CS 273: Machine Learning Final Project

Julius C Aguma (38208345), Matt Dees (30281707), James

December 2019

1 Introduction

2 Dataset

3 Neural Network Models

To analyse the capacity of neural network models on the fashion-mnist data, we chose to use a feedforward and convolutional neural network. While holding all other parameters constant, the number of hidden layers, k was varied for values $1, 2, 4, 8$. Results of this analysis are presented below with a summary of the architectures of both models.

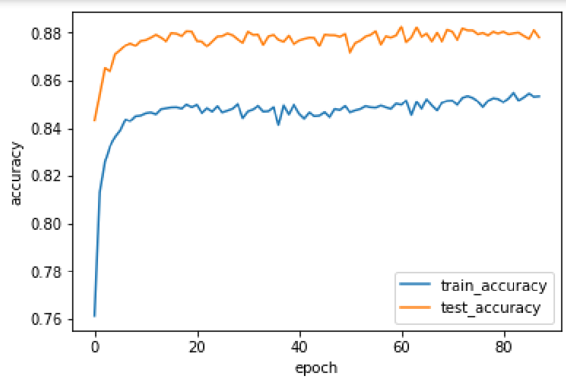
3.1 FeedForward model

3.1.1 Architecture

Using tensorflow with keras, we built a feedforward network that takes fashion mnist data as input read through the mnist-reader.py script. The network has k hidden layers where k is varied for an analysis of capacity, that is a comparison of deep vs shallow networks on the fashion-mnist dataset. The data is split into training and validation data as shown in the model parameters below and the model is trained for 50 epochs where each epoch has 40000 iterations. In addition we use dropout factors less than 1 for regularization and cross entropy as loss function with the adadelta optimizer as a learning rate method for our gradient descent. The hidden layers use relu for an activation function while the output layer uses softmax. All inputs are normalized. Below are parameters and results for each k in the set $1, 2, 4, 8$

1. K = 1

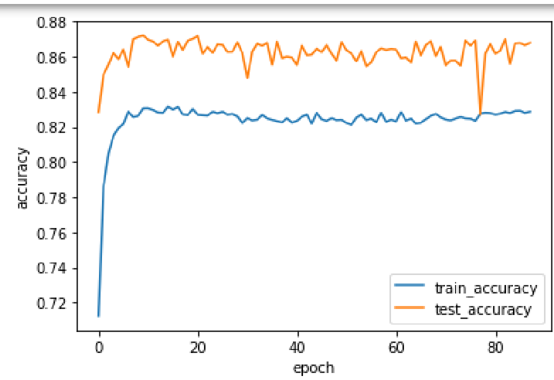
```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_3 (InputLayer)         (None, 784)               0         
_________________________________________________________________
batch_normalization_2 (Batch (None, 784)               3136      
_________________________________________________________________
dropout_4 (Dropout)          (None, 784)               0         
_________________________________________________________________
dense_4 (Dense)              (None, 128)               100480    
_________________________________________________________________
dropout_5 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_5 (Dense)              (None, 10)                1290      
=================================================================
Total params: 104,906
Trainable params: 103,338
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```




```
10000/10000 [==============================] - 0s 37us/step
Out[2]: [0.40412045266628266, 0.8688]
```

2. K = 2

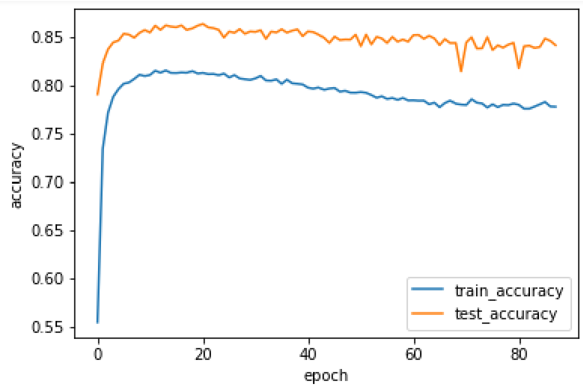
```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_4 (InputLayer)         (None, 784)               0         
_________________________________________________________________
batch_normalization_3 (Batch (None, 784)               3136      
_________________________________________________________________
dropout_6 (Dropout)          (None, 784)               0         
_________________________________________________________________
dense_6 (Dense)              (None, 128)               100480    
_________________________________________________________________
dropout_7 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_7 (Dense)              (None, 256)               33024     
_________________________________________________________________
dropout_8 (Dropout)          (None, 256)               0         
_________________________________________________________________
dense_8 (Dense)              (None, 10)                2570      
=================================================================
Total params: 139,210
Trainable params: 137,642
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```




```
10000/10000 [==============================] - 0s 35us/step
Out[4]: [0.46054650063514707, 0.8587]
```

3. $K = 4$

```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_5 (InputLayer)         (None, 784)               0
_________________________________________________________________
batch_normalization_4 (Batch (None, 784)               3136
_________________________________________________________________
dropout_9 (Dropout)          (None, 784)               0
_________________________________________________________________
dense_9 (Dense)              (None, 128)               100480
_________________________________________________________________
dropout_10 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_10 (Dense)             (None, 256)               33024
_________________________________________________________________
dropout_11 (Dropout)         (None, 256)               0
_________________________________________________________________
dense_11 (Dense)             (None, 128)               32896
_________________________________________________________________
dropout_12 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_12 (Dense)             (None, 128)               16512
_________________________________________________________________
dropout_13 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_13 (Dense)             (None, 10)                1290
=================================================================
Total params: 187,338
Trainable params: 185,770
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```

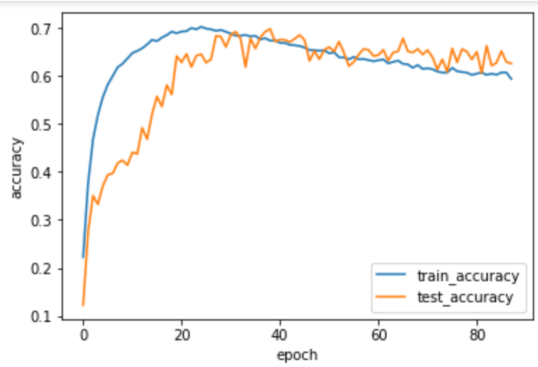



```
10000/10000 [==============================] - 1s 54us/step
```

Out[5]:
```
[0.549263471364975, 0.8291]
```

4. K = 8

```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_6 (InputLayer)         (None, 784)               0         
_________________________________________________________________
batch_normalization_5 (Batch (None, 784)               3136      
_________________________________________________________________
dropout_14 (Dropout)         (None, 784)               0         
_________________________________________________________________
dense_14 (Dense)             (None, 128)               100480    
_________________________________________________________________
dropout_15 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_15 (Dense)             (None, 256)               33024     
_________________________________________________________________
dropout_16 (Dropout)         (None, 256)               0         
_________________________________________________________________
dense_16 (Dense)             (None, 128)               32896     
_________________________________________________________________
dropout_17 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_17 (Dense)             (None, 128)               16512     
_________________________________________________________________
dropout_18 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_18 (Dense)             (None, 128)               16512     
_________________________________________________________________
dropout_19 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_19 (Dense)             (None, 128)               16512     
_________________________________________________________________
dropout_20 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_20 (Dense)             (None, 128)               16512     
_________________________________________________________________
dropout_21 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_21 (Dense)             (None, 128)               16512     
_________________________________________________________________
dropout_22 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_22 (Dense)             (None, 10)                1290      
=================================================================
Total params: 253,386
Trainable params: 251,818
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```




```
10000/10000 [==============================] - 1s 67us/step
```

Out[6]: [1.0562932844161987, 0.6197]

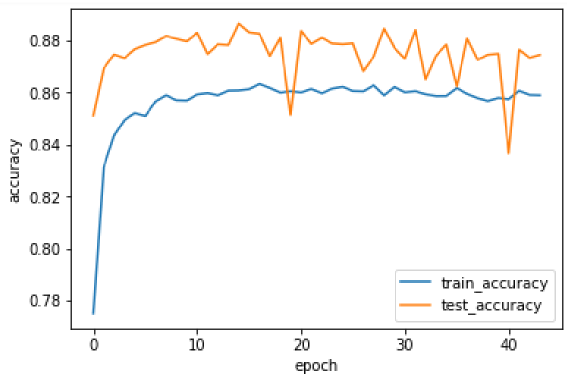
3.2 Convolutional model

3.2.1 Architecture

The convolutional model(cnn) is similar to the feedforward model in number of hidden layers, optimizer, regularization, activation functions, and validation. However, the cnn also has 2 3x3 convolutional filters and a 2x2 max pooling layer for downsampling the convolution layers. Results follow below.

1. K = 1

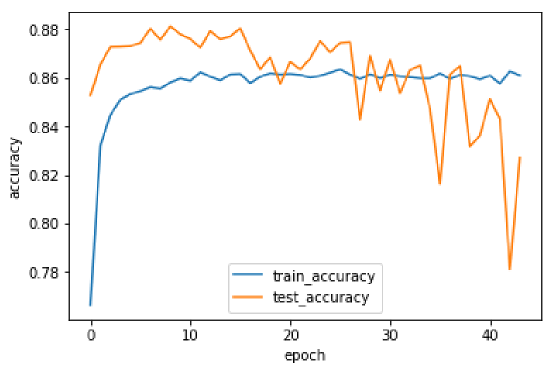
```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_5 (InputLayer)         (None, 784)               0         
_________________________________________________________________
batch_normalization_4 (Batch (None, 784)               3136      
_________________________________________________________________
dropout_15 (Dropout)         (None, 784)               0         
_________________________________________________________________
reshape_4 (Reshape)          (None, 28, 28, 1)         0         
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 26, 26, 32)        320       
_________________________________________________________________
dropout_16 (Dropout)         (None, 26, 26, 32)        0         
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 13, 13, 32)        0         
_________________________________________________________________
flatten_4 (Flatten)          (None, 5408)              0         
_________________________________________________________________
dense_7 (Dense)              (None, 128)               692352    
_________________________________________________________________
dropout_17 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_8 (Dense)              (None, 10)                1290      
=================================================================
Total params: 697,098
Trainable params: 695,530
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```




```
10000/10000 [==============================] - 2s 218us/step
Out[5]: [0.40157343983650207, 0.8689]
```

2. K = 2

```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_3 (InputLayer)         (None, 784)               0         
_________________________________________________________________
batch_normalization_2 (Batch (None, 784)               3136      
_________________________________________________________________
dropout_6 (Dropout)          (None, 784)               0         
_________________________________________________________________
reshape_2 (Reshape)          (None, 28, 28, 1)         0         
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 26, 26, 32)        320       
_________________________________________________________________
dropout_7 (Dropout)          (None, 26, 26, 32)        0         
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 24, 24, 32)        9248      
_________________________________________________________________
dropout_8 (Dropout)          (None, 24, 24, 32)        0         
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 12, 12, 32)        0         
_________________________________________________________________
flatten_2 (Flatten)          (None, 4608)              0         
_________________________________________________________________
dense_2 (Dense)              (None, 128)               589952    
_________________________________________________________________
dropout_9 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_3 (Dense)              (None, 10)                1290      
=================================================================
Total params: 603,946
Trainable params: 602,378
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```

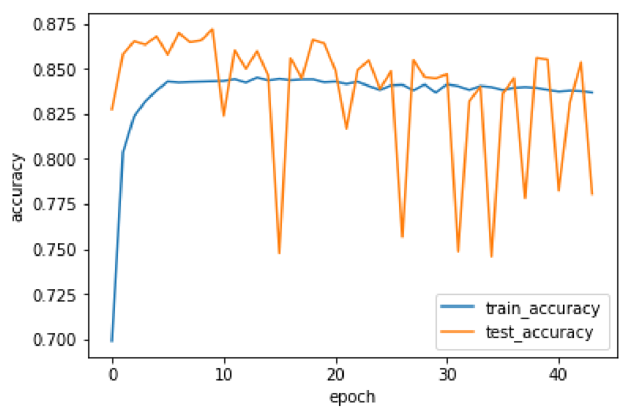



```
10000/10000 [==============================] - 4s 440us/step
```

Out[3]: [0.5941237535476684, 0.8212]

3. K = 4

```
x_train shape: (60000, 784) y_train shape: (60000,)
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_4 (InputLayer)         (None, 784)               0
_________________________________________________________________
batch_normalization_3 (Batch (None, 784)               3136
_________________________________________________________________
dropout_10 (Dropout)         (None, 784)               0
_________________________________________________________________
reshape_3 (Reshape)          (None, 28, 28, 1)         0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 26, 26, 32)        320
_________________________________________________________________
dropout_11 (Dropout)         (None, 26, 26, 32)        0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 24, 24, 32)        9248
_________________________________________________________________
dropout_12 (Dropout)         (None, 24, 24, 32)        0
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 12, 12, 32)        0
_________________________________________________________________
flatten_3 (Flatten)          (None, 4608)              0
_________________________________________________________________
dense_4 (Dense)              (None, 128)               589952
_________________________________________________________________
dropout_13 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_5 (Dense)              (None, 128)               16512
_________________________________________________________________
dropout_14 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_6 (Dense)              (None, 10)                1290
=================================================================
Total params: 620,458
Trainable params: 618,890
Non-trainable params: 1,568
_________________________________________________________________
Train on 40199 samples, validate on 19801 samples
```




```
10000/10000 [==============================] - 5s 460us/step
```

```
4]: [0.7686937316894531, 0.7717]
```

3.3 Discussion of results

The feedforward model gave more consistent predictions with a smoother curve while the cnn gave us higher accuracy with a peak of about 90%. For both architectures, the rate of overfitting increased with increase in the number of hidden layers. More so for the cnn where by k = 4 we see an overwhelming amount of overfitting. We also see that the peak accuracy never increased with increase in number of hidden layers. Further proving the known fact that capacity of a neural network doesnot increase with deeper networks but only the type of functions that can be estimated improves. A good next step in the capacity study would be to look at how the network performance changes with wider hidden layer, that is hidden layers with more neurons.