Task 2

s1703913

April 2019

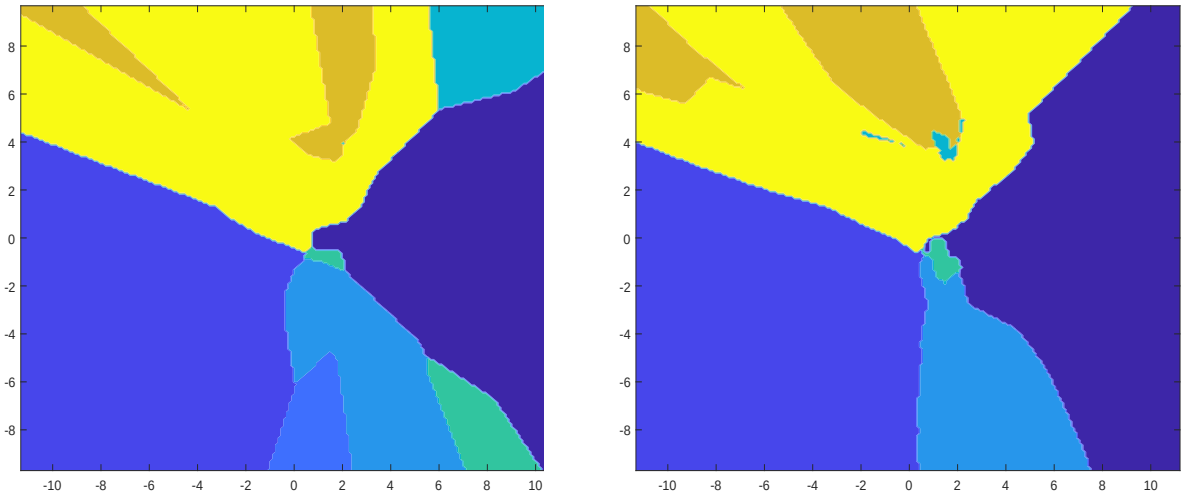
1 Task 2.1

Table 1: Analysis of k-NN classification.

k	N	Number of errors	Accuracy	Time taken
1	3994	145	96.37	6.46 secs
3	3994	141	96.47	6.47 secs
5	3994	145	96.37	6.43 secs
10	3994	147	96.63	6.44 secs
20	3994	166	96.58	6.43 secs

2 Task 2.2

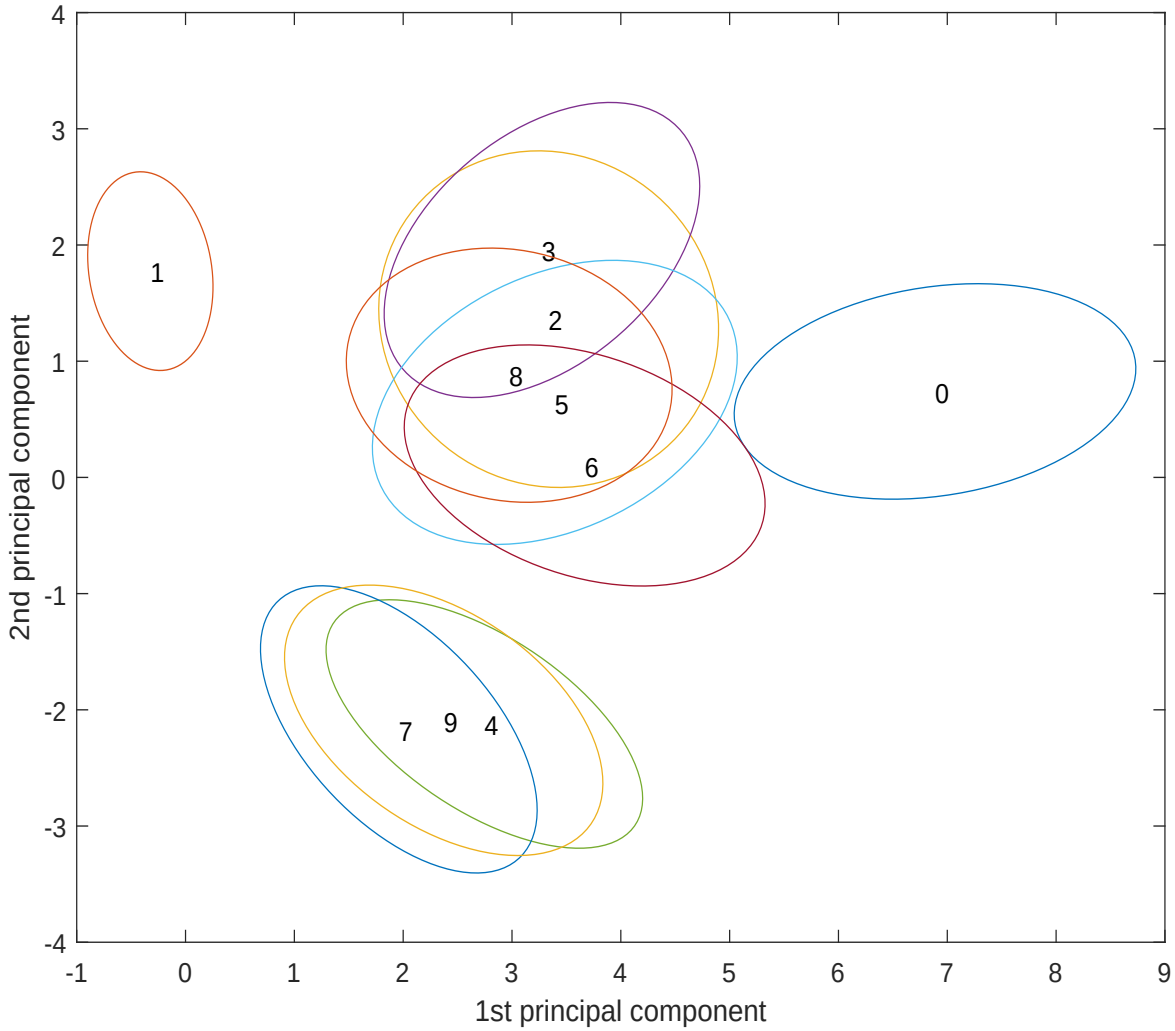
Figure 1: Cross section of decision regions of k-NN with a 2D-PCA plane, where the position of the plane is specified by a point vector. k is the number of nearest neighbours in k-NN.



k=1,3

3 Task 2.3

Figure 2: A contour of Gaussian distribution for each class k = 1,...,10.



4 Task 2.4

Table 2: The correlation on 2D-PCA for each class and for all of the classes (i.e. whole data).

Class	Correlation of data points
1	-0.2114
2	0.1562
3	0.0593
4	-0.4329
5	0.5979
6	-0.3190
7	0.3186
8	0.5613
9	0.1106
10	0.4654
All data	0.0000

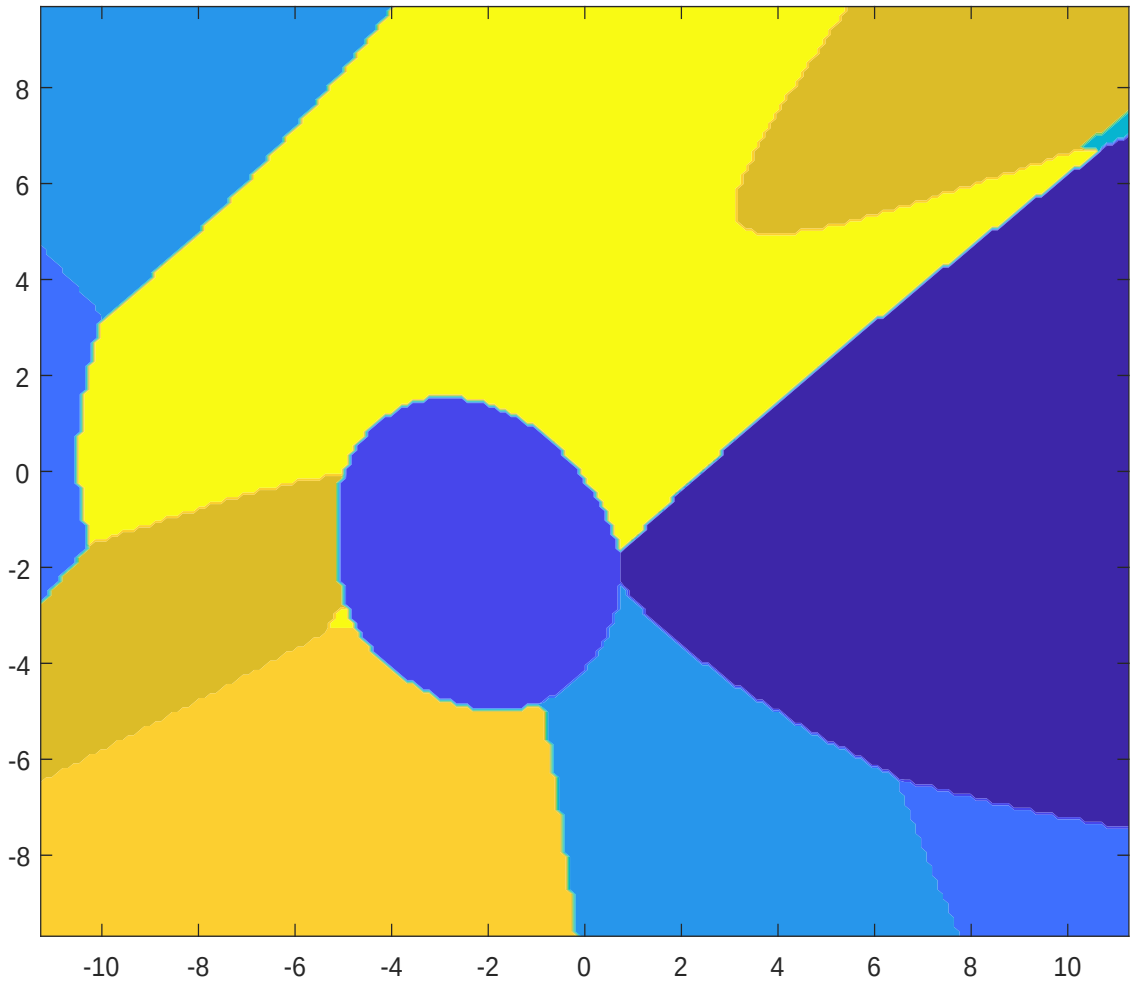
5 Task 2.5

Table 3: Analysis of Gaussian classification.

N	Number of errors	Accuracy	Time taken
3994	201	94.97	5.25 secs

6 Task 2.6

Figure 3: Cross section of decision regions of the Gaussian classifiers with a 2D-PCA plane.



7 Task 2.7

Table 4: Accuracy for each percentage of training data respectively used for Gaussian classification. A generally decreasing trend can be observed.

Percentage of training data	Accuracy
90%	94.89
80%	94.92
70%	94.94
60%	94.82
50%	94.87
40%	94.79
30%	94.64

8 Task 2.8

Table 5: Analysis of classification carried out with k-means clustering applied to each class to obtain L Gaussian classifiers per class.

L	N	Number of errors	Accuracy	Time taken
2	3994	178	95.54	17.03 secs
5	3994	140	96.49	57.82 secs
10	3994	101	97.47	116.83 secs