**Assignment 3**

**6375.002**

*NetID: pem150030*

*Name: Praveen Erode Murugesan*

**Problem statement:**

To implement perceptron algorithm for email classification and K-Means algorithm to compress the image.

**Perceptron Implementation:**

Perceptron is a single unit of Neural Networks. The perceptron takes the features as input and outputs the value. Based on the output value, the email is classified either as Ham or Spam. The perceptron is trained using perceptron training rule. In the perceptron training rule the weights of the features are updated using the following formula.

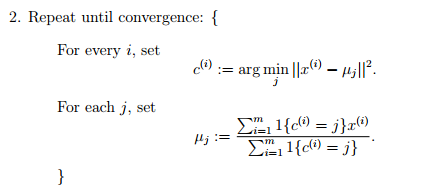


|  |  |  |  |
| --- | --- | --- | --- |
| Learning Rate | Iterations | Accuracy with stopwords | Accuracy without stopwords |
| 0.001 | 300 | 84.52 | 86.19 |
| 0.005 | 300 | 85.98 | 87.65 |
| 0.0001 | 300 | 79.49 | 80.75 |
| 0.0001 | 500 | 80.96 | 81.38 |
| 0.005 | 500 | 85.98 | 87.65 |
| 0.001 | 500 | 84.93 | 87.02 |
| 0.0035 | 500 | 85.56 | 87.44 |
| 0.1 | 500 | 92.25 | 90.79 |
| 0.5 | 500 | 94.35 | 93.30 |
| 1 | 500 | 94.14 | 93.30 |
| 0.0012 | 500 | 85.14 | 87.23 |
| 2.5 | 500 | 93.51 | 91.63 |
| 0.3333 | 500 | 93.72 | 93.93 |
| 0.2 | 250 | 93.09 | 93.93 |
| 0.00001 | 250 | 72.50 | 72.80 |
| 0.065 | 250 | 89.121 | 93.30 |
| 0.005 | 250 | 85.580 | 87.65 |
| 0.0003 | 250 | 82.00 | 83.03 |
| 0.8 | 250 | 93.51 | 95.18 |
| 0.001 | 250 | 85.56 | 84.72 |

This table explains the performance of the implemented algorithm for different values of iterations and learning rate.

**Image Compression K-Means:**

KMeans algorithm is implemented for image compression. This is done by clustering the nearer points on the image. The assignment of image points to a cluster depends on the following formula.



The implemented program is tested for different values of K. It is observed that the image’s clarity increased with increase in the value of K.

Image Compression Ratios:

|  |  |  |
| --- | --- | --- |
| **K Value** | **Compression Ratio for Penguin** | **Compression Ratio for Koala** |
| 1 | 58.823 | 58.823 |
| 2 | 9.09 | 5.988 |
| 5 | 7.69 | 4.54 |
| 10 | 6.66 | 4.672 |
| 15 | 6.944 | 4.784 |
| 25 | 6.66 | 4.926 |
| 40 | 6.667 | 4.975 |
| 50 | 6.86 | 4.926 |
| 75 | 6.451 | 4.878 |
| 100 | 5.882 | 6.802 |

Average Compression ratio for Koala: 10.53

Variance of Compression ratio for Koala: 259.55

Average Compression ratio for Penguin: 12.17

Variance of Compression ratio for Penguin: 242.88

Yes there is trade-off between image quality and compression ratio. As the compression ratio increases, the image quality decreases.

The best K value for penguin: 100

The best K value for koala: 75

WEKA’s performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Learning Rate | Momentum | Iteration | Hidden Layers | Hidden Units | Accuracy |
| 0.03 | 0.1 | 2 | 1 | 2 | 72.0074 % |
| 0.1 | 0.2 | 100 | 1 | 10 | 97.9742 % |
| 0.3 | 0.4 | 100 | 2 | 20 | 95.9484 % |
| 0.9 | 0.1 | 2 | 1 | 4 | 97.0534 % |
| 0.9 | 0.6 | 100 | 3 | 12 | 95.7643 % |

Conclusion:

The assignment helped to learn about perceptron and K-Means clustering algorithms and provided opportunities to work with WEKA.