Homework 8 - Artificial Neural Networks with PyTorch

About

In this homework, you will get your feet wet with deep learning using the PyTorch deep learning platform. This will involve:

- Preparing data
- Learning about the components of a deep learning pipeline
- Setting up a model, a loss function, and an optimizer
- Setting up training and testing loops
- Using a visualizer like tensorboard to monitor logged data

*This homework is due **April 15th 2019**. Training neural networks takes some time, particularly on CPUs so start early.*

Dev Environment

Working on Google Colab

You may choose to work locally or on Google Colaboratory. You have access to free compute through this service.

1. Visit https://colab.research.google.com/drive (https://colab.research.google.com/drive)
2. Navigate to the **`Upload`** tab, and upload your `HW8.ipynb`
3. Now on the top right corner, under the `Comment` and `Share` options, you should see a `Connect` option. Once you are connected, you will have access to a VM with 12GB RAM, 50 GB disk space and a single GPU. The dropdown menu will allow you to connect to a local runtime as well.

Notes:

- **If you do not have a working setup for Python 3, this is your best bet. It will also save you from heavy installations like `tensorflow` if you don't want to deal with those.**
- ***There is a downside*. You can only use this instance for a single 12-hour stretch, after which your data will be deleted, and you would have redownload all your datasets, any libraries not already on the VM, and regenerate your logs.**

Installing PyTorch and Dependencies

The instructions for installing and setting up PyTorch can be found at https://pytorch.org/get-started/locally/ (https://pytorch.org/get-started/locally/). Make sure you follow the instructions for your machine. For any of the remaining libraries used in this assignment:

- We have provided a `hw8_requirements.txt` file on the homework web page.
- Download this file, and in the same directory you can run `pip3 install -r hw8_requirements.txt`

Check that PyTorch installed correctly by running the following:

In [2]:
```
import torch
torch.rand(5, 3)
```

Out[2]:
```
tensor([[0.4335, 0.9857, 0.4848],
        [0.6282, 0.4110, 0.3560],
        [0.1606, 0.4377, 0.2903],
        [0.0763, 0.6127, 0.2210],
        [0.9140, 0.2318, 0.1950]])
```

The output should look something like

```
tensor([[0.3380, 0.3845, 0.3217],
        [0.8337, 0.9050, 0.2650],
        [0.2979, 0.7141, 0.9069],
        [0.1449, 0.1132, 0.1375],
        [0.4675, 0.3947, 0.1426]])
```

Let's get started with the assignment.

Instructions

Part 1 - Datasets and Dataloaders (10 points)

In this section we will download the MNIST dataset using PyTorch's own API.

Helpful Resources:

- https://pytorch.org/docs/stable/torchvision/datasets.html#mnist (https://pytorch.org/docs/stable/torchvision/datasets.html#mnist)
- https://pytorch.org/docs/stable/torchvision/transforms.html (https://pytorch.org/docs/stable/torchvision/transforms.html)
- https://pytorch.org/tutorials/beginner/data_loading_tutorial.html (https://pytorch.org/tutorials/beginner/data_loading_tutorial.html)

The `torchvision` package consists of popular datasets, model architectures, and common image transformations for computer vision. We are particularly concerned with `torchvision.datasets` and `torchvision.transforms`. Check out the API for these modules in the links provided above.

Create a directory named `hw8_data` with the following command.

In [2]:
```
!mkdir hw8_data
```

```
mkdir: hw8_data: File exists
```

Now use `torch.datasets.MNIST` to load the Train and Test data into `hw8_data` .

- ** Use the directory you created above as the `root` directory for your datasets**
- ** Populate the `transformations` variable with any transformations you would like to perform on your data.** (Hint: You will need to do at least one)
- **Pass your `transformations` variable to `torch.datasets.MNIST` . This allows you to perform arbitrary transformations to your data at loading time.**

In [38]:
```
from torchvision import datasets, transforms

transformations = transforms.Compose([
                                transforms.ToTensor(),
                                transforms.Normalize(
                                  (0.1000,), (0.3000,))
                              ])

mnist_train = datasets.MNIST(root='./hw8_data', train=True, download=True,
mnist_test = datasets.MNIST(root='./hw8_data', train=False, download=True,
```

```
0.00B [00:00, ?B/s]

Downloading http://yann.lecun.com/exdb/mnist/train-images-idx3-ubyte.gz
 (http://yann.lecun.com/exdb/mnist/train-images-idx3-ubyte.gz) to ./hw8_d
ata/MNIST/raw/train-images-idx3-ubyte.gz

9.92MB [00:06, 1.62MB/s]

Extracting ./hw8_data/MNIST/raw/train-images-idx3-ubyte.gz

32.8kB [00:00, 200kB/s]
0.00B [00:00, ?B/s]

Downloading http://yann.lecun.com/exdb/mnist/train-labels-idx1-ubyte.gz
 (http://yann.lecun.com/exdb/mnist/train-labels-idx1-ubyte.gz) to ./hw8_d
ata/MNIST/raw/train-labels-idx1-ubyte.gz
Extracting ./hw8_data/MNIST/raw/train-labels-idx1-ubyte.gz
Downloading http://yann.lecun.com/exdb/mnist/t10k-images-idx3-ubyte.gz (h
ttp://yann.lecun.com/exdb/mnist/t10k-images-idx3-ubyte.gz) to ./hw8_data/
MNIST/raw/t10k-images-idx3-ubyte.gz

1.65MB [00:01, 1.29MB/s]
  0%|          | 0.00/4.54k [00:00<?, ?B/s]

Extracting ./hw8_data/MNIST/raw/t10k-images-idx3-ubyte.gz
Downloading http://yann.lecun.com/exdb/mnist/t10k-labels-idx1-ubyte.gz (h
ttp://yann.lecun.com/exdb/mnist/t10k-labels-idx1-ubyte.gz) to ./hw8_data/
MNIST/raw/t10k-labels-idx1-ubyte.gz

8.19kB [00:00, 52.1kB/s]

Extracting ./hw8_data/MNIST/raw/t10k-labels-idx1-ubyte.gz
Processing...
Done!
```

Check that your torch datasets have been successfully downloaded into your data directory by running the next two cells.

- Each will output some metadata about your dataset.
- Check that the training set has 60000 datapoints and a `Root Location: hw8_data`
- Check that the testing (**also validation in our case**) set has 10000 datapoints and `Root Location: hw8_data`

Notice that these datasets implement the python `__len__` and `__getitem__` functions. Each element in the dataset should be a 2-tuple. What does yours look like?

In [40]:
```
print(len(mnist_train))
print(len(mnist_train[0]))
mnist_train
```

```
60000
2
```

Out[40]:
```
Dataset MNIST
    Number of datapoints: 60000
    Split: train
    Root Location: ./hw8_data
    Transforms (if any): Compose(
                             ToTensor()
                             Normalize(mean=(0.1,), std=(0.3,))
                         )
    Target Transforms (if any): None
```

In [39]:
```
print(len(mnist_test))
print(len(mnist_test[0]))
mnist_test
```

```
10000
2
```

Out[39]:
```
Dataset MNIST
    Number of datapoints: 10000
    Split: test
    Root Location: ./hw8_data
    Transforms (if any): Compose(
                             ToTensor()
                             Normalize(mean=(0.1,), std=(0.3,))
                         )
    Target Transforms (if any): None
```

Any file in our dataset will now be read at runtime, and the specified transformations we need on it will be applied when we need it..

We could iterate through these directly using a loop, but this is not idiomatic. PyTorch provides us with this abstraction in the form of `DataLoaders` . The module of interest is `torch.utils.data.DataLoader` .

`DataLoader` allows us to do lots of useful things

- Group our data into batches
- Shuffle our data
- Load the data in parallel using `multiprocessing` workers

Use `DataLoader` to create a loader for the training set and one for the testing set

- **Use a `batch_size` of 32 to start, you may change it if you wish.**
- **Set the `shuffle` parameter to `True`.**

In [41]:
```
from torch.utils.data import DataLoader

train_loader = torch.utils.data.DataLoader(dataset=mnist_train,
                                           batch_size=32,
                                           shuffle=True)
test_loader = torch.utils.data.DataLoader(dataset=mnist_test,
                                          batch_size=len(mnist_test),
                                          shuffle=False)

random_seed = 1
torch.backends.cudnn.enabled = False
torch.manual_seed(random_seed)
```

Out[41]:
```
<torch._C.Generator at 0x11461bf10>
```

The following function is adapted from `show_landmarks_batch` at [https://pytorch.org/tutorials/beginner/data_loading_tutorial.html#iterating-through-the-dataset (https://pytorch.org/tutorials/beginner/data_loading_tutorial.html#iterating-through-the-dataset)](https://pytorch.org/tutorials/beginner/data_loading_tutorial.html#iterating-through-the-dataset) .

Run the following cell to see that your loader provides a random `batch_size` number of data points.

In [42]:
```
import matplotlib.pyplot as plt
from torchvision import utils
%matplotlib inline

def show_mnist_batch(sample_batched):
    """Show images for a batch of samples."""
    images_batch = sample_batched[0]
    batch_size = len(images_batch)
    im_size = images_batch.size(2)

    grid = utils.make_grid(images_batch)
    plt.imshow(grid.numpy().transpose((1, 2, 0)))
    plt.title('Batch from DataLoader')

# Displays the first batch of images
for i, batch in enumerate(train_loader):
    if i==1:
        break
    show_mnist_batch(batch)
```

```
WARNING:matplotlib.image:Clipping input data to the valid range for imsho
w with RGB data ([0..1] for floats or [0..255] for integers).
```

Part 2 - Models, Loss Functions and Optimizers (10 points)

In this section, we will do the following:

- Learn about how to build your deep learning model and define its parameters
- Choose a loss function to optimize
- Choose an optimization method to maximize/minimize the loss

We'll first start with a single layer neural network to do handwritten digit classification. The math may ring some bells from homework 7.

`torch.nn` is the module we will be using here. You can find the API at https://pytorch.org/docs/stable/nn.html (https://pytorch.org/docs/stable/nn.html). There is also a quick summary at https://pytorch.org/tutorials/beginner/nn_tutorial.html#closing_thoughts (https://pytorch.org/tutorials/beginner/nn_tutorial.html#closing_thoughts).

Models

We will use the following python modules in building our one layer model.

- `torch.nn.Module` : Your model will be abstracted as a python class. Your python class must subclass `torch.nn.Module` . It is the base class for all neural network modules in PyTorch (Do not confuse python modules with PyTorch Modules). These implement the `forward()` function which defines how your model handles input and produces an output. Your model class can also have `torch.nn.Module` s as members, allowing nested tree like structures, and it is leveraging this that you are able to build neural networks in PyTorch.
- `torch.nn.Linear` : A unit of computation in neural networks are *Layers* and PyTorch provides abstractions for layers as `nn.Modules` . These come in many forms including *Convolutional*, *Recurrent*, and *Linear*. You can find the API for linear layers here [https://pytorch.org/docs/stable/nn.html#linear-layers (https://pytorch.org/docs/stable/nn.html#linear-layers)](https://pytorch.org/docs/stable/nn.html#linear-layers).

Now use the information provided to define the `OneLayerModel` class below. The superclass constructor has been called for you, and this allows your subclass to access superclass methods and members.

- **Finish the `__init__()` function.**
- **Finish the `forward()` function.** (Hint: Use that fact that layer modules implement their own `forward()` function)

In [43]:
```
from torch import nn
class OneLayerModel(nn.Module):
    def __init__(self, input_dim, output_dim):
        super(OneLayerModel, self).__init__()
        self.flin = nn.Linear(input_dim, output_dim)

    def forward(self, x):
        x = self.flin(x)
        return x
```

Loss Functions and Optimizers

You've defined your model but now what? It's just a black box that takes an input and spits out some numbers. You haven't yet defined what it means to be a good or bad model.

A ***Loss Function*** takes what your model outputs and compares it to what it *should* have put out. It returns some meaningful value used to update your model parameters, and so train your model. Check out Section 21.2.1 of the textbook for more details about types of loss functions. The Loss function represents the overall goal of building this model, and the choice of loss function is very important.

We must examine our model parameters and our problem instance to see about how to choose a loss function.

- We take in a 784-dimensional vector and output 10 real values, giving our model 784 x 10 parameters.
- It is natural given that our problem is an instance of *multi-class classification* that we would want each of our output values to model `P(y==i|x)` .

- If we go this route, we get an added constraint that the sum of all 10 of our output values should be 1 (forming a probability mass distribution).

Turns out there is a very convenient loss function for just our use case known as ***cross-entropy loss***. Check out this reference https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html#cross-entropy (https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html#cross-entropy) for a little more intuition on this.

Once again, PyTorch has abstractions built in for us in the `torch.nn` module, namely `torch.nn.CrossEntropyLoss`. The API can be found at https://pytorch.org/docs/stable/nn.html#crossentropyloss (https://pytorch.org/docs/stable/nn.html#crossentropyloss).

We're still not ready to train our model because while we have some parameters, and we have some measure of how good or bad our predictions are, we have no notion of how to go about updating our parameters in order to improve our loss.

This is where ***Optimizers*** come in. In general, we have one main way of minimizing loss functions (training our models), and that is through *Stochastic Gradient Descent* https://en.wikipedia.org/wiki/Stochastic_gradient_descent (https://en.wikipedia.org/wiki/Stochastic_gradient_descent). There are many variants and optimizations of this method, however, and the `torch.optim` package gives us abstractions for these. The API can be found at https://pytorch.org/docs/stable/optim.html# (https://pytorch.org/docs/stable/optim.html).

In [44]:
```
from torch import optim
```

Part 3 - Training and Validation (45 points)

In this section we will learn how to use the concepts we've learned about so far to train the model we built, and validate how well it does.We also want to monitor how well our training is going while it is happening.

For this we can use a package called `tensorboardX`. You will need to install this package using `pip` or `Anaconda`, based on your dev environment. Additionally, we'll want to use a logging module called `tensorboardX.SummaryWriter`. You can consult the API here https://tensorboardx.readthedocs.io/en/latest/tutorial.html (https://tensorboardx.readthedocs.io/en/latest/tutorial.html). Run the next cell to ensure that all is working well.

In [45]:
```
""" Try uncommenting these commands if you're facing issues here
!pip3 install -U protobuf
!pip3 install -U tensorflow
!pip3 install -U tensorboardX
"""
%load_ext tensorboard.notebook
from tensorboardX import SummaryWriter
```

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

We have provided the code to use `tensorboard` just before calling your `train` function. You don't have to change the top-level log directory, but you can create multiple runs (different parameters or versions of your code) just by creating subdirectories for these within your top-level directory.

Now use the information provided above to do the following:

- ** Instantiate a `OneLayerModel` with the appropriate input/output parameters.**
- ** Define a cross-entropy loss function.**
- ** Define a stochastic gradient descent optimizer based for you model's parameters. Start with a learning rate of 0.001, and adjust as necessary. You can start with the vanilla `optim.SGD` optimizer, and change it if you wish.**
- **Create a `SummaryWriter` object that will be responsible for logging our training progress into a directory called `logs/expt1` (Or whatever you wish your top-level directory to be called).**

In [60]:
```
model = OneLayerModel(1*28*28, 10)

# Loss and optimizer
loss = nn.CrossEntropyLoss()
learning_rate = 0.01
optimizer = optim.SGD(model.parameters(), lr=learning_rate, momentum = 0.5)
writer = SummaryWriter('logs/expt1')
```

We've finally come to the point where we need to write our training set up. We're going to use both our training and testing (validation) sets for this. Note that traditionally, you would separate part of your training data into validation data in order to get an unbiased estimate of how your model performs, but here we'll just pretend that our testing data is our validation data.

Training a model with batches of data broadly involves the following steps:

1. **One `epoch` is defined as a full pass of your dataset through your model. We choose the number of epochs we wish to train our model for.**
2. **In each epoch, set your model to train mode.**
3. **you feed your model `batch_size` examples at a time, and receive `batch_size` number of outputs until you've gotten through your entire dataset.**
4. **Calculate the loss function for those outputs given the labels for that batch.**
5. **Now calculate the gradients for each model parameter.** (Hint: Your loss function object can do this for you)
6. **Update your model parameters** (Hint: The optimizer comes in here)

7. **Set the gradients in your model to zero for the next batch.**
8. **After each epoch, set your model to evaluation mode.**
9. **Now evaluate your model on the validation data. Log the total loss and accuracy over the validation data.** (Note: PyTorch does automatic gradient calculations in the background through its `Autograd` mechanism https://pytorch.org/docs/stable/notes/autograd.html (https://pytorch.org/docs/stable/notes/autograd.html). Make sure to do evaluation in a context where this is turned off!)

Complete the `train()` function below. Try to make it as general as possible, so that it can be used for improved versions of you model. Feel free to define as many helper functions as needed. *Make sure that you do the following: *

- **Log the *training loss* and *training accuracy* on each batch for every epoch, such that it will show up on `tensorboard`.**
- **Log the loss on the validation set and the accuracy on the validation set every epoch**

You will need to produce the plots for these.

You may also want to add some print statements in your training function to report progress in this notebook.

In [55]:

```
def train(model, train_loader, val_loader, loss_func, optimizer,num_epochs=
    test(model, val_loader, loss_func, 0, writer)
    for epoch in range(1, num_epochs + 1):
        train_internal(model, train_loader, loss_func, optimizer, writer, e
        test(model, val_loader, loss_func, epoch, writer)


log_interval = 500

def train_internal(model, train_loader, loss_func, optimizer, writer, epoch
    model.train()
    for batch_id, (data, target) in enumerate(train_loader):
        data=data.reshape(len(data),-1)
        loss = loss_func(model(data), target)
        loss.backward()
        optimizer.step()
        optimizer.zero_grad()
        loss_item = loss.item()

        with torch.no_grad():
                output = model(data)
                predicted = torch.argmax(output, dim=1)
                train_accuracy = predicted.eq(target.data.view_as(predicted
                writer.add_scalars('Training', {'loss':loss_item,
                                                'accuracy': train_accuracy.item

                if batch_id % log_interval == 0:
                    print('Train Epoch: {} [{}/{} ({:.0f}%)]\tLoss: {:.6f}
                        epoch, batch_id * len(data), len(train_loader.datas
                        100. * batch_id / len(train_loader), loss_item, tra


def test(model, val_loader, loss_func, epoch_num, writer):
    model.eval()
    loss_item = 0
    correct = 0
    with torch.no_grad():
        for data, target in val_loader:
            data=data.reshape(len(data),-1)
            output = model(data)
            loss_item += loss_func(output, target)
            pred = torch.argmax(output, dim=1)
            correct += pred.eq(target.data.view_as(pred)).sum()

        accuracy = 100. * correct.item()/ len(val_loader.dataset)
        loss = loss_item.item()/len(val_loader)
        writer.add_scalar('Validation set loss', loss, epoch_num)
        writer.add_scalar('Validation set accuracy', accuracy, epoch_num)

        print('\nTest set: Avg. loss: {:.4f}, Accuracy: {}/{} ({:.2f}%)\n'.
                loss, correct, len(val_loader.dataset),accuracy))
```

Finally call `train` with the relevant parameters. Run the tensorboard command on your top-level logs directory to monitor training. If there is logging data from a previous run, just delete the directory for the run, and reinstantiate the `SummaryWriter` for that run. (You may want to reinstantiate the model itself if you want to clear the model parameters too).

Note : This function may take a while to complete if you're training for many epochs on a cpu. This is where it comes in handy to be running on Google Colab, or just have a GPU on hand.

In [61]:
```
#%tensorboard --logdir=logs

train(model, train_loader, test_loader, loss, optimizer, 15, writer)
```

```
Train Epoch: 13 [16000/60000 (27%)]     Loss: 0.129274 Accuracy: 0.968750
Train Epoch: 13 [32000/60000 (53%)]     Loss: 0.103068 Accuracy: 1.000000
Train Epoch: 13 [48000/60000 (80%)]     Loss: 0.136728 Accuracy: 0.968750

Test set: Avg. loss: 0.2785, Accuracy: 9253/10000 (92.53%)

Train Epoch: 14 [0/60000 (0%)]  Loss: 0.404441 Accuracy: 0.906250
Train Epoch: 14 [16000/60000 (27%)]     Loss: 0.233375 Accuracy: 0.906250
Train Epoch: 14 [32000/60000 (53%)]     Loss: 0.214349 Accuracy: 0.968750
Train Epoch: 14 [48000/60000 (80%)]     Loss: 0.361963 Accuracy: 0.906250

Test set: Avg. loss: 0.2769, Accuracy: 9236/10000 (92.36%)

Train Epoch: 15 [0/60000 (0%)]  Loss: 0.243817 Accuracy: 0.937500
Train Epoch: 15 [16000/60000 (27%)]     Loss: 0.804453 Accuracy: 0.843750

Train Epoch: 15 [32000/60000 (53%)]     Loss: 0.378692 Accuracy: 0.906250
Train Epoch: 15 [48000/60000 (80%)]     Loss: 0.203313 Accuracy: 0.968750

Test set: Avg. loss: 0.2703, Accuracy: 9252/10000 (92.52%)
```

Final Validation Loss: *0.2722*

Final Validation Accuracy: *92.53%*

What is familiar about a 1-layer neural network with cross-entopy loss? Have you seen this before?

Answer: SVM model has similar linear function and uses SGD.

Part 4 - Two Layer Neural Net (20 points)

The thing that makes neural networks really powerful is that they are able to do complex function approximation. As we saw earlier, we can organize the computation done in neural networks into units called *layers*. In a general neural network, there is an *input layer*, and an *output layer*. These may be the same layer as they were in our previous example. When they are not the same, there are intermediate layers known as *hidden layers*. These layers receive input from other layers and send their output to other layers.

We have been dealing with a certain type of neural network known as a **fully connected** network. For our purposes, this just means that the output of the layer is just the dot product of its input `x`, its weights `w` plus a bias term `b`, all wrapped in a non-linear *activation function* `F`.

```
y = F(w^T x + b) .
```

These non-linear activation functions are very important but where in our last neural network did we apply such a function? Implicitly we applied what's known as a **softmax activation** in order to compute cross-entropy loss https://en.wikipedia.org/wiki/Softmax_function (https://en.wikipedia.org/wiki/Softmax_function).

We'll now try to create a neural network with one hidden layer. This means that we have to come up with an activation function for the output of that hidden layer. A famous, simple but powerful activation function is the **Rectified Linear Unit (ReLU)** function defined nas `ReLU(x) = max(x,0)` . We will use this on the output of the hidden layer.

`torch.nn` has a module known as `nn.Sequential` that allows us to chain together other modules. This module implements a `forward()` function that automatically handles input-output connections etc. Check out the API at https://pytorch.org/docs/stable/nn.html#sequential (https://pytorch.org/docs/stable/nn.html#sequential).

Just like you did with the single layer model, define a class `TwoLayerModel` , a neural network with ReLU activation for the hidden layer. `nn.Sequential` may come in handy.

In [62]:
```python
import torch.nn.functional as F

class TwoLayerModel(nn.Module):
    def __init__(self, input_dim, output_dim, hidden_layers):
        super(TwoLayerModel, self).__init__()
        self.relu = nn.ReLU(inplace=True)
        #self.conv1 = nn.Conv2d(1, 20, kernel_size=5)
        self.flin1 = nn.Linear(input_dim, 256)
        self.flin2 = nn.Linear(256, output_dim)

    def forward(self, x):
        #x = self.relu(F.max_pool2d(self.conv1(x), 2))
        x = self.relu(self.flin1(x))
        x = self.flin2(x)
        return x
```

Once again use the information provided above to do the following:

- ** Instantiate a `TwoLayerModel` with the appropriate input/output/hidden layer parameters.**
- ** Define a cross-entropy loss function again.**
- ** Define a stochastic gradient descent optimizer based for you model's parameters. Start with a learning rate of 0.001, and adjust as necessary. You can start with the vanilla `optim.SGD` optimizer, and change it if you wish.**
- **Create a `SummaryWriter` object that will be responsible for logging our training progress into a directory called `logs/expt2` (Or whatever you wish your top-level directory to be called, just make sure the subdirectory is different from your previous SummaryWriter).**

In [63]:
```
model2 = TwoLayerModel(1*28*28, 10, 256)
learning_rate=0.01
# Loss and optimizer
loss2 = nn.CrossEntropyLoss()
optimizer2 = optim.SGD(model2.parameters(), lr=learning_rate, momentum = 0.
writer2 = SummaryWriter('logs/expt2')
```

Call `train` on your two layer neural network.

In [64]:
```
#%tensorboard --logdir=logs

train(model2, train_loader, test_loader, loss2, optimizer2, 15, writer2)
```

```
Train Epoch: 13 [16000/60000 (27%)]     Loss: 0.004959 Accuracy: 1.000000
Train Epoch: 13 [32000/60000 (53%)]     Loss: 0.061009 Accuracy: 1.000000
Train Epoch: 13 [48000/60000 (80%)]     Loss: 0.040152 Accuracy: 1.000000

Test set: Avg. loss: 0.0629, Accuracy: 9796/10000 (97.96%)

Train Epoch: 14 [0/60000 (0%)]  Loss: 0.049145 Accuracy: 1.000000
Train Epoch: 14 [16000/60000 (27%)]     Loss: 0.008735 Accuracy: 1.000000
Train Epoch: 14 [32000/60000 (53%)]     Loss: 0.015478 Accuracy: 1.000000
Train Epoch: 14 [48000/60000 (80%)]     Loss: 0.051917 Accuracy: 1.000000

Test set: Avg. loss: 0.0651, Accuracy: 9801/10000 (98.01%)


Train Epoch: 15 [0/60000 (0%)]  Loss: 0.014761 Accuracy: 1.000000
Train Epoch: 15 [16000/60000 (27%)]     Loss: 0.004340 Accuracy: 1.000000
Train Epoch: 15 [32000/60000 (53%)]     Loss: 0.016982 Accuracy: 1.000000
Train Epoch: 15 [48000/60000 (80%)]     Loss: 0.010811 Accuracy: 1.000000

Test set: Avg. loss: 0.0617, Accuracy: 9815/10000 (98.15%)
```

Final Validation Loss: *0.0618*

Final Validation Accuracy: *98.11%*

Did your accuracy on the validation set improve with multiple layers? Why do you think this is ?

Answer: The problem itself is not linear. Most of the digits' features are not linearly separable. That is why there is a 6% accuracy increase when using 2 linear layers or if adding a third one with convolution and max pooling the accuracy increases to 98% (6% more than using a single layer).

Part 5 - What is being learned at each layer? (10 points)

So what exactly are these weights that our network is learning at each layer? By conveniently picking our layer dimensions as perfect square numbers, we can try to visualize the weights learned at each layer as square images. Use the following function to do so for *all interesting layers* across your models. Feel free to modify the function as you wish.

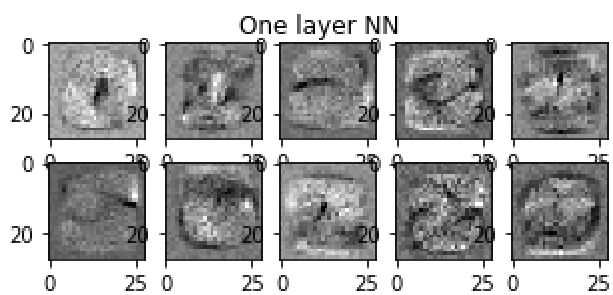
At the very least, you must generate:

1. **The ten 28x28 weight images learned by your one layer model.**
2. **The 256 28x28 weight images learned by the hidden layer in your two-layer model.**

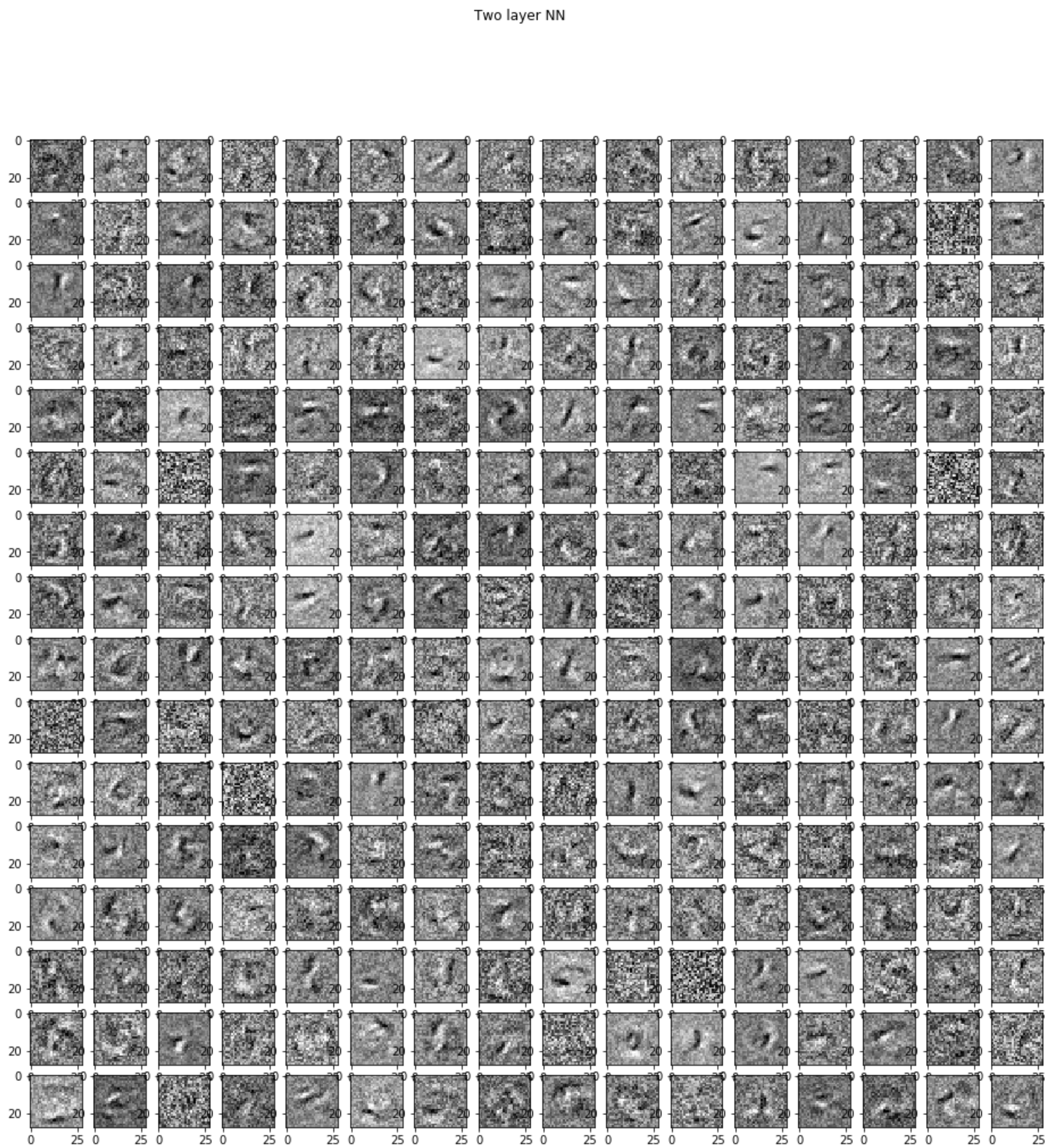
In [65]:
```
def visualize_layer_weights(model, layer_idx, num_images, image_dim, title)
    # Find number of rows and columns based on number of images
    for d in range(1,num_images):
        f = num_images/d
        if int(f)==f:
            dim1 = int(min(f,d))
            dim2 = int(max(f,d))
        if d > f:
            break
    # Plot weights as square images
    fig, ax  = plt.subplots(dim1, dim2)

    # At least 1 inch by 1 inch images
    fig.set_size_inches(dim2, dim1)
    weights = (list(model.parameters())[layer_idx])
    fig.suptitle(title)
    for i in range(dim1):
        for j in range(dim2):
            item = weights[dim2*i+j]
            ax[i][j].imshow(item.reshape(image_dim,image_dim).detach().nump

visualize_layer_weights(model, 0,10,28,'One layer NN')
```



In [66]:
```
visualize_layer_weights(model2, 0,256,28,'Two layer NN')
```

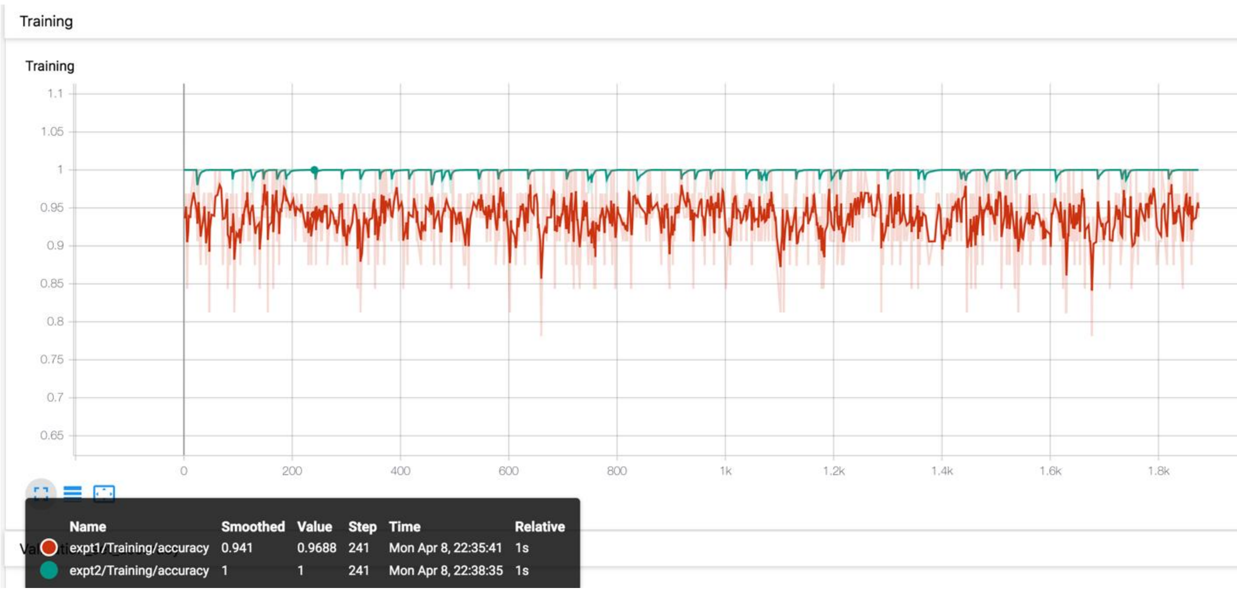


In [ ]:

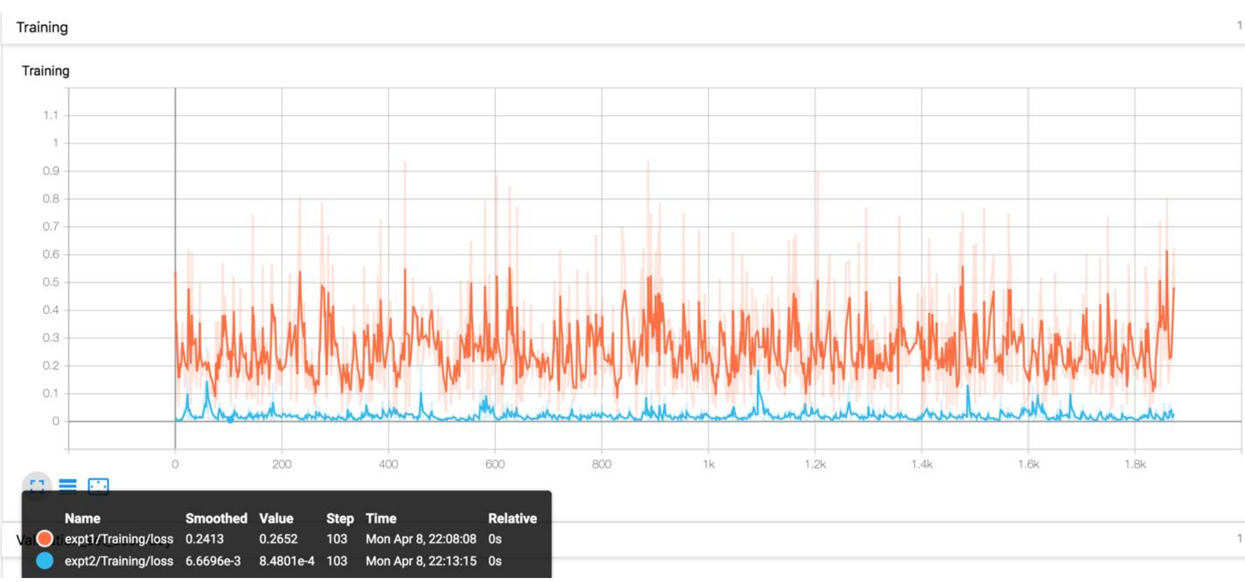
Plots of Accuracy and loss on training and validation sets of both Ona layer and Two layers NNs:

The expt1's folder data represents the logs from single layer network, expt2 has these for 2-layer NN.

On the first graph we can see the differences in terms of training accuracy of both models.



Respectively the loss of the second model is also lower than the loss of the first one.



Validation accuracy and loss confirm that model 2 always performs better than model 1. Mainly because the problem is not linear and has multiple features that can be explained better with multi-layer neural network.