**SEP 786: AI AND ML FUNDAMENTALS COURSE PROJECT**

**RESULT REPORT**

**PCA (Principal Component Analysis):**

**MSE (Mean Square Error) TABLE**

|  |  |
| --- | --- |
| **No. Of Components Reduced** | **Error Magnitude** |
| 0 | 0.0 |
| 1 | 0.0 |
| 2 | 0.0 |
| 3 | 0.0 |
| 4 | 0.001 |
| 5 | 0.002 |
| 6 | 0.004 |
| 7 | 0.006 |
| 8 | 0.008 |
| 9 | 0.014 |
| 10 | 0.021 |
| 11 | 0.03 |
| 12 | 0.044 |
| 13 | 0.072 |
| 14 | 0.116 |
| 15 | 0.175 |
| 16 | 0.279 |
| 17 | 0.388 |
| 18 | 0.642 |

INFERENCE:

1. Clearly visible as we remove the components or we reduce the dimensions, the error magnitude increases.
2. The error magnitude remains 0(approximately) till removal of 4 components. After that it increases by a factor of 1.5 (approximately) and reaches to the peak of 0.642 as we remove 18 components.

**CONFUSION MATRICES FOR LDA (Linear Discriminant Analysis)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dimensions reduced** | **Confusion Matrices** | | | | | | | |
| 0 |  | 0 | 1 |  | 10 |  | 0 | 1 |
| 0 | 113 | 18 | 0 | 92 | 39 |
| 1 | 65 | 92 | 1 | 56 | 101 |
|  | | | |  | | | |
| 1 |  | 0 | 1 | 11 |  | 0 | 1 |
| 0 | 114 | 17 | 0 | 81 | 50 |
| 1 | 66 | 91 | 1 | 60 | 97 |
|  | | | |  | | | |
| 2 |  | 0 | 1 | 12 |  | 0 | 1 |
| 0 | 107 | 24 | 0 | 72 | 59 |
| 1 | 61 | 96 | 1 | 58 | 99 |
|  | | | |  | | | |
| 3 |  | 0 | 1 | 13 |  | 0 | 1 |
| 0 | 110 | 21 | 0 | 70 | 61 |
| 1 | 61 | 96 | 1 | 56 | 101 |
|  | | | |  | | | |
| 4 |  | 0 | 1 | 14 |  | 0 | 1 |
| 0 | 100 | 31 | 0 | 68 | 63 |
| 1 | 64 | 93 | 1 | 59 | 98 |
|  | | | |  | | | |
| 5 |  | 0 | 1 | 15 |  | 0 | 1 |
| 0 | 97 | 34 | 0 | 67 | 64 |
| 1 | 63 | 94 | 1 | 58 | 99 |
|  | | | |  | | | |
| 6 |  | 0 | 1 | 16 |  | 0 | 1 |
| 0 | 96 | 35 | 0 | 68 | 63 |
| 1 | 63 | 94 | 1 | 58 | 99 |
|  | | | |  | | | |
| 7 |  | 0 | 1 | 17 |  | 0 | 1 |
| 0 | 97 | 34 | 0 | 68 | 63 |
| 1 | 64 | 93 | 1 | 58 | 99 |
|  | | | |  | | | |
| 8 |  | 0 | 1 | 18 |  | 0 | 1 |
| 0 | 96 | 35 | 0 | 66 | 65 |
| 1 | 63 | 94 | 1 | 62 | 95 |
|  | | | |  | | | |
| 9 |  | 0 | 1 |
| 0 | 97 | 34 |
| 1 | 64 | 93 |
|  | | | |

**CONFUSION MATRICES FOR RANDOM FOREST CLASSIFIER**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dimensions reduced** | **Confusion Matrices** | | | | | | | |
| 0 |  | 0 | 1 |  | 10 |  | 0 | 1 |
| 0 | 102 | 29 | 0 | 92 | 39 |
| 1 | 38 | 119 | 1 | 57 | 100 |
|  | | | |  | | | |
| 1 |  | 0 | 1 | 11 |  | 0 | 1 |
| 0 | 102 | 29 | 0 | 93 | 38 |
| 1 | 45 | 112 | 1 | 53 | 102 |
|  | | | |  | | | |
| 2 |  | 0 | 1 | 12 |  | 0 | 1 |
| 0 | 99 | 32 | 0 | 85 | 46 |
| 1 | 41 | 116 | 1 | 55 | 102 |
|  | | | |  | | | |
| 3 |  | 0 | 1 | 13 |  | 0 | 1 |
| 0 | 103 | 28 | 0 | 88 | 43 |
| 1 | 48 | 109 | 1 | 55 | 102 |
|  | | | |  | | | |
| 4 |  | 0 | 1 | 14 |  | 0 | 1 |
| 0 | 99 | 32 | 0 | 83 | 48 |
| 1 | 50 | 107 | 1 | 64 | 93 |
|  | | | |  | | | |
| 5 |  | 0 | 1 | 15 |  | 0 | 1 |
| 0 | 96 | 35 | 0 | 82 | 49 |
| 1 | 50 | 107 | 1 | 58 | 99 |
|  | | | |  | | | |
| 6 |  | 0 | 1 | 16 |  | 0 | 1 |
| 0 | 95 | 36 | 0 | 81 | 50 |
| 1 | 54 | 103 | 1 | 63 | 94 |
|  | | | |  | | | |
| 7 |  | 0 | 1 | 17 |  | 0 | 1 |
| 0 | 93 | 38 | 0 | 79 | 52 |
| 1 | 57 | 100 | 1 | 61 | 96 |
|  | | | |  | | | |
| 8 |  | 0 | 1 | 18 |  | 0 | 1 |
| 0 | 95 | 36 | 0 | 71 | 60 |
| 1 | 59 | 98 | 1 | 73 | 84 |
|  | | | |  | | | |
| 9 |  | 0 | 1 |
| 0 | 97 | 34 |
| 1 | 58 | 99 |
|  | | | |

**ACCURACY:LINEAR DISCRIMANT ANALYSIS vs RANDOM FOREST CLASSIFIER**

|  |  |  |
| --- | --- | --- |
| **No. Of Components Reduced** | **Linear Discrimant analysis** | **Random Forest Classifier** |
| 0 | 0.711806 | 0.767361 |
| 1 | 0.711806 | 0.743056 |
| 2 | 0.704861 | 0.746528 |
| 3 | 0.715278 | 0.736111 |
| 4 | 0.670139 | 0.715278 |
| 5 | 0.663194 | 0.704861 |
| 6 | 0.659722 | 0.687500 |
| 7 | 0.659722 | 0.670139 |
| 8 | 0.659722 | 0.670139 |
| 9 | 0.659722 | 0.680556 |
| 10 | 0.670139 | 0.666667 |
| 11 | 0.618056 | 0.677083 |
| 12 | 0.593750 | 0.656250 |
| 13 | 0.593750 | 0.659722 |
| 14 | 0.576389 | 0.611111 |
| 15 | 0.576389 | 0.628472 |
| 16 | 0.579861 | 0.607639 |
| 17 | 0.579861 | 0.607639 |
| 18 | 0.559028 | 0.538194 |

**\*100 for percentage**

**BACKWARD SEARCH (FEATURE SELECTION)**

**CONFUSION MATRICES FOR LDA (Linear Discriminant Analysis)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dimensions reduced** | **Confusion Matrices** | | | | | | | |
| 0 |  | 0 | 1 |  | 10 |  | 0 | 1 |
| 0 | 31 | 92 | 0 | 19 | 93 |
| 1 | 108 | 57 | 1 | 120 | 56 |
|  | | | |  | | | |
| 1 |  | 0 | 1 | 11 |  | 0 | 1 |
| 0 | 30 | 97 | 0 | 21 | 94 |
| 1 | 109 | 52 | 1 | 118 | 55 |
|  | | | |  | | | |
| 2 |  | 0 | 1 | 12 |  | 0 | 1 |
| 0 | 30 | 98 | 0 | 21 | 92 |
| 1 | 109 | 51 | 1 | 118 | 57 |
|  | | | |  | | | |
| 3 |  | 0 | 1 | 13 |  | 0 | 1 |
| 0 | 23 | 91 | 0 | 18 | 90 |
| 1 | 116 | 58 | 1 | 121 | 59 |
|  | | | |  | | | |
| 4 |  | 0 | 1 | 14 |  | 0 | 1 |
| 0 | 24 | 95 | 0 | 20 | 89 |
| 1 | 115 | 54 | 1 | 119 | 60 |
|  | | | |  | | | |
| 5 |  | 0 | 1 | 15 |  | 0 | 1 |
| 0 | 23 | 95 | 0 | 19 | 90 |
| 1 | 116 | 54 | 1 | 120 | 59 |
|  | | | |  | | | |
| 6 |  | 0 | 1 | 16 |  | 0 | 1 |
| 0 | 23 | 95 | 0 | 25 | 89 |
| 1 | 116 | 54 | 1 | 114 | 60 |
|  | | | |  | | | |
| 7 |  | 0 | 1 | 17 |  | 0 | 1 |
| 0 | 22 | 95 | 0 | 23 | 83 |
| 1 | 117 | 54 | 1 | 116 | 66 |
|  | | | |  | | | |
| 8 |  | 0 | 1 | 18 |  | 0 | 1 |
| 0 | 21 | 94 | 0 | 62 | 92 |
| 1 | 118 | 55 | 1 | 77 | 57 |
|  | | | |  | | | |
| 9 |  | 0 | 1 |
| 0 | 19 | 94 |
| 1 | 120 | 55 |
|  | | | |

**CONFUSION MATRICES FOR RANDOM FOREST CLASSIFIER**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dimensions reduced** | **Confusion Matrices** | | | | | | | |
| 0 |  | 0 | 1 |  | 10 |  | 0 | 1 |
| 0 | 50 | 99 | 0 | 57 | 108 |
| 1 | 89 | 50 | 1 | 82 | 41 |
|  | | | |  | | | |
| 1 |  | 0 | 1 | 11 |  | 0 | 1 |
| 0 | 57 | 101 | 0 | 53 | 108 |
| 1 | 82 | 48 | 1 | 86 | 41 |
|  | | | |  | | | |
| 2 |  | 0 | 1 | 12 |  | 0 | 1 |
| 0 | 57 | 103 | 0 | 56 | 105 |
| 1 | 82 | 46 | 1 | 83 | 44 |
|  | | | |  | | | |
| 3 |  | 0 | 1 | 13 |  | 0 | 1 |
| 0 | 51 | 111 | 0 | 56 | 106 |
| 1 | 88 | 38 | 1 | 83 | 43 |
|  | | | |  | | | |
| 4 |  | 0 | 1 | 14 |  | 0 | 1 |
| 0 | 53 | 104 | 0 | 57 | 104 |
| 1 | 86 | 45 | 1 | 82 | 45 |
|  | | | |  | | | |
| 5 |  | 0 | 1 | 15 |  | 0 | 1 |
| 0 | 51 | 108 | 0 | 55 | 108 |
| 1 | 88 | 41 | 1 | 84 | 41 |
|  | | | |  | | | |
| 6 |  | 0 | 1 | 16 |  | 0 | 1 |
| 0 | 51 | 109 | 0 | 56 | 97 |
| 1 | 88 | 40 | 1 | 83 | 52 |
|  | | | |  | | | |
| 7 |  | 0 | 1 | 17 |  | 0 | 1 |
| 0 | 52 | 113 | 0 | 47 | 97 |
| 1 | 87 | 36 | 1 | 92 | 52 |
|  | | | |  | | | |
| 8 |  | 0 | 1 | 18 |  | 0 | 1 |
| 0 | 52 | 107 | 0 | 67 | 95 |
| 1 | 87 | 42 | 1 | 72 | 54 |
|  | | | |  | | | |
| 9 |  | 0 | 1 |
| 0 | 53 | 108 |
| 1 | 86 | 41 |
|  | | | |

**ACCURACY:LINEAR DISCRIMANT ANALYSIS vs RANDOM FOREST CLASSIFIER**

|  |  |  |
| --- | --- | --- |
| **No. of Components Reduced** | **Linear Discriminant Analysis** | **Random Forest Classifier** |
| 0 | 69.44444444444444 | 65.27777777777779 |
| 1 | 71.52777777777779 | 68.05555555555556 |
| 2 | 71.875 | 68.05555555555556 |
| 3 | 71.875 | 69.09722222222221 |
| 4 | 72.91666666666666 | 70.13888888888889 |
| 5 | 73.26388888888889 | 72.56944444444444 |
| 6 | 73.26388888888889 | 70.48611111111111 |
| 7 | 73.61111111111111 | 70.48611111111111 |
| 8 | 73.61111111111111 | 68.75 |
| 9 | 74.30555555555556 | 69.44444444444444 |
| 10 | 73.95833333333334 | 69.09722222222221 |
| 11 | 73.61111111111111 | 69.44444444444444 |
| 12 | 72.91666666666666 | 67.70833333333334 |
| 13 | 73.26388888888889 | 68.75 |
| 14 | 72.22222222222221 | 65.27777777777779 |
| 15 | 72.91666666666666 | 64.58333333333334 |
| 16 | 70.48611111111111 | 62.15277777777778 |
| 17 | 69.09722222222221 | 64.23611111111111 |
| 18 | 58.68055555555556 | 58.68055555555556 |

**All are in percentage**

**Project Members**

|  |  |  |
| --- | --- | --- |
| Akashdeep Singh | singa190@mcmaster.ca | 400353314 |
| Ankur Bassi | bassia8@mcmaster.ca | 400347849 |
| Sulaiman Ghouse Mohiuddin | mohius3@mcmaster.ca | 400418055 |
| Vikram Chawla | chawlv3@mcmaster.ca | 400354764 |