Senthilkumar Karunakaran

Assignment 3:

Used digits 3 & 4 from training set to build the classifier and used test data to test the classifier.

1. Data set used for training by filtering for digits 3 & 4.

Total size: 60000

Number of 3s: 6131

Number of 4s: 5842

1. After PCA, original and reconstructed images. Reconstructed image doesn’t compromise the identity of the samples, they are still readable to human eye.

Original:





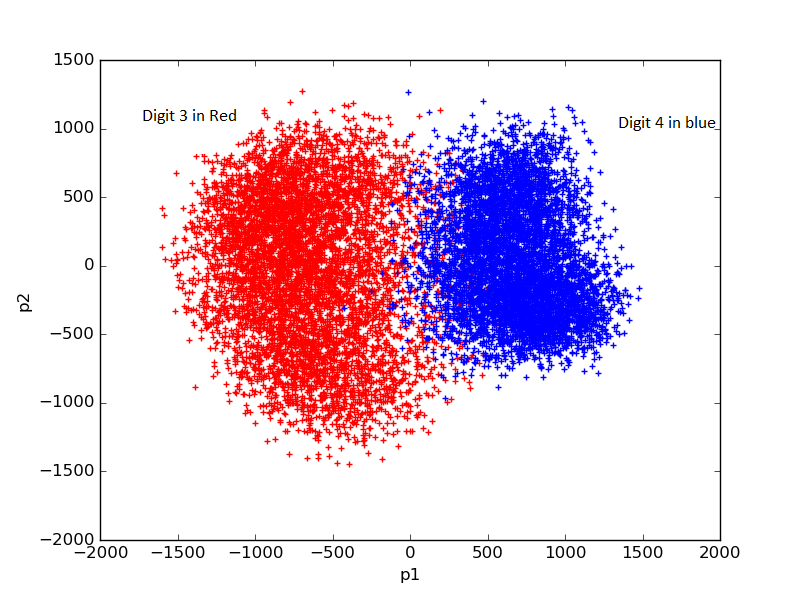
Reconstructed:





1. Scatter plot for the PCA of digits 3 & 4

Digit 3 is plotted in RED and 4 in BLUE



1. Classifier results with histogram & Bayesian

Test data has 1010 number of 3s and 982 number of 4s

|  |  |  |  |
| --- | --- | --- | --- |
| From Testing Set | Histogram ( Correct / Wrong / Indeterminate) | Bayesian ( Correct / Wrong / Indeterminate) | |
| Digit – 3 | 979 / 28 / 3 | 986 / 24 /0 |
| Digit – 4 | 972 / 8 / 2 | 975 / 7 /0 |
| Digits 3 & 4 | 1951 / 36 /5 | 1961 / 31 /0 |
| Digits except 3 & 4 | 0 / 7994 / 14 | 0/ 8008 / 0 |
| All digits in test set | 1951 / 8030 / 19  Out of 8030; 4137 are FP | 1961 / 8039 / 0  Out of 8039; 3607 are FP |

For digits 3 & 4 (first 2 rows above), both histogram & Bayesian classifier were able to classify the digits with error. Bayesian is able to classify all without any indeterminate state.

For all other digits excluding 3 & 4 (Row 4), both histogram & Bayesian classifiers failed. They classified all the inputs to either 3 or 4. Since the input is not 3 & 4, all are classified to wrong class labels.

1. Examples of correct & incorrect classifications

|  |  |  |
| --- | --- | --- |
| Histogram output | Bayesian output | Image tested |
| Incorrectly classified as 4 | Incorrectly classified as 4 |  |
| Correctly classified as 3 | Incorrectly classified as 4 |  |
| Incorrectly classified as 3 | Correctly classified as 4 |  |
| Correctly classified as 3 | Correctly classified as 3 |  |
| Correctly classified as 4 | Correctly classified as 4 |  |

1. Metrics

Ground truth: 3 is (-) , 4 is (+)

Details of the calculations are inferred from the classifiers results table under heading 4 (Classifier results with histogram & Bayesian):

Accuracy: Refer row 3 in the table

Sensitivity: Refer row 2 in the table

Specificity : Refer row 1 in the table

|  |  |  |
| --- | --- | --- |
| Metric | Histogram | Bayesian |
| Accuracy | 0.9794 | 0.9844 (1961 / 1992) |
| Sensitivity | 0.9898 | 0.9928 |
| Specificity | 0.9693 | 0.9762 |
| PPV | 0.1898 | 0.2127 |

Inference from PPV, Bayesian is able to find TP slightly better than histograms.