CS 4442 Artificial Intelligence II - Assignment 1

Ian Wood - 250690592

February 12th 2017

# Question 3:

b) Tested using the full data set, not the tiny one.

|  |  |
| --- | --- |
| k Value | Error |
| 1 | 0.0835 |
| 3 | 0.0790 |
| 5 | 0.0850 |
| 7 | 0.0885 |

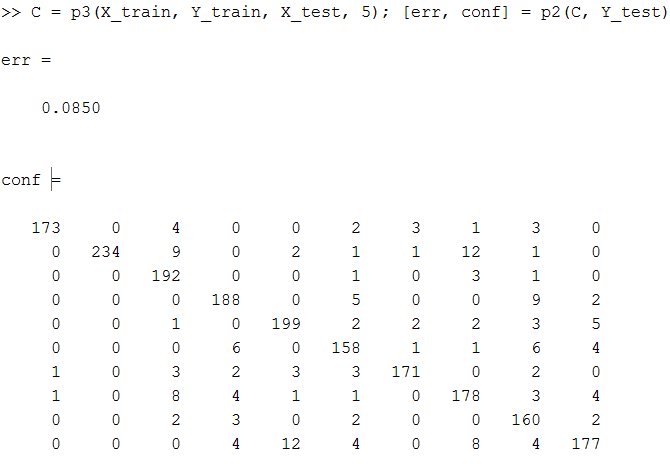
The error is consistently low, no matter the value of K. Therefore, in this test, there don’t seem to be an outstandingly large effect on the error when K is changed. This could be a consequence of the fact that error is consistently low and therefore increasing the k value can’t make a significant improvement to the accuracy of the classifier.

c) The six most common confusions are 2 classified as 8, 0 misclassified as 5, 2 misclassified as 3, 4 misclassified as 9, 8 misclassified as 3 and 8 misclassified as 2 at a rate of 12, 12, 9, 9, 8 and 8, respectively. About half of the remaining cells are 0, with the rest being below 6.

The biggest single confusions are 2 classified as 8 and 0 misclassified as 5 at 12 misclassifications each.

The biggest sum of confusions is 2 classified as 8 and 8 misclassified as 2 at a total of 20 misclassifications.

The 2 and 8 misclassifications are not symmetrical, one is 12 and the other is 8. However, it is possible that the reason that number of misclassifications are identical simply because of the distribution of data in the training set. It’s hard to say with certainty whether the difference is just noise, but there is still a reasonable difference.



# Question 5:

b)

Weights were determined using the training data and then tested on both the training and testing data.

|  |  |  |
| --- | --- | --- |
| Iterations | Training Error | Test Error |
| 100 | 0.2515 | 0.3212 |
| 1000 | 0.2078 | 0.2336 |
| 10000 | 0.1990 | 0.2725 |

From this set of data, there are two trends that can be observed. The first trend is that the number of iterations and the error are inversely proportionate (with the exception of test error for 10000 iterations). That is, more iterations will yield a lower error. The second trend is that training error is lower than test error.

The first trend can be explained be the fact that more iterations will yield more chances for a better set of weights to occur. Since the weights are randomly being generating, there is a chance that the best possible set of weights is generated. The best way to achieve that optimal set of weights is to attempt more iterations. The second trend is harder to explain. My best guess is that the method being used to determine if the weights are optimal is too primitive to give an accurate representation of the entire data set, not just the training set. However, I imagine something more complex is to the cause of this trend.

# Question 6:

b)



Weights are trained with the training data (x510train and y510train) and tested with the testing data (x510test and y510test).

With 30 iterations, a learning rate of 0.1 and initial weights of all ones, this is the error:

|  |  |
| --- | --- |
| Training Error | Test Error |
| 0.0272 | 0.0535 |

Average errors from 5b):

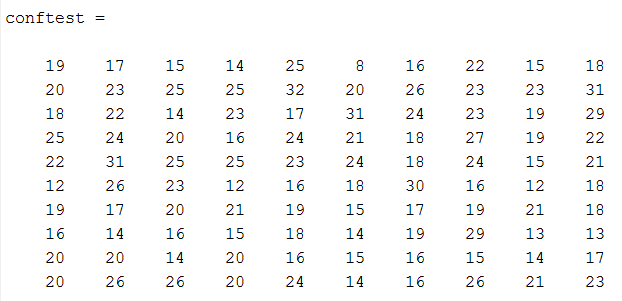
|  |  |
| --- | --- |
| Training Error | Test Error |
| 0.2194 | 0.2758 |

The error when using logistic regression is significantly lower than the previous method. This is because the weight is intelligently being tuned to the need of the data instead of just randomly guessing what might work.

# Question 8:

b)

|  |  |
| --- | --- |
| Training Error | Test Error |
| 0.0696 | 0.9020 |



# Question 9:

# Question 10: