Minimum Euclidian Distance (Pink) – This approach works out the way we expect but is not a great classifier since we can see a fair number of points that would be misclassified.

Generalized Euclidian Distance (Black) – This approach works out slightly better than the minimum Euclidian distance, with less points being misclassified.

Maximum Posteriori Probability (Teal) – Looks the same as the generalized Euclidian Distance since both Classes have equal number of points and therefore equal probability given the priors.

Minimum Euclidian Distance (Pink) - Divides things into three as we would expect with straight lines roughly splitting up the three classes but with a fair bit of visible error in the classifications.

Generalized Euclidian Distance (Black) – Using the contours we generated we can see that the line passes through the point of intersection of the contours as we would expect. Because the standard deviation of Class D is smaller we see a parabolic shape around its mean. While Class C and E, each having a higher standard deviation in one direction, form a cross.

Maximum Posteriori Probability (Teal) – Follows the Generalized Euclidian Distance very closely however it has a more parabolic shape when it comes to the boundary between Class C and Class E. The border between class C and Class E tends to favour Class E more than Class C since Class E has 50 more points. Class D has the same parabolic shape as before however the boundary is further out as we would expect given that Class D has 200 points while the other only have 100 and 150.

Nearest Neighbour (Black) - The classification system seems to work well and form a good boundary however there are large pockets where outliers form collections of incorrectly classified points.

K Nearest Neighbours (Teal) – The classification boundary is much straighter and seems to be much better at removing outliers and the errors caused by them. The improvement is Hard to see since a nearest neighbour method of classification does not misclassify any of the training points, however it provides better classification boundaries.

Nearest Neighbour (Black) – Behaviour is very close to that in the 2 class example however there seem to be far more pockets of classifications due to higher standard deviations in the data.

K Nearest Neighbours (Teal) – This helps to smooth out the Nearest Neighbour approach the same way it did in the previous 2 class case.

One thing that is fairly apparent is the misclassification between C and D is very unlikely in any of our methods. This is explained by the large distance between the means and the small standard deviation for these two classes. The largest errors seem to be between Class D and Class E, which is accounted for by the small distance between means and large standard deviations. The weakness of the NN method is apparent in these matrices since the diagonal elements are all the smallest using the NN method, this is explained by the fact that using the nearest neighbour will not preference a specific method but creates a convoluted classification boundary.

|  |  |  |
| --- | --- | --- |
| Error Analysis | | |
| Method | Experimental Error Rate | |
| Case 1 | Case 2 |
| MED | 0.0925 | 0.2111 |
| GED | 0.075 | 0.2422 |
| MAP | 0.075 | 0.2200 |
| NN | 0.115 | 0.3044 |
| kNN | 0.0725 | 0.2378 |

Based on the experimental error we measured we can see that in case 1 the kNN method of classification provided the least error. The NN classifier provided the highest error rate.

In case 2 MED provided the lowest Error Rate while the NN provided the highest Error Rate.