ITU Computer and Informatics Faculty
BLG 454E Learning From Data, Spring 2018
Homework #4
Due May 27, 2018 11pm

Kadir Emre Oto
(150140032)

1: Multilayer Perceptron

(a) (25 pts) Report cross entropy error value for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 50, 100, 200th iteration in the backpropagation(training) phase.

According the results which is given below, if iteration count is incresed, the accuracy increases and the cross entropy error value decreases.

Iteration Count: 1
Cross Entropy Error: 5560.259742484045
Accuracy: 0.75
———-
Iteration Count: 2
Cross Entropy Error: 2455.591097218306
Accuracy: 0.8333333333333334
———-
Iteration Count: 3
Cross Entropy Error: 1553.4939942454293
Accuracy: 0.9166666666666666
———-
Iteration Count: 4
Cross Entropy Error: 1430.0461926499293
Accuracy: 0.8333333333333334
———-
Iteration Count: 5
Cross Entropy Error: 1923.7371876472791
Accuracy: 0.75
———-
Iteration Count: 6
Cross Entropy Error: 1077.1906751938932
Accuracy: 0.8333333333333334
———-
Iteration Count: 7
Cross Entropy Error: 1059.2269895314132
Accuracy: 0.8333333333333334
———-
Iteration Count: 8
Cross Entropy Error: 968.299161454117
Accuracy: 0.9166666666666666
———-
Iteration Count: 9
Cross Entropy Error: 1005.9774139521438
Accuracy: 0.8333333333333334
———-
Iteration Count: 10
Cross Entropy Error: 775.3261478428205
Accuracy: 0.8333333333333334
———-
Iteration Count: 50
Cross Entropy Error: 437.0891897898743
Accuracy: 0.8333333333333334
———-
Iteration Count: 100
Cross Entropy Error: 280.4186414102801
Accuracy: 1.0
———-
Iteration Count: 150
Cross Entropy Error: 281.14420842642227
Accuracy: 0.9166666666666666
———-
Iteration Count: 200
Cross Entropy Error: 196.63660369996987
Accuracy: 0.9166666666666666

(b) (25pts)Use trained multilayer perceptron model and calculate the accuracy for test set(dataTest.csv) and report it.

The accuracy valus for test dataset is also given at **part a** for different iteration count. Average accuracy is higher than 90 percent. After 200 iterations, the accuracy may be 100 percent as Figure 1.

```
[Emre:Question1 KEO$ ls
dataTest.csv    dataTrain.csv   kod.py
[Emre:Question1 KEO$ time python3 kod.py
Iteration Count: 200
Cross Entropy Error: 298.5275463841384
Accuracy: 1.0
----------

real    0m34.797s
user    0m34.490s
sys     0m0.151s
```

Figure 1: Example usage of program and the result

2: Clustering

In this question, the first field of the data is discarded because the label of the field is not determined!

(a) Calculate Sum of Squared Error(SSE) value for k=1, 5, 10 and 20 in this dataset and report them. Decide which k value is the best clustering parameter for this dataset and report it.

According to the results, the SSE value decreases while k value increases.

```
[Emre:Question2 KEO$ ls
Cluster.csv     cluster.py      kod.py
[Emre:Question2 KEO$ python3 kod.py
k: 1    error: 1303746.0612357464
k: 5    error: 266995.1221811069
k: 10   error: 128358.20813854622
k: 20   error: 75131.30436781942
```

Figure 2: Example usage of program and the results

(b) (25 pts) Cluster this dataset using finded the best k value in section a. Draw the decision boundaries and report it.

I choose the k value as 20, beacuse its sum of squared error is lowest.

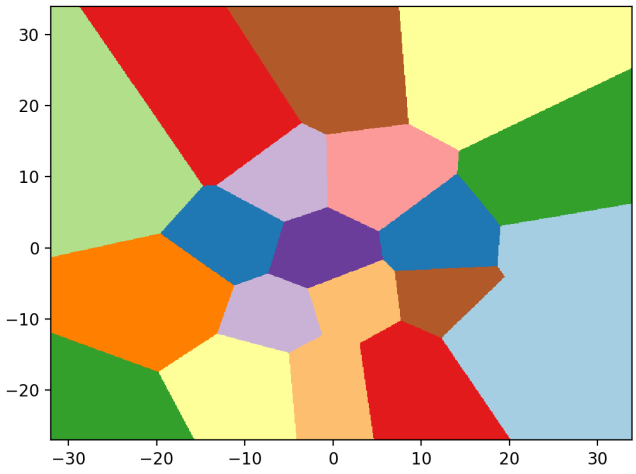

Figure 3: The decision boundaries