

Face Recognition using K Nearest Neighbours and PCA

Ubair Nisar

15 April 2023

1 Introduction :

In this report i am presenting the result of face images classification using knn algorithm using my own implementation in python , the code for the experiment is provided along in jupyter notebook. The averages are based on different k values range from 1 - 10

2 Results :

Following Results are achieved after experimenting with dataset (The averages are based on different k values range from 1 - 10)

2.1 Experimenting with distances

Using different distance measures for KNN classifier

2.1.1 Whole Data and Euclidean distance

Table 1: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Average Standard Deviation
1	0.9644444444444444	0.01817269692644013
2	0.9622222222222222	0.016348821077291305
3	0.9583333333333334	0.01943650631615099
4	0.9411111111111112	0.0275658582890117
5	0.9472222222222223	0.02749859704614349

2.1.2 Whole Data and Cosine distance

Table 2: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.07166666666666668	0.012472191289246475
2	0.04555555555555555	0.008642416214502248
3	0.04611111111111111	0.006136311676215145
4	0.02666666666666667	0.00666666666666667
5	0.05611111111111111	0.009362388636862621

2.1.3 Whole Data and Mahalanobis distance

Table 3: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.7766666666666667	0.08576453553512405
2	0.8066666666666666	0.08141116494546542
3	0.8083333333333332	0.07767453465154028
4	0.7966666666666666	0.09823441352194251
5	0.7694444444444444	0.08480187512485414

2.2 Experimenting with number of classes

2.2.1 3 classes and euclidean distance

Table 4: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Average Standard Deviation
1	0.7766666666666667	0.08576453553512405
2	0.8066666666666666	0.08141116494546542
3	0.8083333333333332	0.07767453465154028
4	0.7966666666666666	0.09823441352194251
5	0.7694444444444444	0.08480187512485414

2.2.2 5 classes and euclidean distance

Table 5: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Average Standard Deviation
1	0.952222222222223	0.029355210696939773
2	0.98	0.015634719199411447
3	0.9766666666666666	0.019436506316151018
4	0.9599999999999999	0.025385910352879702
5	0.962222222222222	0.030469576001782413

2.3 Experimenting with less training examples

2.3.1 100 training examples and 70 test examples

Table 6: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Average Standard Deviation
1	0.9182539682539682	0.03486142165891439
2	0.9001587301587303	0.038216737464411905
3	0.9203174603174603	0.033885078407386464
4	0.9026984126984127	0.03310427724756497
5	0.9266666666666666	0.02859522818286177

2.4 Dimensional reduction of features using PCA

The key impact seen in this experiment is the reduction of computational time because of reduce in number of features of the data set. The use of cosine similarity for the distance measure is not giving any significance results even after applying PCA

2.4.1 n-components = 20 / distance : euclidean

Table 7: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.9327777777777779	0.03101174608211745
2	0.9161111111111111	0.03619221661089406
3	0.9266666666666666	0.04672615256292061
4	0.9088888888888889	0.032979604621457395
5	0.9477777777777777	0.017014880923459486

2.4.2 n-components = 20 / distance : mahalanobis

Table 8: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.9661111111111111	0.005152010275275396
2	0.9700000000000001	0.009428090415820642
3	0.9716666666666668	0.012692955176439858
4	0.9611111111111111	0.012422599874998842
5	0.9611111111111111	0.016461702655587832

2.4.3 n-components = 20 / distance : cosine

Table 9: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.02111111111111112	0.003928371006591932
2	0.01666666666666667	0.004082482904638631
3	0.02277777777777775	0.0047790695928014585
4	0.02055555555555556	0.006849348892187752
5	0.00722222222222221	0.005826715823167509

2.4.4 n-components = 50 / distance : mahalanobis

Table 10: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.02111111111111112	0.003928371006591932
2	0.01666666666666667	0.004082482904638631
3	0.02277777777777775	0.0047790695928014585
4	0.02055555555555556	0.006849348892187752
5	0.00722222222222221	0.005826715823167509

2.5 what are optimal number of components for the given data set?

From the experiment i have found out that the best is ranging from 100 to 150 n-components. Here are the results with using 120 n-components

Table 11: Average Accuracy and Standard Deviation for Cross-validation Splits

Split	Average Accuracy	Standard Deviation
1	0.02111111111111112	0.003928371006591932
2	0.01666666666666667	0.004082482904638631
3	0.022777777777777775	0.0047790695928014585
4	0.020555555555555556	0.006849348892187752
5	0.007222222222222221	0.005826715823167509

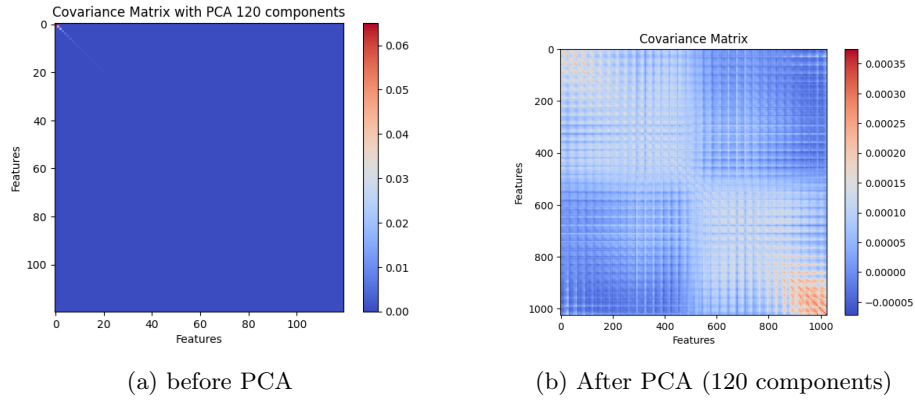


Figure 1: Comparison of Covariance Matrix before and after PCA