ECE C147/247 HW4 Q3: Dropout

In this notebook, you will implement dropout. Then we will ask you to train a network with batchnorm and dropout, and acheive over 55% accuracy on CIFAR-10.

`utils` 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, and their layer structure. This also includes `nndl.fc_net`, `nndl.layers`, and `nndl.layer_utils`.

In [ ]:
```
## Import and setups

import time
import numpy as np
import matplotlib.pyplot as plt
from nndl.fc_net import *
from nndl.layers import *
from utils.data_utils import get_CIFAR10_data
from utils.gradient_check import eval_numerical_gradient, eval_numerical_gradient_arr
from utils.solver import Solver

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

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

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

In [ ]:
```
# 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,)
```

Dropout forward pass

Implement the training and test time dropout forward pass, `dropout_forward`, in `nndl/layers.py`. After that, test your implementation by running the following cell.

In [ ]:
```
x = np.random.randn(500, 500) + 10

for p in [0.3, 0.6, 0.75]:
  out, _ = dropout_forward(x, {'mode': 'train', 'p': p})
  out_test, _ = dropout_forward(x, {'mode': 'test', 'p': p})

  print('Running tests with p = ', p)
  print('Mean of input: ', x.mean())
  print('Mean of train-time output: ', out.mean())
  print('Mean of test-time output: ', out_test.mean())
  print('Fraction of train-time output set to zero: ', (out == 0).mean())
  print('Fraction of test-time output set to zero: ', (out_test == 0).mean())
```

```
Running tests with p =  0.3
Mean of input:  9.997149956912688
Mean of train-time output:  9.995329989880013
Mean of test-time output:  9.997149956912688
Fraction of train-time output set to zero:  0.30018
Fraction of test-time output set to zero:  0.0
Running tests with p =  0.6
Mean of input:  9.997149956912688
Mean of train-time output:  10.004803612775156
Mean of test-time output:  9.997149956912688
Fraction of train-time output set to zero:  0.59964
Fraction of test-time output set to zero:  0.0
Running tests with p =  0.75
Mean of input:  9.997149956912688
Mean of train-time output:  9.956688011868048
Mean of test-time output:  9.997149956912688
Fraction of train-time output set to zero:  0.750956
Fraction of test-time output set to zero:  0.0
```

Dropout backward pass

Implement the backward pass, `dropout_backward`, in `nndl/layers.py`. After that, test your gradients by running the following cell:

In [ ]:
```
x = np.random.randn(10, 10) + 10
dout = np.random.randn(*x.shape)

dropout_param = {'mode': 'train', 'p': 0.8, 'seed': 123}
out, cache = dropout_forward(x, dropout_param)
dx = dropout_backward(dout, cache)
dx_num = eval_numerical_gradient_array(lambda xx: dropout_forward(xx, dropout_param)[

print('dx relative error: ', rel_error(dx, dx_num))
```

```
dx relative error:  1.8929089298490347e-11
```

Implement a fully connected neural network with dropout layers

Modify the `FullyConnectedNet()` class in `nndl/fc_net.py` to incorporate dropout. A dropout layer should be incorporated after every ReLU layer. Concretely, there shouldn't be a dropout at the output layer since there is no ReLU at the output layer. You will need to modify the class in the following areas:

(1) In the forward pass, you will need to incorporate a dropout layer after every relu layer.

(2) In the backward pass, you will need to incorporate a dropout backward pass layer.

Check your implementation by running the following code. Our W1 gradient relative error is on the order of 1e-6 (the largest of all the relative errors).

In [ ]:

```
N, D, H1, H2, C = 2, 15, 20, 30, 10
X = np.random.randn(N, D)
y = np.random.randint(C, size=(N,))

for dropout in [0, 0.25, 0.5]:
  print('Running check with dropout = ', dropout)
  model = FullyConnectedNet([H1, H2], input_dim=D, num_classes=C,
                            weight_scale=5e-2, dtype=np.float64,
                            dropout=dropout, seed=123)

  loss, grads = model.loss(X, y)
  print('Initial loss: ', loss)

  for name in sorted(grads):
    f = lambda _: model.loss(X, y)[0]
    grad_num = eval_numerical_gradient(f, model.params[name], verbose=False, h=1e-5)
    print('{} relative error: {}'.format(name, rel_error(grad_num, grads[name])))
  print('\n')
```

```
Running check with dropout =  0
Initial loss:  2.3051948273987857
W1 relative error: 2.5272575344376073e-07
W2 relative error: 1.5034484929313676e-05
W3 relative error: 2.753446833630168e-07
b1 relative error: 2.936957476400148e-06
b2 relative error: 5.051339805546953e-08
b3 relative error: 1.1740467838205477e-10


Running check with dropout =  0.25
Initial loss:  2.3052077546540826
W1 relative error: 2.613846944812385e-07
W2 relative error: 5.022056536108928e-07
W3 relative error: 4.456316077044505e-08
b1 relative error: 7.39711723790801e-08
b2 relative error: 7.151678402730031e-10
b3 relative error: 1.003974732116764e-10


Running check with dropout =  0.5
Initial loss:  2.3035667586595423
W1 relative error: 1.1401257458777745e-06
W2 relative error: 1.847669681023635e-07
W3 relative error: 6.5966195253431734e-09
b1 relative error: 7.71639621892128e-08
b2 relative error: 1.1975910493629166e-09
b3 relative error: 1.4558471033827801e-10
```

Dropout as a regularizer

In class, we claimed that dropout acts as a regularizer by effectively bagging. To check this, we will train two small networks, one with dropout and one without dropout.

In [ ]:

```
# Train two identical nets, one with dropout and one without

num_train = 500
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'],
}

solvers = {}
dropout_choices = [0, 0.6]
for dropout in dropout_choices:
  model = FullyConnectedNet([100, 100, 100], dropout=dropout)

  solver = Solver(model, small_data,
                  num_epochs=25, batch_size=100,
                  update_rule='adam',
                  optim_config={
                    'learning_rate': 5e-4,
                  },
                  verbose=True, print_every=100)
```

```
  solver.train()
  solvers[dropout] = solver
```

```
(Iteration 1 / 125) loss: 2.300804
(Epoch 0 / 25) train acc: 0.220000; val_acc: 0.168000
(Epoch 1 / 25) train acc: 0.188000; val_acc: 0.147000
(Epoch 2 / 25) train acc: 0.266000; val_acc: 0.200000
(Epoch 3 / 25) train acc: 0.338000; val_acc: 0.262000
(Epoch 4 / 25) train acc: 0.378000; val_acc: 0.278000
(Epoch 5 / 25) train acc: 0.428000; val_acc: 0.297000
(Epoch 6 / 25) train acc: 0.468000; val_acc: 0.323000
(Epoch 7 / 25) train acc: 0.494000; val_acc: 0.287000
(Epoch 8 / 25) train acc: 0.566000; val_acc: 0.328000
(Epoch 9 / 25) train acc: 0.572000; val_acc: 0.322000
(Epoch 10 / 25) train acc: 0.622000; val_acc: 0.324000
(Epoch 11 / 25) train acc: 0.670000; val_acc: 0.279000
(Epoch 12 / 25) train acc: 0.710000; val_acc: 0.338000
(Epoch 13 / 25) train acc: 0.746000; val_acc: 0.319000
(Epoch 14 / 25) train acc: 0.792000; val_acc: 0.307000
(Epoch 15 / 25) train acc: 0.834000; val_acc: 0.297000
(Epoch 16 / 25) train acc: 0.876000; val_acc: 0.327000
(Epoch 17 / 25) train acc: 0.886000; val_acc: 0.320000
(Epoch 18 / 25) train acc: 0.918000; val_acc: 0.314000
(Epoch 19 / 25) train acc: 0.922000; val_acc: 0.290000
(Epoch 20 / 25) train acc: 0.944000; val_acc: 0.306000
(Iteration 101 / 125) loss: 0.156105
(Epoch 21 / 25) train acc: 0.968000; val_acc: 0.302000
(Epoch 22 / 25) train acc: 0.978000; val_acc: 0.302000
(Epoch 23 / 25) train acc: 0.976000; val_acc: 0.289000
(Epoch 24 / 25) train acc: 0.986000; val_acc: 0.285000
(Epoch 25 / 25) train acc: 0.978000; val_acc: 0.311000
(Iteration 1 / 125) loss: 2.298716
(Epoch 0 / 25) train acc: 0.132000; val_acc: 0.146000
(Epoch 1 / 25) train acc: 0.118000; val_acc: 0.131000
(Epoch 2 / 25) train acc: 0.220000; val_acc: 0.214000
(Epoch 3 / 25) train acc: 0.206000; val_acc: 0.180000
(Epoch 4 / 25) train acc: 0.220000; val_acc: 0.193000
(Epoch 5 / 25) train acc: 0.264000; val_acc: 0.229000
(Epoch 6 / 25) train acc: 0.268000; val_acc: 0.203000
(Epoch 7 / 25) train acc: 0.266000; val_acc: 0.212000
(Epoch 8 / 25) train acc: 0.282000; val_acc: 0.236000
(Epoch 9 / 25) train acc: 0.310000; val_acc: 0.255000
(Epoch 10 / 25) train acc: 0.320000; val_acc: 0.267000
(Epoch 11 / 25) train acc: 0.338000; val_acc: 0.273000
(Epoch 12 / 25) train acc: 0.346000; val_acc: 0.278000
(Epoch 13 / 25) train acc: 0.332000; val_acc: 0.279000
(Epoch 14 / 25) train acc: 0.328000; val_acc: 0.284000
(Epoch 15 / 25) train acc: 0.354000; val_acc: 0.271000
(Epoch 16 / 25) train acc: 0.386000; val_acc: 0.277000
(Epoch 17 / 25) train acc: 0.388000; val_acc: 0.297000
(Epoch 18 / 25) train acc: 0.402000; val_acc: 0.280000
(Epoch 19 / 25) train acc: 0.388000; val_acc: 0.274000
(Epoch 20 / 25) train acc: 0.386000; val_acc: 0.274000
(Iteration 101 / 125) loss: 1.919649
(Epoch 21 / 25) train acc: 0.402000; val_acc: 0.272000
(Epoch 22 / 25) train acc: 0.440000; val_acc: 0.286000
(Epoch 23 / 25) train acc: 0.458000; val_acc: 0.295000
(Epoch 24 / 25) train acc: 0.462000; val_acc: 0.311000
(Epoch 25 / 25) train acc: 0.466000; val_acc: 0.297000
```

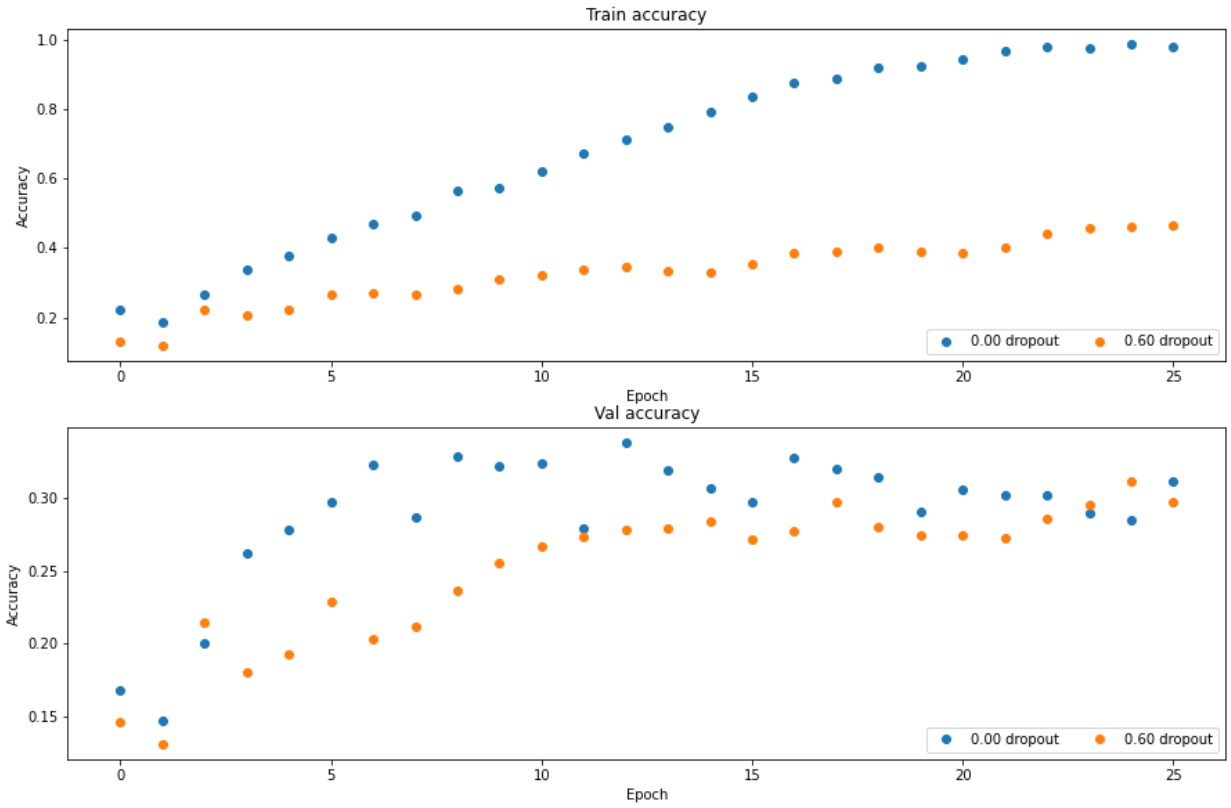
In [ ]:
```
# Plot train and validation accuracies of the two models

train_accs = []
val_accs = []
for dropout in dropout_choices:
  solver = solvers[dropout]
  train_accs.append(solver.train_acc_history[-1])
  val_accs.append(solver.val_acc_history[-1])

plt.subplot(3, 1, 1)
for dropout in dropout_choices:
  plt.plot(solvers[dropout].train_acc_history, 'o', label='%.2f dropout' % dropout)
plt.title('Train accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend(ncol=2, loc='lower right')

plt.subplot(3, 1, 2)
for dropout in dropout_choices:
  plt.plot(solvers[dropout].val_acc_history, 'o', label='%.2f dropout' % dropout)
plt.title('Val accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend(ncol=2, loc='lower right')

plt.gcf().set_size_inches(15, 15)
plt.show()
```



Question

Based off the results of this experiment, is dropout performing regularization? Explain your answer.

The results of the experiment show that dropout is performing regularization. We can see that the 0.00 dropout training model achieves training accuracy of nearly 1, but only a validation accuracy of about 0.3. This means that this model is overfit to the training data. On the contrary, we can see that the accuracy of the 0.6 dropout stays fairly consistent between the training and validation runs. This suggests that dropout is preventing the parameters from becoming too tailored to the trianing data and thus making theem more generalizable--this is regularization. It also prevents the model from rapidly reaching 100% accuracy in training.

Final part of the assignment

Get over 55% validation accuracy on CIFAR-10 by using the layers you have implemented. You will be graded according to the following equation:

```
min(floor((X - 32%)) / 28%, 1)
```

where if you get 60% or higher validation accuracy, you get full points.

In [ ]:

```
# ================================================================ #
# YOUR CODE HERE:
#   Implement a FC-net that achieves at least 55% validation accuracy
#   on CIFAR-10.
# ================================================================ #

optimizers = ['adam', 'rmsprop','sgd_nesterov_momentum']
best_optimizer_index = 1
best_model = None

layer_dims = [(200,200,200),(300,300,300),(400,400,400),(500, 500, 500),(600,600,600)
best_architecture_index =2 #1 also good on some runs
learning_rate = [1e-5,1e-4, 1e-3,1e-2]
best_learning_rate_index = 2
lr_decay = [0.8, 0.85, 0.9, 0.95, 0.99]
best_lr_decay_index = 1 #0 by exp
dropouts = [0.1, 0.2, 0.3, 0.4, 0.5]
best_dropout_index = 1
weight_scale = [0.01, 2/(np.mean(layer_dims[best_architecture_index])+3072), 1/np.mea
best_weight_index = 2
# for i,weight in enumerate(lr_decay):
best_model = FullyConnectedNet(layer_dims[best_architecture_index], weight_scale=weig
                           use_batchnorm=True,
                           dropout=dropouts[best_dropout_index])

solver = Solver(best_model, data,
                num_epochs=20, batch_size=100,
                update_rule=optimizers[best_optimizer_index],
                optim_config={
                  'learning_rate': learning_rate[best_learning_rate_index],
                },
                lr_decay=lr_decay[best_lr_decay_index],
                verbose=False)
solver.train()
# print('Model {} finished training...'.format(i))
print('Best Validation Accuracy: {}'.format(solver.best_val_acc))
# ================================================================ #
```

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

```
Best Validation Accuracy: 0.604
```

Note for the above result we are querying the inbuilt variable in the solver for the best validation accuracy.