

6th-sem-mini-Project

Project title - Incremental Dimensionality Reduction And one class classification on NHRR data set

- NHRR data set is a dataset of handwritten digits generated by our own
- Each image is of size 50x50 pixels
- Image_taken_0 folder contains the original image taken
- Original image is resized to 50x50 pixels image which is stored in 0_50by50
- Resized image is cropped to make a smooth boundary around the digit and that is stored in folder Crop_0
- Cropped image is again upscaled to 50x50 pixels stored in Last_0 folder
- 200 Features are extracted from each image stored in file features_0.txt
- Each digit has 600 images

NHRR Dataset - On taking 200 features in an image of size 50x50 the accuracy obtained on the built in classifier and our own classifier is as follows -

Digit no	Testing data	Training data	Built in classifier accuracy (%)	Our own classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	99.69	96.67
1	600 - 1 5400 - other digits	500 - 1	97.67	90.74
2	600 - 2 5400 - other digits	500 - 2	98.36	95.94
3	600 - 3 5400 - other digits	500 - 3	98.08	96.07
4	600 - 4 5400 - other digits	500 - 4	98.99	95.48
5	600 - 5 5400 - other digits	500 - 5	98.42	83.50
6	600 - 6 5400 - other digits	500 - 6	99.47	96.21
7	600 - 7 5400 - other digits	500 - 7	99.12	95.77
8	600 - 8 5400 - other digits	500 - 8	88.78	80.96
9	600 - 9 5400 - other digits	500 - 9	90.71	74.18

NHRR Dataset - On adding more features like adding maximum continuous 0's in each row,column,diagonal accuracy obtained was - (total features - 494)

Digit no	Testing data	Training data	Built in classifier accuracy (%)	Our own classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	99.19	96.07
1	600 - 1 5400 - other digits	500 - 1	97.28	95.19
2	600 - 2 5400 - other digits	500 - 2	98.04	96.05
3	600 - 3 5400 - other digits	500 - 3	98.10	95.87
4	600 - 4 5400 - other digits	500 - 4	97.76	95.68
5	600 - 5 5400 - other digits	500 - 5	97.93	90.09
6	600 - 6 5400 - other digits	500 - 6	99.22	95.65
7	600 - 7 5400 - other digits	500 - 7	98.47	95.74
8	600 - 8 5400 - other digits	500 - 8	98.45	94.99
9	600 - 9 5400 - other digits	500 - 9	97.73	85.53

NHRR Dataset - Zone Based Feature Extraction divided 50x50 image into 25 zones of 10x10 each - (total features - 1000)

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	97.95
1	600 - 1 5400 - other digits	500 - 1	98.27
2	600 - 2 5400 - other digits	500 - 2	97.67
3	600 - 3 5400 - other digits	500 - 3	97.95
4	600 - 4 5400 - other digits	500 - 4	97.79
5	600 - 5 5400 - other digits	500 - 5	98.36
6	600 - 6 5400 - other digits	500 - 6	97.78
7	600 - 7 5400 - other digits	500 - 7	97.82
8	600 - 8 5400 - other digits	500 - 8	96.47
9	600 - 9 5400 - other digits	500 - 9	96.62

NHRR Dataset - PCA with no of Eigen vectors as 20 that implies that the dimensions were reduced from 200 to 20 -

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	99.39
1	600 - 1 5400 - other digits	500 - 1	98.42
2	600 - 2 5400 - other digits	500 - 2	98.32
3	600 - 3 5400 - other digits	500 - 3	99.16
4	600 - 4 5400 - other digits	500 - 4	99.46
5	600 - 5 5400 - other digits	500 - 5	99.16
6	600 - 6 5400 - other digits	500 - 6	99.55
7	600 - 7 5400 - other digits	500 - 7	99.56
8	600 - 8 5400 - other digits	500 - 8	90.06
9	600 - 9 5400 - other digits	500 - 9	92.97

NHRR Dataset - PCA with no of Eigen vectors as 20 that implies that the dimensions were reduced from 1000 to 20 -

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	99.75
1	600 - 1 5400 - other digits	500 - 1	99.84
2	600 - 2 5400 - other digits	500 - 2	98.55
3	600 - 3 5400 - other digits	500 - 3	98.67
4	600 - 4 5400 - other digits	500 - 4	99.27
5	600 - 5 5400 - other digits	500 - 5	98.95
6	600 - 6 5400 - other digits	500 - 6	99.46
7	600 - 7 5400 - other digits	500 - 7	99.56
8	600 - 8 5400 - other digits	500 - 8	98.38
9	600 - 9 5400 - other digits	500 - 9	94.96

NHRR Dataset - PCA with no of Eigen vectors as 20 that implies that the dimensions were reduced from 494 to 20 -

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	600 - 0 5400 - other digits	500 - 0	99.52
1	600 - 1 5400 - other digits	500 - 1	98.38
2	600 - 2 5400 - other digits	500 - 2	98.76
3	600 - 3 5400 - other digits	500 - 3	99.12
4	600 - 4 5400 - other digits	500 - 4	98.69
5	600 - 5 5400 - other digits	500 - 5	99.15
6	600 - 6 5400 - other digits	500 - 6	99.55
7	600 - 7 5400 - other digits	500 - 7	99.43
8	600 - 8 5400 - other digits	500 - 8	99.13
9	600 - 9 5400 - other digits	500 - 9	98.05

MNIST Dataset - PCA with no of Eigen vectors as 20 that implies that the dimensions were reduced from 112 to 20 -

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	6000 - 0 54000 - other digits	5000 - 0	95.78
1	6000 - 1 54000 - other digits	5000 - 1	95.78
2	6000 - 2 54000 - other digits	5000 - 2	93.92
3	6000 - 3 54000 - other digits	5000 - 3	94.35
4	6000 - 4 54000 - other digits	5000 - 4	93.98
5	6000 - 5 54000 - other digits	5000 - 5	94.37
6	6000 - 6 54000 - other digits	5000 - 6	94.74
7	6000 - 7 54000 - other digits	5000 - 7	94.64
8	6000 - 8 54000 - other digits	5000 - 8	94.62
9	6000 - 9 54000 - other digits	5000 - 9	95.02

MNIST Dataset - PCA with no of Eigen vectors as 20 that implies that the dimensions were reduced from 274 to 20 -

Digit no	Testing data	Training data	Built in classifier accuracy (%)
0	6000 - 0 54000 - other digits	5000 - 0	95.78
1	6000 - 1 54000 - other digits	5000 - 1	95.74
2	6000 - 2 54000 - other digits	5000 - 2	93.83
3	6000 - 3 54000 - other digits	5000 - 3	94.24
4	6000 - 4 54000 - other digits	5000 - 4	94.01
5	6000 - 5 54000 - other digits	5000 - 5	94.39
6	6000 - 6 54000 - other digits	5000 - 6	95.07
7	6000 - 7 54000 - other digits	5000 - 7	94.59
8	6000 - 8 54000 - other digits	5000 - 8	94.57
9	6000 - 9 54000 - other digits	5000 - 9	95.05