1 CS 5644: Assignment 4

1.1 Question 1 (90 points) Multilayer Perceptron Classifier

Consider the MNIST dataset, a downsampled portion of which is provided alongside this document. The data consists of a set of handwritten digits (0-9) that are downsampled to 14x14 pixels. The first column represents the digit value (0-9) and the remaining columns represent the pixels. Just like the previous assignment, spend some time understanding the structure of the dataset, how the instances are organized, how the features are organized, what the various features mean, and so on. Do not attempt to run any machine learning algorithm before understanding the structure of the dataset. A portion of this data can be viewed using the following code:

```python
from matplotlib import pyplot as plt
fig = plt.figure(figsize=(16,8))

for i in range(10):
    for j in range(y.shape[0]):
        if y[j] == i:
            ax = fig.add_subplot(2, 5, i+1)
            ax.imshow(X[i].reshape(14,14), cmap=plt.cm.gray_r)
            break
```

The assignment is to create a MLP classifier that can correctly determine the digits in the dataset. The classifier object should use the arguments of `max_iter=1000` and `solver=lbfgs`. You will need to use `MinMaxScaler` to scale the data for optimum results. Implement different MLP architectures for the following scenarios and report the precision, recall, and fscore for 5-fold cross validation:

1. Layers:
    - (a) Use 2 hidden layers with 20 nodes in each layer.
    - (b) Use 2 hidden layers with 100 nodes in each layer.
    - (c) Use 5 hidden layers with 20 nodes in each layer.
    - (d) Use 5 hidden layers with 100 nodes in each layer.
2. Activation Function:
    - (a) Use 'relu'.
    - (b) Use 'tanh'.

Alongside these metrics you should observe the folowing:

1. The time it takes to fit each model as size of layers increase for various layer sizes (recommend number of layers = 2 and size of layers = [20, 40, 60, 80, 100]). Provide graphs.
2. The time it takes to fit each model as number of layers increase for various numbers of layers (recommend number of layers = [1, 2, 3, 4, 5] and size of layers = 20). Provide graphs.
3. What type of growth do you see for each? Why?
4. What happens in simple/smallish network if you stop using `MinMaxScaler`. Why?

Write a report on your findings including answers to each question above and graphs where applicable.

For measuring time spent, the following code will be useful:

```python
import time
```

```
start = time.time()
# do something
time_spent = time.time() - start
```

1.2 Question 2 (10 points)

In the scikit learn package, the default value for the learning rate of an MLP is 0.001. Explain what will happen to the classification result if we set this parameter to 0.5 and why.

1.3 What to submit:

A zipped file containing:

1. a PDF document summarizing answers to questions 1 and 2. Distill your lessons and experiences succinctly.
2. Either hyperlinks to or actual attachments of your data files and your iPython notebook(s).