Your submission for Part 3 should be a zip file that includes;
1. a pdf with 2sections, - YES
2. the file for the model of the best network found and
3. the .arff file that includes data set 2

Programming Assignment 5 - Part 3

Section 1

Oleg Tielushko, Tyler Brown.
Then It should include the results that WEKA provides when, using data set 1 to train the network selecting “use training set”option in the “test options” window, you experiment trying to find the “best” network by varying the number of neurons in the hidden layers, varying the number of epochs, and increasing the number of hidden layers.

You should at least try up to 3 hidden layers. For each of 3 different networks that you build searching for the “best”, include the following items.

Network 1 Information

Percentage of data used as a training set: 80%, as a testing set 20%
Noticeable change: the number of training epochs increased from 500 to 2000.

a. Number of layers in the network and number of neurons in each layer
   Input Layer: Attributes (16 nodes due to 16 attributes of each letter).
   Hidden Layer: Node 26 through Node 46 (21 nodes total)
   Output Layer: Node 0 through Node 25 (26 nodes total)

b. The overall accuracy achieved by the network on the training set (data set 1) and on the test set (data set 2).
   Training Set Accuracy: 84.8125%

```
Correctly Classified Instances       13570               84.8125 %
Incorrectly Classified Instances      2430               15.1875 %
Kappa statistic                          0.842
Mean absolute error                      0.0135
Root mean squared error                  0.1018
Relative absolute error                 18.2321 %
Root relative squared error             52.9541 %
Total Number of Instances            16000
```

Testing Set Accuracy: 82.775%

```
=== Summary ===

Correctly Classified Instances        3311               82.775  %
Incorrectly Classified Instances       689               17.225  %
Kappa statistic                          0.8208
Mean absolute error                      0.0148
Root mean squared error                  0.1074
Total Number of Instances            4000
```

c. The accuracy by which each letter was classified.

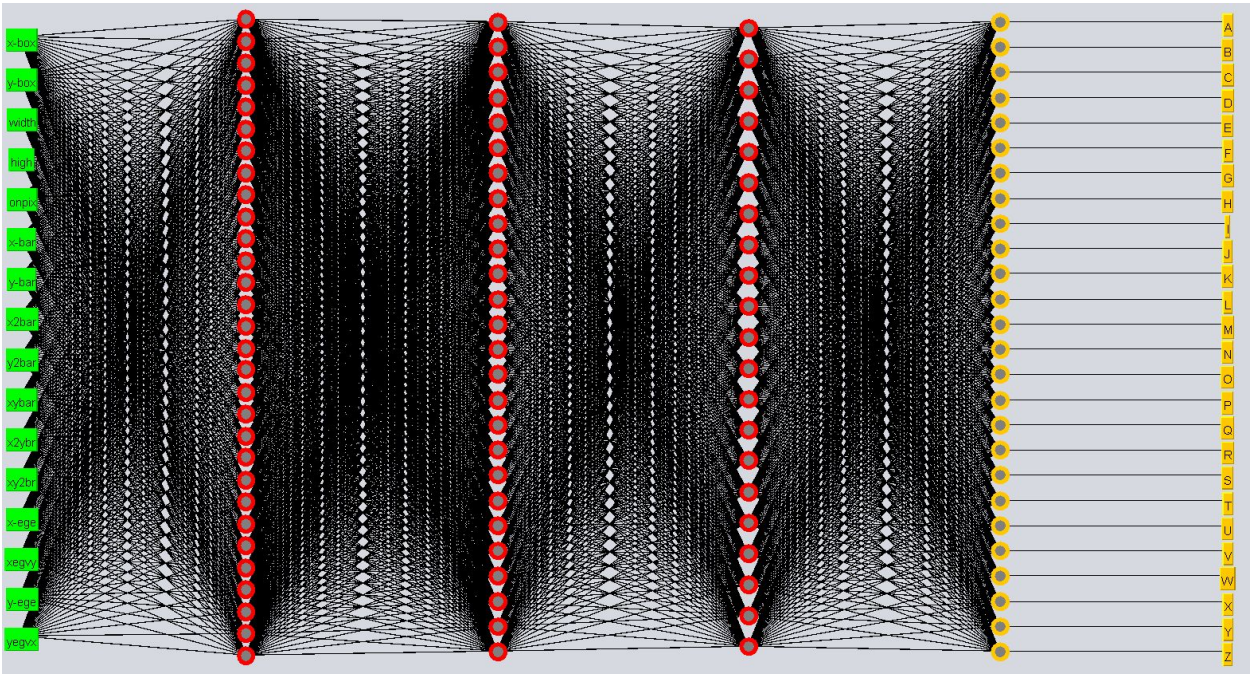
Letter	Accuracy Training Set	Accuracy Testing Set
A	88.9%	89.8%
B	95.7%	91.9%
C	88.9%	84.9%
D	85.2%	83.4%
E	82.5%	72.7%
F	86.7%	82.8%
G	64.8%	79.4%
H	72.9%	63.9%
I	86.2%	86.4%
J	78.8%	74.5%
K	83.5%	73.3%

L	83.3%	91.4%
M	93.0%	92.0%
N	87.8%	82.0%
O	84.7%	81.0%
P	87.4%	82.8%
Q	82.6%	76.4%
R	82.9%	83.8%
S	69.6%	58.7%
T	79.3%	79.1%
U	90.9%	90.2%
V	88.6%	89.5%
W	92.4%	94.1%
X	91.7%	91.8%
Y	92.7%	92.9%
Z	81.7%	81.3%

Network 2 Information

Percentage of data used as a training set: 80%, as a testing set 20%
Noticeable change: the number of training epochs increased from 500 to 1000. Added additional hidden layers (details in part a.)

a. The number of layers in the network and the number of neurons in each layer:
   Input Layer: Attributes (16 nodes due to 16 attributes of each letter).
   Hidden Layer 1: 30 nodes total.
   Hidden Layer 2: 26 nodes total.
   Hidden Layer 3: 21 nodes total.
   Output Layer: 26 nodes total.



b. The overall accuracy achieved by the network on the training set (data set 1) and on the test set (data set 2).
Training Set Accuracy: 88.9688%

```
=== Summary ===

Correctly Classified Instances       14235               88.9688 %
Incorrectly Classified Instances      1765               11.0313 %
Kappa statistic                          0.8853
Mean absolute error                      0.0099
Root mean squared error                  0.0859
Relative absolute error                 13.4146 %
Root relative squared error             44.6512 %
Total Number of Instances            16000
```

Testing Set Accuracy: 86.85%

```
=== Summary ===

Correctly Classified Instances        3474               86.85   %
Incorrectly Classified Instances       526               13.15   %
Kappa statistic                          0.8632
Mean absolute error                      0.0115
Root mean squared error                  0.0937
Total Number of Instances             4000
```

c. The accuracy by which each letter was classified. Use a format similar to what you used in sections Parts 1 and 2.

Letter	Accuracy Training Set	Accuracy Testing Set
A	89.6%	89.2%
B	89.3%	86.3%
C	80.3%	71.1%
D	90.5%	85.4%
E	93.1%	88.2%
F	85.8%	82.2%
G	74.1%	83.5%
H	78.1%	69.6%
I	82.8%	92.0%
J	90.5%	85.9%
K	93.2%	85.3%
L	84.2%	90.8%
M	93.9%	92.8%
N	95.3%	86.2%
O	94.5%	92.2%

P	92.6%	91.7%
Q	81.3%	82.7%
R	87.1%	85.0%
S	83.6%	74.1%
T	84.6%	86.9%
U	95.7%	93.9%
V	90.7%	88.2%
W	96.3%	95.6%
X	93.2%	91.2%
Y	96.7%	93.6%
Z	94.7%	94.7%

Network 3 Information

Percentage of data used as a training set: 80%, as a testing set 20%
Noticeable change: the number of training epochs increased from 500 to 6560. Added additional hidden layers (more details in part a).

a. The number of layers in the network and the number of neurons in each layer
   Input Layer: Attributes (16 nodes due to 16 attributes of each letter).
   Hidden Layer 1: 41 nodes total.
   Hidden Layer 2: 31 nodes total.
   Hidden Layer 3: 26 nodes total.
   Output Layer: 26 nodes total.

b. The overall accuracy achieved by the network on the training set (data set 1) and on the test set (data set 2).

Training Set Accuracy: 90.8063%

```
=== Summary ===

Correctly Classified Instances       14529               90.8063 %
Incorrectly Classified Instances      1471                9.1937 %
Kappa statistic                          0.9044
Mean absolute error                      0.0081
Root mean squared error                  0.0801
Relative absolute error                 10.8859 %
Root relative squared error             41.6531 %
Total Number of Instances            16000
```

Testing Set Accuracy: 85.575%

```
=== Summary ===

Correctly Classified Instances        3423               85.575  %
Incorrectly Classified Instances       577               14.425  %
Kappa statistic                          0.85
Mean absolute error                      0.0117
Root mean squared error                  0.0991
Total Number of Instances            4000
```

c. The accuracy by which each letter was classified. Use a format similar to what you used in sections Parts 1 and 2.

Letter	Accuracy Training Set	Accuracy Testing Set
A	93.1%	93.4%
B	91.7%	85.0%
C	71.1%	62.5%
D	94.6%	83.4%
E	88.0%	78.3%
F	82.8%	75.8%
G	82.8%	80.0%

H	77.8%	65.8%
I	92.1%	92.6%
J	95.0%	89.9%
K	94.2%	82.7%
L	85.6%	88.8%
M	95.0%	94.9%
N	95.0%	89.2%
O	93.3%	88.2%
P	88.8%	85.5%
Q	88.9%	90.6%
R	89.8%	82.5%
S	92.9%	77.6%
T	95.8%	89.5%
U	95.7%	93.3%
V	91.3%	87.5%
W	96.1%	94.8%
X	94.5%	88.8%
Y	95.7%	92.9%
Z	97.3%	93.3%

d.At the end of the section, include a paragraph discussing the results obtained.The discussion should show depth of understanding of the process of building neural networks.

For the creation of the first network, our group decided to increase the number of epochs in the training cycle. While having more than an automatic number of training epochs allowed for a slight increase in the accuracy in the network, it has proven to not be a complete success after all, as the overall accuracy improved just by 1 percent.

For the second network, our group decided to increase the number of epochs to a reasonable amount of 1000, as well as add some additional hidden layers. This allowed us to increase the accuracy of training and testing data sets by over 5%, proving that proper formation of hidden layers can achieve much greater accuracy.

For the third network, we decided to make sure that the amount of nodes in the last hidden layer is equal to the number of output classes of the network (26). Making sure that as the network progresses, it is able to shrink and get smaller to output. In addition, the number of epochs was decided to be increased to 6500. Such a network allowed for another increase in the accuracy on the training set. Some of the most noticeable parts being that there were no drastic outliers among the letters, many accuracies were clustered tightly together. Unfortunately, when the network was tested using the 20% training set, it has not shown any improvement despite the size of the NN, and the outliers were prevalent once again.

Section 2

Should include your selection of the best network that you found. The following should be included and clearly described in this section:

i.A description of the network, including number of layers and neurons in each layer, and the overall accuracy of the network on the testing set.

Input Layer: Attributes (16 nodes due to 16 attributes of each letter).
Hidden Layer 1: 52 nodes total
Hidden Layer 2: 26 nodes total.
Output Layer: 26 nodes total.
Number of Epochs: 2000.

Accuracy on the Testing Set: 93.31%
Accuracy on the Training Set: 90.067%

```
=== Summary ===

Correctly Classified Instances        2702               90.0667 %
Incorrectly Classified Instances       298                9.9333 %
Kappa statistic                          0.8967
Mean absolute error                      0.0084
Root mean squared error                  0.0836
Total Number of Instances             3000
```

j.The accuracy by which each letter was classified by this best network. One column for the accuracy on the training set (data set 1) and one column for the accuracy on the test set (data set 2). Use a format similar to what you used in Part 1.

Letter	Accuracy Testing Set
A	91.2%
B	85.5%
C	91.0%
D	87.5%
E	90.0%
F	87.2%
G	92.8%
H	73.0%
I	90.7%
J	92.5%
K	87.1%
L	92.9%
M	94.6%
N	86.7%
O	86.4%

P	89.2%
Q	93.0%
R	90.2%
S	89.6%
T	90.8%
U	92.3%
V	89.6%
W	96.5%
X	95.2%
Y	96.7%
Z	90.3%

k.A paragraph with a clear justification for why you selected the network as being “best”.

During the many hours of searching for the best possible network, I decided that reaching the 90% accuracy was good enough considering the hardware that was at my disposal. With the current “best” network, it took me around 1.5 hours to train the network. In addition, the small difference between the training and testing accuracies proved that the network can be viewed as a reliable network. I just wish that my computer had higher capabilities in building such networks at a quicker pace. Otherwise the network presents itself as being quite reliable and is suited for the future use.

l.A paragraph discussing the strengths and weaknesses of the entire method followed,as well as strengths and weaknesses of WEKA as a tool.
Some of the strengths that we have found is that a person with zero programming knowledge will be able to use WEKA, if they attempt to wrap their head around it. It presents the user with a variety of tools they can use to build and modify their neural network, as well as display statistical data.
Some of the disadvantages that we have found include cumbersome interface that takes a second to figure out, lack of great documentation about the tool, relatively slow performance, for section 3 of the project each neural network took 30 min -1:30 hr to train, as well as inability to export the statistical data from the neural network as excel or .csv file, which would serve nicely for the

scope of this project. I believe that overall it was an easy tool and a good project to get the hang of neural networks!