loss: 0.875924
(Iteration 1069 / 1960) loss: 0.873949
(Iteration 1070 / 1960) loss: 0.776492
(Iteration 1071 / 1960) loss: 0.968471
(Iteration 1072 / 1960) loss: 0.862557
(Iteration 1073 / 1960) loss: 0.841609
(Iteration 1074 / 1960) loss: 0.918812
(Iteration 1075 / 1960) loss: 0.811155
(Iteration 1076 / 1960) loss: 0.917559
(Iteration 1077 / 1960) loss: 0.799803
(Iteration 1078 / 1960) loss: 0.783457
(Epoch 11 / 20) train acc: 0.779000; val_acc: 0.656000
(Iteration 1079 / 1960) loss: 0.786387
(Iteration 1080 / 1960) loss: 0.869078
(Iteration 1081 / 1960) loss: 0.936926
(Iteration 1082 / 1960) loss: 0.927863
(Iteration 1083 / 1960) loss: 0.807175
(Iteration 1084 / 1960) loss: 0.915289
(Iteration 1085 / 1960) loss: 0.912239

(Iteration 1086 / 1960) loss: 0.787537
(Iteration 1087 / 1960) loss: 0.909889
(Iteration 1088 / 1960) loss: 0.815467
(Iteration 1089 / 1960) loss: 0.912177
(Iteration 1090 / 1960) loss: 0.814017
(Iteration 1091 / 1960) loss: 0.856726
(Iteration 1092 / 1960) loss: 0.882012
(Iteration 1093 / 1960) loss: 0.854181
(Iteration 1094 / 1960) loss: 0.878517
(Iteration 1095 / 1960) loss: 0.890108
(Iteration 1096 / 1960) loss: 0.872275
(Iteration 1097 / 1960) loss: 0.943946
(Iteration 1098 / 1960) loss: 0.778529
(Iteration 1099 / 1960) loss: 0.908809
(Iteration 1100 / 1960) loss: 0.759213
(Iteration 1101 / 1960) loss: 0.850638
(Iteration 1102 / 1960) loss: 0.917873
(Iteration 1103 / 1960) loss: 0.855221
(Iteration 1104 / 1960) loss: 0.891957
(Iteration 1105 / 1960) loss: 0.825542
(Iteration 1106 / 1960) loss: 0.904092
(Iteration 1107 / 1960) loss: 0.869561
(Iteration 1108 / 1960) loss: 0.851285
(Iteration 1109 / 1960) loss: 0.782219
(Iteration 1110 / 1960) loss: 0.883431
(Iteration 1111 / 1960) loss: 0.877805
(Iteration 1112 / 1960) loss: 0.867440
(Iteration 1113 / 1960) loss: 0.755699
(Iteration 1114 / 1960) loss: 0.879248
(Iteration 1115 / 1960) loss: 0.909587
(Iteration 1116 / 1960) loss: 0.817395
(Iteration 1117 / 1960) loss: 0.883223
(Iteration 1118 / 1960) loss: 0.861037
(Iteration 1119 / 1960) loss: 0.840195
(Iteration 1120 / 1960) loss: 0.901421
(Iteration 1121 / 1960) loss: 0.784727
(Iteration 1122 / 1960) loss: 0.919509
(Iteration 1123 / 1960) loss: 0.868023
(Iteration 1124 / 1960) loss: 0.852532
(Iteration 1125 / 1960) loss: 0.812966
(Iteration 1126 / 1960) loss: 0.800166
(Iteration 1127 / 1960) loss: 0.875493
(Iteration 1128 / 1960) loss: 0.902112
(Iteration 1129 / 1960) loss: 0.817904
(Iteration 1130 / 1960) loss: 0.913012
(Iteration 1131 / 1960) loss: 0.837177
(Iteration 1132 / 1960) loss: 0.766938
(Iteration 1133 / 1960) loss: 0.792163
(Iteration 1134 / 1960) loss: 0.875690
(Iteration 1135 / 1960) loss: 0.774266
(Iteration 1136 / 1960) loss: 0.850184
(Iteration 1137 / 1960) loss: 0.826087
(Iteration 1138 / 1960) loss: 0.790931
(Iteration 1139 / 1960) loss: 0.803421
(Iteration 1140 / 1960) loss: 0.783283
(Iteration 1141 / 1960) loss: 0.873174
(Iteration 1142 / 1960) loss: 0.811663

(Iteration 1143 / 1960) loss: 0.875639
(Iteration 1144 / 1960) loss: 0.846794
(Iteration 1145 / 1960) loss: 0.784196
(Iteration 1146 / 1960) loss: 0.745550
(Iteration 1147 / 1960) loss: 0.820771
(Iteration 1148 / 1960) loss: 0.754515
(Iteration 1149 / 1960) loss: 0.867397
(Iteration 1150 / 1960) loss: 0.874436
(Iteration 1151 / 1960) loss: 0.758544
(Iteration 1152 / 1960) loss: 0.772283
(Iteration 1153 / 1960) loss: 0.805181
(Iteration 1154 / 1960) loss: 0.759835
(Iteration 1155 / 1960) loss: 0.817818
(Iteration 1156 / 1960) loss: 0.740610
(Iteration 1157 / 1960) loss: 0.830154
(Iteration 1158 / 1960) loss: 0.744585
(Iteration 1159 / 1960) loss: 0.889892
(Iteration 1160 / 1960) loss: 0.793553
(Iteration 1161 / 1960) loss: 0.950921
(Iteration 1162 / 1960) loss: 0.794264
(Iteration 1163 / 1960) loss: 0.782919
(Iteration 1164 / 1960) loss: 0.822265
(Iteration 1165 / 1960) loss: 0.793122
(Iteration 1166 / 1960) loss: 0.781805
(Iteration 1167 / 1960) loss: 0.793377
(Iteration 1168 / 1960) loss: 0.797568
(Iteration 1169 / 1960) loss: 0.778405
(Iteration 1170 / 1960) loss: 0.892498
(Iteration 1171 / 1960) loss: 0.854975
(Iteration 1172 / 1960) loss: 0.765348
(Iteration 1173 / 1960) loss: 0.809195
(Iteration 1174 / 1960) loss: 0.825403
(Iteration 1175 / 1960) loss: 0.721922
(Iteration 1176 / 1960) loss: 0.777454
(Epoch 12 / 20) train acc: 0.807000; val_acc: 0.659000
(Iteration 1177 / 1960) loss: 0.828966
(Iteration 1178 / 1960) loss: 0.807991
(Iteration 1179 / 1960) loss: 0.792402
(Iteration 1180 / 1960) loss: 0.781812
(Iteration 1181 / 1960) loss: 0.818617
(Iteration 1182 / 1960) loss: 0.764742
(Iteration 1183 / 1960) loss: 0.824753
(Iteration 1184 / 1960) loss: 0.878347
(Iteration 1185 / 1960) loss: 0.764473
(Iteration 1186 / 1960) loss: 0.840444
(Iteration 1187 / 1960) loss: 0.856437
(Iteration 1188 / 1960) loss: 0.752248
(Iteration 1189 / 1960) loss: 0.828286
(Iteration 1190 / 1960) loss: 0.725257
(Iteration 1191 / 1960) loss: 0.745069
(Iteration 1192 / 1960) loss: 0.736307
(Iteration 1193 / 1960) loss: 0.749504
(Iteration 1194 / 1960) loss: 0.790491
(Iteration 1195 / 1960) loss: 0.849939
(Iteration 1196 / 1960) loss: 0.782961
(Iteration 1197 / 1960) loss: 0.804828
(Iteration 1198 / 1960) loss: 0.794407

(Iteration 1199 / 1960) loss: 0.784562
(Iteration 1200 / 1960) loss: 0.838723
(Iteration 1201 / 1960) loss: 0.747431
(Iteration 1202 / 1960) loss: 0.720817
(Iteration 1203 / 1960) loss: 0.800388
(Iteration 1204 / 1960) loss: 0.766536
(Iteration 1205 / 1960) loss: 0.821910
(Iteration 1206 / 1960) loss: 0.792761
(Iteration 1207 / 1960) loss: 0.775444
(Iteration 1208 / 1960) loss: 0.722958
(Iteration 1209 / 1960) loss: 0.722164
(Iteration 1210 / 1960) loss: 0.792704
(Iteration 1211 / 1960) loss: 0.817305
(Iteration 1212 / 1960) loss: 0.751436
(Iteration 1213 / 1960) loss: 0.750427
(Iteration 1214 / 1960) loss: 0.810245
(Iteration 1215 / 1960) loss: 0.752549
(Iteration 1216 / 1960) loss: 0.855240
(Iteration 1217 / 1960) loss: 0.794623
(Iteration 1218 / 1960) loss: 0.740681
(Iteration 1219 / 1960) loss: 0.836710
(Iteration 1220 / 1960) loss: 0.750560
(Iteration 1221 / 1960) loss: 0.731021
(Iteration 1222 / 1960) loss: 0.846866
(Iteration 1223 / 1960) loss: 0.794466
(Iteration 1224 / 1960) loss: 0.856790
(Iteration 1225 / 1960) loss: 0.796858
(Iteration 1226 / 1960) loss: 0.791951
(Iteration 1227 / 1960) loss: 0.770449
(Iteration 1228 / 1960) loss: 0.800652
(Iteration 1229 / 1960) loss: 0.804048
(Iteration 1230 / 1960) loss: 0.839547
(Iteration 1231 / 1960) loss: 0.767593
(Iteration 1232 / 1960) loss: 0.881357
(Iteration 1233 / 1960) loss: 0.774889
(Iteration 1234 / 1960) loss: 0.765572
(Iteration 1235 / 1960) loss: 0.745656
(Iteration 1236 / 1960) loss: 0.840509
(Iteration 1237 / 1960) loss: 0.832914
(Iteration 1238 / 1960) loss: 0.735222
(Iteration 1239 / 1960) loss: 0.754890
(Iteration 1240 / 1960) loss: 0.810685
(Iteration 1241 / 1960) loss: 0.757844
(Iteration 1242 / 1960) loss: 0.887208
(Iteration 1243 / 1960) loss: 0.820222
(Iteration 1244 / 1960) loss: 0.839504
(Iteration 1245 / 1960) loss: 0.872256
(Iteration 1246 / 1960) loss: 0.744877
(Iteration 1247 / 1960) loss: 0.782450
(Iteration 1248 / 1960) loss: 0.851243
(Iteration 1249 / 1960) loss: 0.838333
(Iteration 1250 / 1960) loss: 0.770599
(Iteration 1251 / 1960) loss: 0.777791
(Iteration 1252 / 1960) loss: 0.748782
(Iteration 1253 / 1960) loss: 0.817001
(Iteration 1254 / 1960) loss: 0.824134
(Iteration 1255 / 1960) loss: 0.722126

(Iteration 1256 / 1960) loss: 0.740285
(Iteration 1257 / 1960) loss: 0.828412
(Iteration 1258 / 1960) loss: 0.695989
(Iteration 1259 / 1960) loss: 0.844979
(Iteration 1260 / 1960) loss: 0.790760
(Iteration 1261 / 1960) loss: 0.803408
(Iteration 1262 / 1960) loss: 0.785954
(Iteration 1263 / 1960) loss: 0.758211
(Iteration 1264 / 1960) loss: 0.739070
(Iteration 1265 / 1960) loss: 0.903578
(Iteration 1266 / 1960) loss: 0.769388
(Iteration 1267 / 1960) loss: 0.780114
(Iteration 1268 / 1960) loss: 0.768914
(Iteration 1269 / 1960) loss: 0.763761
(Iteration 1270 / 1960) loss: 0.880180
(Iteration 1271 / 1960) loss: 0.850672
(Iteration 1272 / 1960) loss: 0.790095
(Iteration 1273 / 1960) loss: 0.782243
(Iteration 1274 / 1960) loss: 0.882084
(Epoch 13 / 20) train acc: 0.796000; val_acc: 0.669000
(Iteration 1275 / 1960) loss: 0.723038
(Iteration 1276 / 1960) loss: 0.779556
(Iteration 1277 / 1960) loss: 0.796796
(Iteration 1278 / 1960) loss: 0.701339
(Iteration 1279 / 1960) loss: 0.852958
(Iteration 1280 / 1960) loss: 0.761390
(Iteration 1281 / 1960) loss: 0.702144
(Iteration 1282 / 1960) loss: 0.706411
(Iteration 1283 / 1960) loss: 0.771733
(Iteration 1284 / 1960) loss: 0.757150
(Iteration 1285 / 1960) loss: 0.833492
(Iteration 1286 / 1960) loss: 0.749504
(Iteration 1287 / 1960) loss: 0.709573
(Iteration 1288 / 1960) loss: 0.737266
(Iteration 1289 / 1960) loss: 0.778980
(Iteration 1290 / 1960) loss: 0.710155
(Iteration 1291 / 1960) loss: 0.725927
(Iteration 1292 / 1960) loss: 0.870076
(Iteration 1293 / 1960) loss: 0.774415
(Iteration 1294 / 1960) loss: 0.723028
(Iteration 1295 / 1960) loss: 0.823266
(Iteration 1296 / 1960) loss: 0.706591
(Iteration 1297 / 1960) loss: 0.801743
(Iteration 1298 / 1960) loss: 0.818714
(Iteration 1299 / 1960) loss: 0.702525
(Iteration 1300 / 1960) loss: 0.804085
(Iteration 1301 / 1960) loss: 0.806759
(Iteration 1302 / 1960) loss: 0.767468
(Iteration 1303 / 1960) loss: 0.862753
(Iteration 1304 / 1960) loss: 0.718483
(Iteration 1305 / 1960) loss: 0.786275
(Iteration 1306 / 1960) loss: 0.813196
(Iteration 1307 / 1960) loss: 0.757265
(Iteration 1308 / 1960) loss: 0.889634
(Iteration 1309 / 1960) loss: 0.750188
(Iteration 1310 / 1960) loss: 0.805531
(Iteration 1311 / 1960) loss: 0.789479

(Iteration 1312 / 1960) loss: 0.696628
(Iteration 1313 / 1960) loss: 0.687092
(Iteration 1314 / 1960) loss: 0.760899
(Iteration 1315 / 1960) loss: 0.817235
(Iteration 1316 / 1960) loss: 0.805957
(Iteration 1317 / 1960) loss: 0.773363
(Iteration 1318 / 1960) loss: 0.751899
(Iteration 1319 / 1960) loss: 0.722917
(Iteration 1320 / 1960) loss: 0.719071
(Iteration 1321 / 1960) loss: 0.788236
(Iteration 1322 / 1960) loss: 0.775363
(Iteration 1323 / 1960) loss: 0.797024
(Iteration 1324 / 1960) loss: 0.760920
(Iteration 1325 / 1960) loss: 0.732225
(Iteration 1326 / 1960) loss: 0.790000
(Iteration 1327 / 1960) loss: 0.798031
(Iteration 1328 / 1960) loss: 0.770577
(Iteration 1329 / 1960) loss: 0.736136
(Iteration 1330 / 1960) loss: 0.746132
(Iteration 1331 / 1960) loss: 0.764873
(Iteration 1332 / 1960) loss: 0.736505
(Iteration 1333 / 1960) loss: 0.647580
(Iteration 1334 / 1960) loss: 0.676160
(Iteration 1335 / 1960) loss: 0.757877
(Iteration 1336 / 1960) loss: 0.693491
(Iteration 1337 / 1960) loss: 0.762273
(Iteration 1338 / 1960) loss: 0.806483
(Iteration 1339 / 1960) loss: 0.783177
(Iteration 1340 / 1960) loss: 0.766596
(Iteration 1341 / 1960) loss: 0.747819
(Iteration 1342 / 1960) loss: 0.760528
(Iteration 1343 / 1960) loss: 0.737613
(Iteration 1344 / 1960) loss: 0.700674
(Iteration 1345 / 1960) loss: 0.753196
(Iteration 1346 / 1960) loss: 0.747852
(Iteration 1347 / 1960) loss: 0.749818
(Iteration 1348 / 1960) loss: 0.743333
(Iteration 1349 / 1960) loss: 0.725897
(Iteration 1350 / 1960) loss: 0.788427
(Iteration 1351 / 1960) loss: 0.774542
(Iteration 1352 / 1960) loss: 0.732356
(Iteration 1353 / 1960) loss: 0.842367
(Iteration 1354 / 1960) loss: 0.708658
(Iteration 1355 / 1960) loss: 0.773446
(Iteration 1356 / 1960) loss: 0.772474
(Iteration 1357 / 1960) loss: 0.714875
(Iteration 1358 / 1960) loss: 0.759278
(Iteration 1359 / 1960) loss: 0.753605
(Iteration 1360 / 1960) loss: 0.709001
(Iteration 1361 / 1960) loss: 0.770838
(Iteration 1362 / 1960) loss: 0.779731
(Iteration 1363 / 1960) loss: 0.740672
(Iteration 1364 / 1960) loss: 0.731437
(Iteration 1365 / 1960) loss: 0.773090
(Iteration 1366 / 1960) loss: 0.750952
(Iteration 1367 / 1960) loss: 0.750799
(Iteration 1368 / 1960) loss: 0.758837

(Iteration 1369 / 1960) loss: 0.759637
(Iteration 1370 / 1960) loss: 0.732007
(Iteration 1371 / 1960) loss: 0.733125
(Iteration 1372 / 1960) loss: 0.738467
(Epoch 14 / 20) train acc: 0.809000; val_acc: 0.650000
(Iteration 1373 / 1960) loss: 0.711891
(Iteration 1374 / 1960) loss: 0.773936
(Iteration 1375 / 1960) loss: 0.792264
(Iteration 1376 / 1960) loss: 0.782392
(Iteration 1377 / 1960) loss: 0.851789
(Iteration 1378 / 1960) loss: 0.759115
(Iteration 1379 / 1960) loss: 0.799514
(Iteration 1380 / 1960) loss: 0.704748
(Iteration 1381 / 1960) loss: 0.759859
(Iteration 1382 / 1960) loss: 0.710474
(Iteration 1383 / 1960) loss: 0.756272
(Iteration 1384 / 1960) loss: 0.681637
(Iteration 1385 / 1960) loss: 0.874452
(Iteration 1386 / 1960) loss: 0.738510
(Iteration 1387 / 1960) loss: 0.732714
(Iteration 1388 / 1960) loss: 0.716838
(Iteration 1389 / 1960) loss: 0.774342
(Iteration 1390 / 1960) loss: 0.689741
(Iteration 1391 / 1960) loss: 0.707669
(Iteration 1392 / 1960) loss: 0.738850
(Iteration 1393 / 1960) loss: 0.741688
(Iteration 1394 / 1960) loss: 0.758668
(Iteration 1395 / 1960) loss: 0.706094
(Iteration 1396 / 1960) loss: 0.764233
(Iteration 1397 / 1960) loss: 0.728259
(Iteration 1398 / 1960) loss: 0.683738
(Iteration 1399 / 1960) loss: 0.721625
(Iteration 1400 / 1960) loss: 0.700200
(Iteration 1401 / 1960) loss: 0.701728
(Iteration 1402 / 1960) loss: 0.701721
(Iteration 1403 / 1960) loss: 0.689155
(Iteration 1404 / 1960) loss: 0.748431
(Iteration 1405 / 1960) loss: 0.786471
(Iteration 1406 / 1960) loss: 0.709515
(Iteration 1407 / 1960) loss: 0.737741
(Iteration 1408 / 1960) loss: 0.763883
(Iteration 1409 / 1960) loss: 0.752022
(Iteration 1410 / 1960) loss: 0.723687
(Iteration 1411 / 1960) loss: 0.669827
(Iteration 1412 / 1960) loss: 0.800983
(Iteration 1413 / 1960) loss: 0.718867
(Iteration 1414 / 1960) loss: 0.769830
(Iteration 1415 / 1960) loss: 0.708405
(Iteration 1416 / 1960) loss: 0.677790
(Iteration 1417 / 1960) loss: 0.814777
(Iteration 1418 / 1960) loss: 0.704149
(Iteration 1419 / 1960) loss: 0.742875
(Iteration 1420 / 1960) loss: 0.742169
(Iteration 1421 / 1960) loss: 0.730594
(Iteration 1422 / 1960) loss: 0.728139
(Iteration 1423 / 1960) loss: 0.681720
(Iteration 1424 / 1960) loss: 0.710055

(Iteration 1425 / 1960) loss: 0.859868
(Iteration 1426 / 1960) loss: 0.672638
(Iteration 1427 / 1960) loss: 0.747939
(Iteration 1428 / 1960) loss: 0.672987
(Iteration 1429 / 1960) loss: 0.730480
(Iteration 1430 / 1960) loss: 0.768306
(Iteration 1431 / 1960) loss: 0.718741
(Iteration 1432 / 1960) loss: 0.711271
(Iteration 1433 / 1960) loss: 0.714998
(Iteration 1434 / 1960) loss: 0.751574
(Iteration 1435 / 1960) loss: 0.744269
(Iteration 1436 / 1960) loss: 0.744085
(Iteration 1437 / 1960) loss: 0.767885
(Iteration 1438 / 1960) loss: 0.710323
(Iteration 1439 / 1960) loss: 0.727570
(Iteration 1440 / 1960) loss: 0.769940
(Iteration 1441 / 1960) loss: 0.810212
(Iteration 1442 / 1960) loss: 0.761242
(Iteration 1443 / 1960) loss: 0.689922
(Iteration 1444 / 1960) loss: 0.761958
(Iteration 1445 / 1960) loss: 0.723767
(Iteration 1446 / 1960) loss: 0.668325
(Iteration 1447 / 1960) loss: 0.754125
(Iteration 1448 / 1960) loss: 0.707059
(Iteration 1449 / 1960) loss: 0.753938
(Iteration 1450 / 1960) loss: 0.648387
(Iteration 1451 / 1960) loss: 0.707460
(Iteration 1452 / 1960) loss: 0.734975
(Iteration 1453 / 1960) loss: 0.716065
(Iteration 1454 / 1960) loss: 0.679094
(Iteration 1455 / 1960) loss: 0.741920
(Iteration 1456 / 1960) loss: 0.710991
(Iteration 1457 / 1960) loss: 0.666102
(Iteration 1458 / 1960) loss: 0.779186
(Iteration 1459 / 1960) loss: 0.653461
(Iteration 1460 / 1960) loss: 0.714664
(Iteration 1461 / 1960) loss: 0.733225
(Iteration 1462 / 1960) loss: 0.751435
(Iteration 1463 / 1960) loss: 0.687593
(Iteration 1464 / 1960) loss: 0.693783
(Iteration 1465 / 1960) loss: 0.709442
(Iteration 1466 / 1960) loss: 0.678932
(Iteration 1467 / 1960) loss: 0.698733
(Iteration 1468 / 1960) loss: 0.745466
(Iteration 1469 / 1960) loss: 0.763851
(Iteration 1470 / 1960) loss: 0.670406
(Epoch 15 / 20) train acc: 0.843000; val_acc: 0.663000
(Iteration 1471 / 1960) loss: 0.672973
(Iteration 1472 / 1960) loss: 0.688509
(Iteration 1473 / 1960) loss: 0.741592
(Iteration 1474 / 1960) loss: 0.653267
(Iteration 1475 / 1960) loss: 0.663511
(Iteration 1476 / 1960) loss: 0.700297
(Iteration 1477 / 1960) loss: 0.691567
(Iteration 1478 / 1960) loss: 0.640038
(Iteration 1479 / 1960) loss: 0.674340
(Iteration 1480 / 1960) loss: 0.626324

(Iteration 1481 / 1960) loss: 0.669756
(Iteration 1482 / 1960) loss: 0.636622
(Iteration 1483 / 1960) loss: 0.674387
(Iteration 1484 / 1960) loss: 0.649861
(Iteration 1485 / 1960) loss: 0.691079
(Iteration 1486 / 1960) loss: 0.715158
(Iteration 1487 / 1960) loss: 0.724020
(Iteration 1488 / 1960) loss: 0.780820
(Iteration 1489 / 1960) loss: 0.698629
(Iteration 1490 / 1960) loss: 0.674810
(Iteration 1491 / 1960) loss: 0.786208
(Iteration 1492 / 1960) loss: 0.641823
(Iteration 1493 / 1960) loss: 0.778298
(Iteration 1494 / 1960) loss: 0.723188
(Iteration 1495 / 1960) loss: 0.693149
(Iteration 1496 / 1960) loss: 0.783951
(Iteration 1497 / 1960) loss: 0.709114
(Iteration 1498 / 1960) loss: 0.687749
(Iteration 1499 / 1960) loss: 0.770403
(Iteration 1500 / 1960) loss: 0.651654
(Iteration 1501 / 1960) loss: 0.705598
(Iteration 1502 / 1960) loss: 0.707666
(Iteration 1503 / 1960) loss: 0.704260
(Iteration 1504 / 1960) loss: 0.717780
(Iteration 1505 / 1960) loss: 0.705659
(Iteration 1506 / 1960) loss: 0.754948
(Iteration 1507 / 1960) loss: 0.713149
(Iteration 1508 / 1960) loss: 0.693015
(Iteration 1509 / 1960) loss: 0.630471
(Iteration 1510 / 1960) loss: 0.698810
(Iteration 1511 / 1960) loss: 0.805243
(Iteration 1512 / 1960) loss: 0.657907
(Iteration 1513 / 1960) loss: 0.682453
(Iteration 1514 / 1960) loss: 0.736155
(Iteration 1515 / 1960) loss: 0.739846
(Iteration 1516 / 1960) loss: 0.708074
(Iteration 1517 / 1960) loss: 0.682877
(Iteration 1518 / 1960) loss: 0.684249
(Iteration 1519 / 1960) loss: 0.744003
(Iteration 1520 / 1960) loss: 0.691791
(Iteration 1521 / 1960) loss: 0.710579
(Iteration 1522 / 1960) loss: 0.742822
(Iteration 1523 / 1960) loss: 0.819097
(Iteration 1524 / 1960) loss: 0.727303
(Iteration 1525 / 1960) loss: 0.687381
(Iteration 1526 / 1960) loss: 0.772831
(Iteration 1527 / 1960) loss: 0.724350
(Iteration 1528 / 1960) loss: 0.685294
(Iteration 1529 / 1960) loss: 0.712999
(Iteration 1530 / 1960) loss: 0.706040
(Iteration 1531 / 1960) loss: 0.736891
(Iteration 1532 / 1960) loss: 0.691114
(Iteration 1533 / 1960) loss: 0.724970
(Iteration 1534 / 1960) loss: 0.655943
(Iteration 1535 / 1960) loss: 0.684397
(Iteration 1536 / 1960) loss: 0.725500
(Iteration 1537 / 1960) loss: 0.660774

```
(Iteration 1538 / 1960) loss: 0.639744
(Iteration 1539 / 1960) loss: 0.712551
(Iteration 1540 / 1960) loss: 0.669418
(Iteration 1541 / 1960) loss: 0.756638
(Iteration 1542 / 1960) loss: 0.642598
(Iteration 1543 / 1960) loss: 0.660802
(Iteration 1544 / 1960) loss: 0.644207
(Iteration 1545 / 1960) loss: 0.721392
(Iteration 1546 / 1960) loss: 0.651712
(Iteration 1547 / 1960) loss: 0.731904
(Iteration 1548 / 1960) loss: 0.707442
(Iteration 1549 / 1960) loss: 0.684836
(Iteration 1550 / 1960) loss: 0.706017
(Iteration 1551 / 1960) loss: 0.711558
(Iteration 1552 / 1960) loss: 0.656665
(Iteration 1553 / 1960) loss: 0.706814
(Iteration 1554 / 1960) loss: 0.724124
(Iteration 1555 / 1960) loss: 0.674539
(Iteration 1556 / 1960) loss: 0.608439
(Iteration 1557 / 1960) loss: 0.764276
(Iteration 1558 / 1960) loss: 0.644679
(Iteration 1559 / 1960) loss: 0.767085
(Iteration 1560 / 1960) loss: 0.736924
(Iteration 1561 / 1960) loss: 0.661914
(Iteration 1562 / 1960) loss: 0.673170
(Iteration 1563 / 1960) loss: 0.712002
(Iteration 1564 / 1960) loss: 0.697587
(Iteration 1565 / 1960) loss: 0.740635
(Iteration 1566 / 1960) loss: 0.696243
(Iteration 1567 / 1960) loss: 0.706442
(Iteration 1568 / 1960) loss: 0.675124
(Epoch 16 / 20) train acc: 0.861000; val_acc: 0.682000
(Iteration 1569 / 1960) loss: 0.655841
(Iteration 1570 / 1960) loss: 0.602668
(Iteration 1571 / 1960) loss: 0.671229
(Iteration 1572 / 1960) loss: 0.700893
(Iteration 1573 / 1960) loss: 0.686412
(Iteration 1574 / 1960) loss: 0.665930
(Iteration 1575 / 1960) loss: 0.700338
(Iteration 1576 / 1960) loss: 0.710718
(Iteration 1577 / 1960) loss: 0.628692
(Iteration 1578 / 1960) loss: 0.666314
(Iteration 1579 / 1960) loss: 0.687997
(Iteration 1580 / 1960) loss: 0.711332
(Iteration 1581 / 1960) loss: 0.658501
(Iteration 1582 / 1960) loss: 0.686918
(Iteration 1583 / 1960) loss: 0.737420
(Iteration 1584 / 1960) loss: 0.760767
(Iteration 1585 / 1960) loss: 0.701757
(Iteration 1586 / 1960) loss: 0.703558
(Iteration 1587 / 1960) loss: 0.680734
(Iteration 1588 / 1960) loss: 0.743737
(Iteration 1589 / 1960) loss: 0.723501
(Iteration 1590 / 1960) loss: 0.653156
(Iteration 1591 / 1960) loss: 0.671266
(Iteration 1592 / 1960) loss: 0.771111
(Iteration 1593 / 1960) loss: 0.695194
```

(Iteration 1594 / 1960) loss: 0.707476
(Iteration 1595 / 1960) loss: 0.628307
(Iteration 1596 / 1960) loss: 0.831224
(Iteration 1597 / 1960) loss: 0.714803
(Iteration 1598 / 1960) loss: 0.640318
(Iteration 1599 / 1960) loss: 0.715629
(Iteration 1600 / 1960) loss: 0.661048
(Iteration 1601 / 1960) loss: 0.608202
(Iteration 1602 / 1960) loss: 0.667790
(Iteration 1603 / 1960) loss: 0.656272
(Iteration 1604 / 1960) loss: 0.628959
(Iteration 1605 / 1960) loss: 0.754326
(Iteration 1606 / 1960) loss: 0.707628
(Iteration 1607 / 1960) loss: 0.733853
(Iteration 1608 / 1960) loss: 0.695885
(Iteration 1609 / 1960) loss: 0.615795
(Iteration 1610 / 1960) loss: 0.723973
(Iteration 1611 / 1960) loss: 0.760600
(Iteration 1612 / 1960) loss: 0.710079
(Iteration 1613 / 1960) loss: 0.676100
(Iteration 1614 / 1960) loss: 0.664931
(Iteration 1615 / 1960) loss: 0.656115
(Iteration 1616 / 1960) loss: 0.708469
(Iteration 1617 / 1960) loss: 0.715094
(Iteration 1618 / 1960) loss: 0.673284
(Iteration 1619 / 1960) loss: 0.623415
(Iteration 1620 / 1960) loss: 0.697515
(Iteration 1621 / 1960) loss: 0.671354
(Iteration 1622 / 1960) loss: 0.680121
(Iteration 1623 / 1960) loss: 0.706531
(Iteration 1624 / 1960) loss: 0.672365
(Iteration 1625 / 1960) loss: 0.658804
(Iteration 1626 / 1960) loss: 0.662789
(Iteration 1627 / 1960) loss: 0.699534
(Iteration 1628 / 1960) loss: 0.671015
(Iteration 1629 / 1960) loss: 0.708028
(Iteration 1630 / 1960) loss: 0.656915
(Iteration 1631 / 1960) loss: 0.637129
(Iteration 1632 / 1960) loss: 0.743378
(Iteration 1633 / 1960) loss: 0.676235
(Iteration 1634 / 1960) loss: 0.695620
(Iteration 1635 / 1960) loss: 0.660691
(Iteration 1636 / 1960) loss: 0.691184
(Iteration 1637 / 1960) loss: 0.599474
(Iteration 1638 / 1960) loss: 0.582916
(Iteration 1639 / 1960) loss: 0.639048
(Iteration 1640 / 1960) loss: 0.681285
(Iteration 1641 / 1960) loss: 0.711625
(Iteration 1642 / 1960) loss: 0.700397
(Iteration 1643 / 1960) loss: 0.700633
(Iteration 1644 / 1960) loss: 0.663670
(Iteration 1645 / 1960) loss: 0.658383
(Iteration 1646 / 1960) loss: 0.670402
(Iteration 1647 / 1960) loss: 0.668219
(Iteration 1648 / 1960) loss: 0.660251
(Iteration 1649 / 1960) loss: 0.722930
(Iteration 1650 / 1960) loss: 0.650871

(Iteration 1651 / 1960) loss: 0.632961
(Iteration 1652 / 1960) loss: 0.685432
(Iteration 1653 / 1960) loss: 0.679876
(Iteration 1654 / 1960) loss: 0.698697
(Iteration 1655 / 1960) loss: 0.618837
(Iteration 1656 / 1960) loss: 0.709099
(Iteration 1657 / 1960) loss: 0.707617
(Iteration 1658 / 1960) loss: 0.666759
(Iteration 1659 / 1960) loss: 0.687388
(Iteration 1660 / 1960) loss: 0.744078
(Iteration 1661 / 1960) loss: 0.624966
(Iteration 1662 / 1960) loss: 0.665272
(Iteration 1663 / 1960) loss: 0.692051
(Iteration 1664 / 1960) loss: 0.616535
(Iteration 1665 / 1960) loss: 0.706358
(Iteration 1666 / 1960) loss: 0.645691
(Epoch 17 / 20) train acc: 0.866000; val_acc: 0.670000
(Iteration 1667 / 1960) loss: 0.707185
(Iteration 1668 / 1960) loss: 0.657506
(Iteration 1669 / 1960) loss: 0.617042
(Iteration 1670 / 1960) loss: 0.642616
(Iteration 1671 / 1960) loss: 0.624377
(Iteration 1672 / 1960) loss: 0.667702
(Iteration 1673 / 1960) loss: 0.636915
(Iteration 1674 / 1960) loss: 0.654231
(Iteration 1675 / 1960) loss: 0.702759
(Iteration 1676 / 1960) loss: 0.634614
(Iteration 1677 / 1960) loss: 0.635743
(Iteration 1678 / 1960) loss: 0.698192
(Iteration 1679 / 1960) loss: 0.656131
(Iteration 1680 / 1960) loss: 0.696133
(Iteration 1681 / 1960) loss: 0.681634
(Iteration 1682 / 1960) loss: 0.660362
(Iteration 1683 / 1960) loss: 0.717756
(Iteration 1684 / 1960) loss: 0.757861
(Iteration 1685 / 1960) loss: 0.673987
(Iteration 1686 / 1960) loss: 0.662198
(Iteration 1687 / 1960) loss: 0.656492
(Iteration 1688 / 1960) loss: 0.652413
(Iteration 1689 / 1960) loss: 0.743006
(Iteration 1690 / 1960) loss: 0.733370
(Iteration 1691 / 1960) loss: 0.700963
(Iteration 1692 / 1960) loss: 0.766612
(Iteration 1693 / 1960) loss: 0.680562
(Iteration 1694 / 1960) loss: 0.655724
(Iteration 1695 / 1960) loss: 0.708660
(Iteration 1696 / 1960) loss: 0.672633
(Iteration 1697 / 1960) loss: 0.675636
(Iteration 1698 / 1960) loss: 0.706384
(Iteration 1699 / 1960) loss: 0.623147
(Iteration 1700 / 1960) loss: 0.661169
(Iteration 1701 / 1960) loss: 0.574689
(Iteration 1702 / 1960) loss: 0.626006
(Iteration 1703 / 1960) loss: 0.614477
(Iteration 1704 / 1960) loss: 0.678779
(Iteration 1705 / 1960) loss: 0.594305
(Iteration 1706 / 1960) loss: 0.669895

(Iteration 1707 / 1960) loss: 0.676871
(Iteration 1708 / 1960) loss: 0.670982
(Iteration 1709 / 1960) loss: 0.706890
(Iteration 1710 / 1960) loss: 0.663242
(Iteration 1711 / 1960) loss: 0.667616
(Iteration 1712 / 1960) loss: 0.596342
(Iteration 1713 / 1960) loss: 0.640216
(Iteration 1714 / 1960) loss: 0.640067
(Iteration 1715 / 1960) loss: 0.653853
(Iteration 1716 / 1960) loss: 0.664200
(Iteration 1717 / 1960) loss: 0.591878
(Iteration 1718 / 1960) loss: 0.628537
(Iteration 1719 / 1960) loss: 0.661203
(Iteration 1720 / 1960) loss: 0.638321
(Iteration 1721 / 1960) loss: 0.633889
(Iteration 1722 / 1960) loss: 0.631249
(Iteration 1723 / 1960) loss: 0.678080
(Iteration 1724 / 1960) loss: 0.639317
(Iteration 1725 / 1960) loss: 0.629281
(Iteration 1726 / 1960) loss: 0.667683
(Iteration 1727 / 1960) loss: 0.612787
(Iteration 1728 / 1960) loss: 0.643852
(Iteration 1729 / 1960) loss: 0.592208
(Iteration 1730 / 1960) loss: 0.608365
(Iteration 1731 / 1960) loss: 0.649753
(Iteration 1732 / 1960) loss: 0.646724
(Iteration 1733 / 1960) loss: 0.589285
(Iteration 1734 / 1960) loss: 0.642265
(Iteration 1735 / 1960) loss: 0.613786
(Iteration 1736 / 1960) loss: 0.648891
(Iteration 1737 / 1960) loss: 0.638395
(Iteration 1738 / 1960) loss: 0.716426
(Iteration 1739 / 1960) loss: 0.667567
(Iteration 1740 / 1960) loss: 0.600275
(Iteration 1741 / 1960) loss: 0.583032
(Iteration 1742 / 1960) loss: 0.713361
(Iteration 1743 / 1960) loss: 0.620422
(Iteration 1744 / 1960) loss: 0.630211
(Iteration 1745 / 1960) loss: 0.656224
(Iteration 1746 / 1960) loss: 0.643366
(Iteration 1747 / 1960) loss: 0.629964
(Iteration 1748 / 1960) loss: 0.615189
(Iteration 1749 / 1960) loss: 0.606310
(Iteration 1750 / 1960) loss: 0.672451
(Iteration 1751 / 1960) loss: 0.628116
(Iteration 1752 / 1960) loss: 0.672304
(Iteration 1753 / 1960) loss: 0.589369
(Iteration 1754 / 1960) loss: 0.544999
(Iteration 1755 / 1960) loss: 0.660190
(Iteration 1756 / 1960) loss: 0.634370
(Iteration 1757 / 1960) loss: 0.633508
(Iteration 1758 / 1960) loss: 0.662029
(Iteration 1759 / 1960) loss: 0.610762
(Iteration 1760 / 1960) loss: 0.586926
(Iteration 1761 / 1960) loss: 0.675942
(Iteration 1762 / 1960) loss: 0.581656
(Iteration 1763 / 1960) loss: 0.632583

(Iteration 1764 / 1960) loss: 0.598559
(Epoch 18 / 20) train acc: 0.869000; val_acc: 0.652000
(Iteration 1765 / 1960) loss: 0.624110
(Iteration 1766 / 1960) loss: 0.596845
(Iteration 1767 / 1960) loss: 0.610988
(Iteration 1768 / 1960) loss: 0.573706
(Iteration 1769 / 1960) loss: 0.590023
(Iteration 1770 / 1960) loss: 0.603161
(Iteration 1771 / 1960) loss: 0.635027
(Iteration 1772 / 1960) loss: 0.613844
(Iteration 1773 / 1960) loss: 0.657714
(Iteration 1774 / 1960) loss: 0.676618
(Iteration 1775 / 1960) loss: 0.620750
(Iteration 1776 / 1960) loss: 0.589047
(Iteration 1777 / 1960) loss: 0.586504
(Iteration 1778 / 1960) loss: 0.714188
(Iteration 1779 / 1960) loss: 0.561016
(Iteration 1780 / 1960) loss: 0.673400
(Iteration 1781 / 1960) loss: 0.595225
(Iteration 1782 / 1960) loss: 0.678023
(Iteration 1783 / 1960) loss: 0.643241
(Iteration 1784 / 1960) loss: 0.647341
(Iteration 1785 / 1960) loss: 0.583297
(Iteration 1786 / 1960) loss: 0.585162
(Iteration 1787 / 1960) loss: 0.571219
(Iteration 1788 / 1960) loss: 0.664321
(Iteration 1789 / 1960) loss: 0.649608
(Iteration 1790 / 1960) loss: 0.605812
(Iteration 1791 / 1960) loss: 0.596035
(Iteration 1792 / 1960) loss: 0.638730
(Iteration 1793 / 1960) loss: 0.631708
(Iteration 1794 / 1960) loss: 0.621739
(Iteration 1795 / 1960) loss: 0.658398
(Iteration 1796 / 1960) loss: 0.639533
(Iteration 1797 / 1960) loss: 0.639115
(Iteration 1798 / 1960) loss: 0.608987
(Iteration 1799 / 1960) loss: 0.641618
(Iteration 1800 / 1960) loss: 0.556847
(Iteration 1801 / 1960) loss: 0.594556
(Iteration 1802 / 1960) loss: 0.642324
(Iteration 1803 / 1960) loss: 0.638934
(Iteration 1804 / 1960) loss: 0.592191
(Iteration 1805 / 1960) loss: 0.621562
(Iteration 1806 / 1960) loss: 0.539380
(Iteration 1807 / 1960) loss: 0.575931
(Iteration 1808 / 1960) loss: 0.633319
(Iteration 1809 / 1960) loss: 0.563323
(Iteration 1810 / 1960) loss: 0.649370
(Iteration 1811 / 1960) loss: 0.663921
(Iteration 1812 / 1960) loss: 0.564028
(Iteration 1813 / 1960) loss: 0.670924
(Iteration 1814 / 1960) loss: 0.627419
(Iteration 1815 / 1960) loss: 0.582259
(Iteration 1816 / 1960) loss: 0.551018
(Iteration 1817 / 1960) loss: 0.652877
(Iteration 1818 / 1960) loss: 0.567827
(Iteration 1819 / 1960) loss: 0.537292

(Iteration 1820 / 1960) loss: 0.604642
(Iteration 1821 / 1960) loss: 0.610145
(Iteration 1822 / 1960) loss: 0.625813
(Iteration 1823 / 1960) loss: 0.618853
(Iteration 1824 / 1960) loss: 0.659915
(Iteration 1825 / 1960) loss: 0.627485
(Iteration 1826 / 1960) loss: 0.627627
(Iteration 1827 / 1960) loss: 0.580776
(Iteration 1828 / 1960) loss: 0.591040
(Iteration 1829 / 1960) loss: 0.581282
(Iteration 1830 / 1960) loss: 0.579132
(Iteration 1831 / 1960) loss: 0.569143
(Iteration 1832 / 1960) loss: 0.625171
(Iteration 1833 / 1960) loss: 0.593758
(Iteration 1834 / 1960) loss: 0.583042
(Iteration 1835 / 1960) loss: 0.546782
(Iteration 1836 / 1960) loss: 0.634370
(Iteration 1837 / 1960) loss: 0.601586
(Iteration 1838 / 1960) loss: 0.643241
(Iteration 1839 / 1960) loss: 0.646640
(Iteration 1840 / 1960) loss: 0.619559
(Iteration 1841 / 1960) loss: 0.664952
(Iteration 1842 / 1960) loss: 0.598807
(Iteration 1843 / 1960) loss: 0.633911
(Iteration 1844 / 1960) loss: 0.610461
(Iteration 1845 / 1960) loss: 0.616077
(Iteration 1846 / 1960) loss: 0.657680
(Iteration 1847 / 1960) loss: 0.573560
(Iteration 1848 / 1960) loss: 0.622198
(Iteration 1849 / 1960) loss: 0.625380
(Iteration 1850 / 1960) loss: 0.597841
(Iteration 1851 / 1960) loss: 0.594031
(Iteration 1852 / 1960) loss: 0.546777
(Iteration 1853 / 1960) loss: 0.619363
(Iteration 1854 / 1960) loss: 0.608965
(Iteration 1855 / 1960) loss: 0.551641
(Iteration 1856 / 1960) loss: 0.638825
(Iteration 1857 / 1960) loss: 0.605772
(Iteration 1858 / 1960) loss: 0.544667
(Iteration 1859 / 1960) loss: 0.639409
(Iteration 1860 / 1960) loss: 0.584444
(Iteration 1861 / 1960) loss: 0.610413
(Iteration 1862 / 1960) loss: 0.615803
(Epoch 19 / 20) train acc: 0.872000; val_acc: 0.678000
(Iteration 1863 / 1960) loss: 0.528473
(Iteration 1864 / 1960) loss: 0.599923
(Iteration 1865 / 1960) loss: 0.584615
(Iteration 1866 / 1960) loss: 0.581169
(Iteration 1867 / 1960) loss: 0.620451
(Iteration 1868 / 1960) loss: 0.552258
(Iteration 1869 / 1960) loss: 0.617291
(Iteration 1870 / 1960) loss: 0.607924
(Iteration 1871 / 1960) loss: 0.613216
(Iteration 1872 / 1960) loss: 0.639325
(Iteration 1873 / 1960) loss: 0.650543
(Iteration 1874 / 1960) loss: 0.572874
(Iteration 1875 / 1960) loss: 0.594442

(Iteration 1876 / 1960)	loss: 0.606672
(Iteration 1877 / 1960)	loss: 0.550667
(Iteration 1878 / 1960)	loss: 0.570674
(Iteration 1879 / 1960)	loss: 0.602913
(Iteration 1880 / 1960)	loss: 0.561317
(Iteration 1881 / 1960)	loss: 0.591927
(Iteration 1882 / 1960)	loss: 0.599683
(Iteration 1883 / 1960)	loss: 0.584088
(Iteration 1884 / 1960)	loss: 0.605605
(Iteration 1885 / 1960)	loss: 0.574979
(Iteration 1886 / 1960)	loss: 0.540026
(Iteration 1887 / 1960)	loss: 0.588381
(Iteration 1888 / 1960)	loss: 0.615201
(Iteration 1889 / 1960)	loss: 0.562335
(Iteration 1890 / 1960)	loss: 0.564247
(Iteration 1891 / 1960)	loss: 0.638745
(Iteration 1892 / 1960)	loss: 0.577460
(Iteration 1893 / 1960)	loss: 0.608551
(Iteration 1894 / 1960)	loss: 0.550391
(Iteration 1895 / 1960)	loss: 0.576879
(Iteration 1896 / 1960)	loss: 0.569594
(Iteration 1897 / 1960)	loss: 0.573881
(Iteration 1898 / 1960)	loss: 0.633293
(Iteration 1899 / 1960)	loss: 0.578776
(Iteration 1900 / 1960)	loss: 0.567274
(Iteration 1901 / 1960)	loss: 0.624730
(Iteration 1902 / 1960)	loss: 0.572424
(Iteration 1903 / 1960)	loss: 0.569173
(Iteration 1904 / 1960)	loss: 0.527592
(Iteration 1905 / 1960)	loss: 0.577532
(Iteration 1906 / 1960)	loss: 0.564464
(Iteration 1907 / 1960)	loss: 0.581675
(Iteration 1908 / 1960)	loss: 0.594935
(Iteration 1909 / 1960)	loss: 0.576347
(Iteration 1910 / 1960)	loss: 0.602785
(Iteration 1911 / 1960)	loss: 0.596645
(Iteration 1912 / 1960)	loss: 0.578425
(Iteration 1913 / 1960)	loss: 0.601921
(Iteration 1914 / 1960)	loss: 0.567646
(Iteration 1915 / 1960)	loss: 0.567353
(Iteration 1916 / 1960)	loss: 0.520739
(Iteration 1917 / 1960)	loss: 0.565597
(Iteration 1918 / 1960)	loss: 0.564995
(Iteration 1919 / 1960)	loss: 0.551179
(Iteration 1920 / 1960)	loss: 0.564372
(Iteration 1921 / 1960)	loss: 0.594811
(Iteration 1922 / 1960)	loss: 0.584596
(Iteration 1923 / 1960)	loss: 0.557993
(Iteration 1924 / 1960)	loss: 0.617314
(Iteration 1925 / 1960)	loss: 0.571974
(Iteration 1926 / 1960)	loss: 0.617921
(Iteration 1927 / 1960)	loss: 0.580649
(Iteration 1928 / 1960)	loss: 0.542720
(Iteration 1929 / 1960)	loss: 0.545295
(Iteration 1930 / 1960)	loss: 0.538512
(Iteration 1931 / 1960)	loss: 0.569153
(Iteration 1932 / 1960)	loss: 0.671134

(Iteration 1933 / 1960) loss: 0.595565
(Iteration 1934 / 1960) loss: 0.605014
(Iteration 1935 / 1960) loss: 0.594055
(Iteration 1936 / 1960) loss: 0.598601
(Iteration 1937 / 1960) loss: 0.617905
(Iteration 1938 / 1960) loss: 0.628643
(Iteration 1939 / 1960) loss: 0.605354
(Iteration 1940 / 1960) loss: 0.592557
(Iteration 1941 / 1960) loss: 0.611027
(Iteration 1942 / 1960) loss: 0.587949
(Iteration 1943 / 1960) loss: 0.564501
(Iteration 1944 / 1960) loss: 0.628398
(Iteration 1945 / 1960) loss: 0.574424
(Iteration 1946 / 1960) loss: 0.608729
(Iteration 1947 / 1960) loss: 0.558325
(Iteration 1948 / 1960) loss: 0.541109
(Iteration 1949 / 1960) loss: 0.636479
(Iteration 1950 / 1960) loss: 0.532941
(Iteration 1951 / 1960) loss: 0.549662
(Iteration 1952 / 1960) loss: 0.571513
(Iteration 1953 / 1960) loss: 0.550808
(Iteration 1954 / 1960) loss: 0.600695
(Iteration 1955 / 1960) loss: 0.545769
(Iteration 1956 / 1960) loss: 0.618362
(Iteration 1957 / 1960) loss: 0.619695
(Iteration 1958 / 1960) loss: 0.534447
(Iteration 1959 / 1960) loss: 0.607649
(Iteration 1960 / 1960) loss: 0.558238
(Epoch 20 / 20) train acc: 0.894000; val_acc: 0.667000

