

Task 3 – Bayes classification with Gaussian distributions

The Algorithm

The algorithm starts by going through a for loop going through all 26 classes to calculate the mean, covariance, inverse covariance and log determinant.

It then goes through a nested for loop, one loop going through all the rows in the test data and one loop going through all 26 classes. Here in the nested for loop, for each row, the algorithm calculates the probability for each class and then finds the index of maximum probability which is the predicted class for the respected row.

Standard Output

For my_gaussian_classify.m, standard output is as follows :

Time taken	157.237667
N	7800
Nerrs	1447
acc	0.8145

Improvements

To improve my_gaussian_classify.m, I tried to implement PCAs into my_improved_gaussian_classify.m . This meant the dimensions of the matrix were lower. To do this we first minus the mean of the columns from the training data. We then find the covariance of the matrix and with this we calculate the eigenvectors and eigenvalues. We then sort the eigenvalues in descending order and take k eigenvectors to get a $d \times k$ matrix. To project the training and test data onto the new subspace. We matrix multiply the training data and test data by the $d \times k$ matrix so we get $m \times k$ and $n \times k$ matrix respectively. After this we can this run my_gaussian_classify.m.

Accuracy can also be determined by the k eigenvectors you take.

