

Intelligent Systems Lab

Lab No- 3

Name- Adhyyan Tripathi

Roll no -8

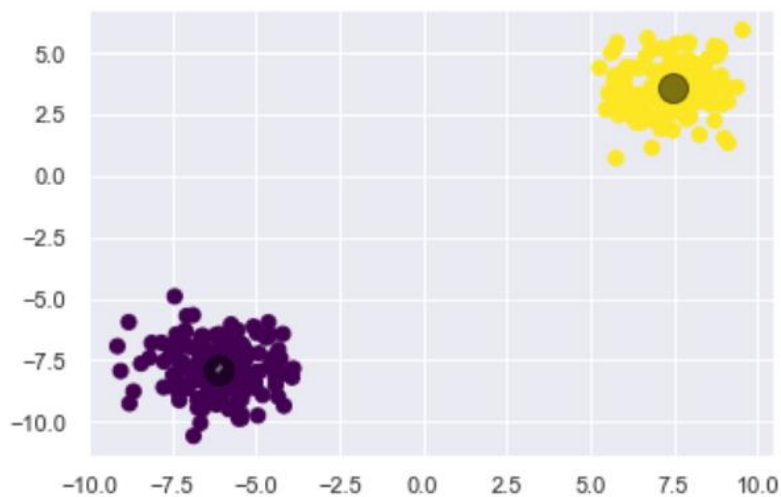
Sec C

Reg no – 201700403

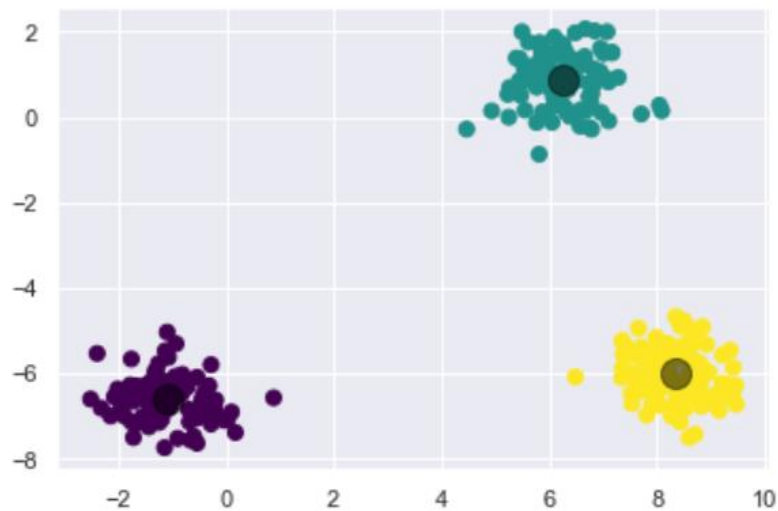
Q1) Data Visualization.

Ans- Data visualization of randomly generated dataset using makeblobs.

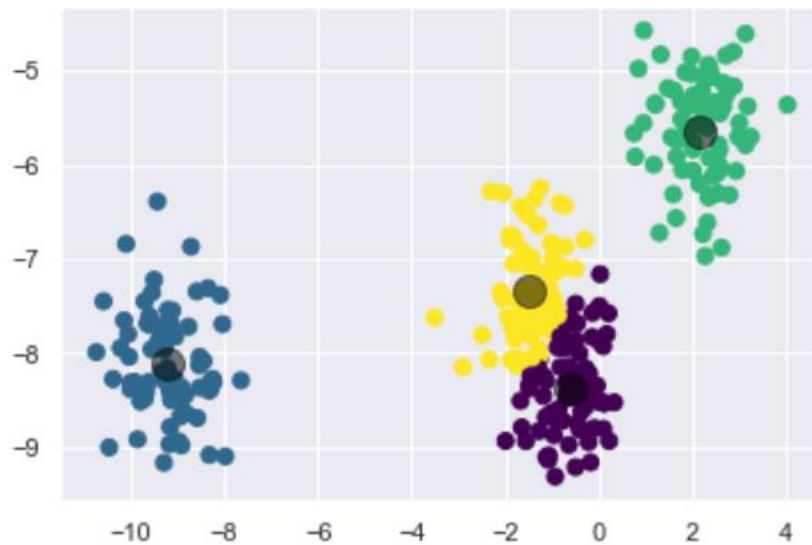
2 centers-



3 centers-



4 centers-

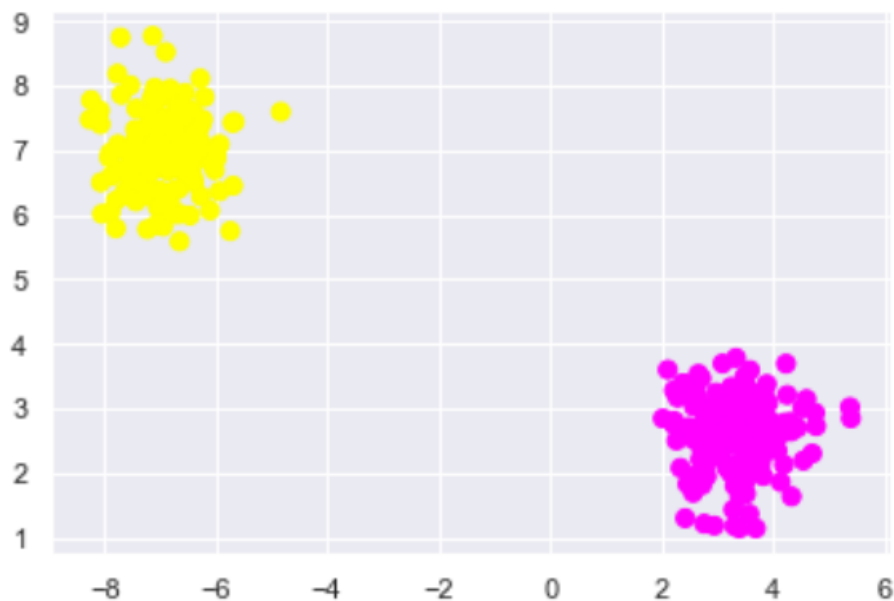


Q2) Binary classification.

Ans- For binary classification I'm generating random data with 2 clusters.

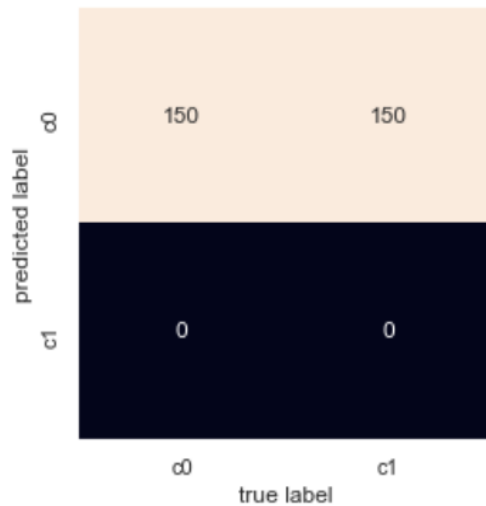
The hyperparameters that I'll be changing in this section are- `n_init`, `max_iter` and `n_clusters`.

The data used once for a configuration will remain same for all the other configurations.



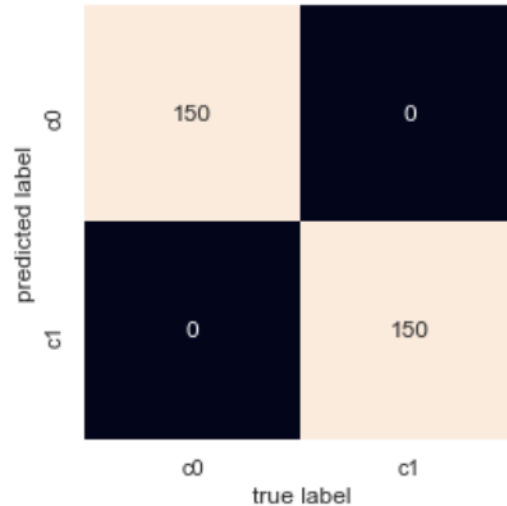
Changing the n_clusters hyperparameter-

For n_clusters=1:



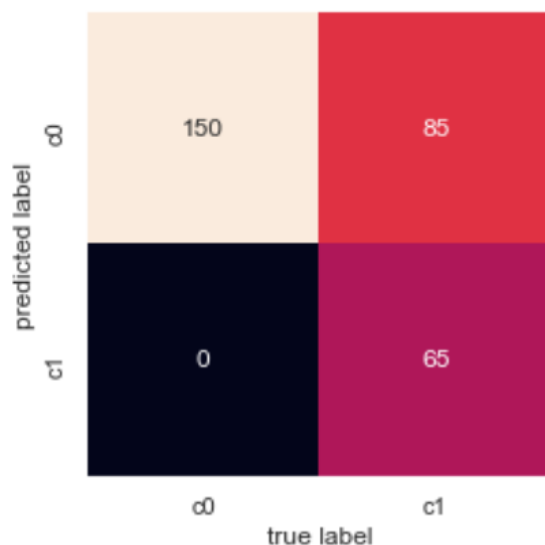
Accuracy = 0.5

For n_clusters=2:



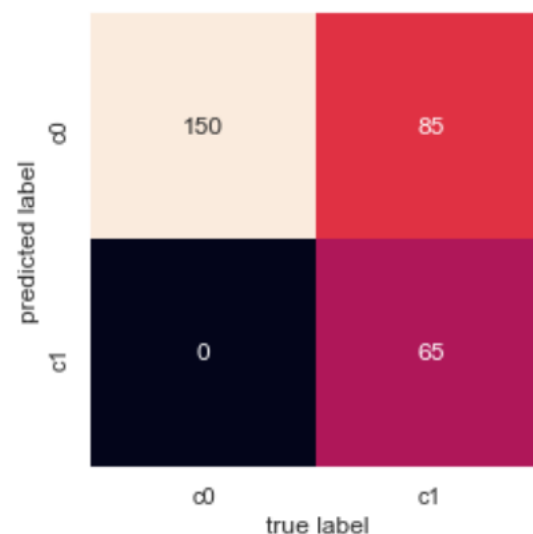
Accuracy = 1

For n_clusters=5:



Accuracy= 0.7166666666666667

For n_clusters=10:



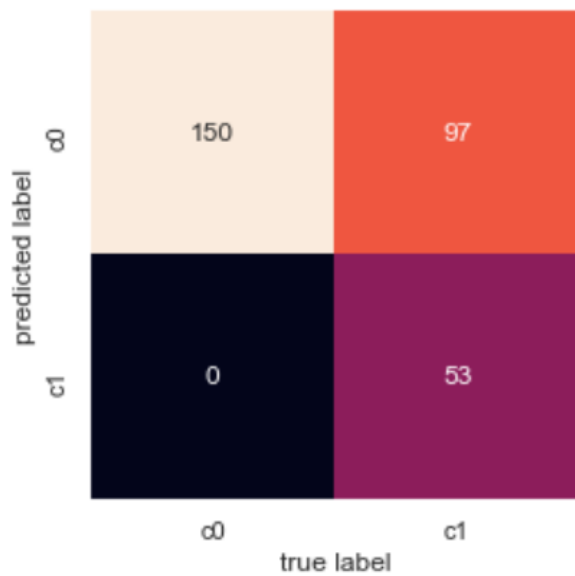
Accuracy= 0.6766666666666666

The accuracy increases if the n_clusters value is less than the number of clusters in the dataset. When the value of n_clusters is equal to the number of clusters the accuracy is the maximum. If the n_clusters value is increased further, the accuracy starts to fall.

Changing the max_iter and n_init hyperparameter-

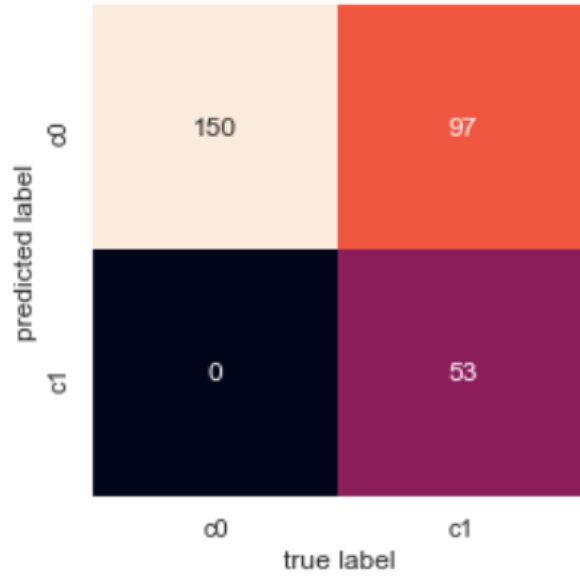
For these hyperparameters if we keep n_clusters= 2 then, the difference is very minimal, so to make these differences clear I'll be using n_cluster value as 10.

For n_init=10 and max_iter=100:



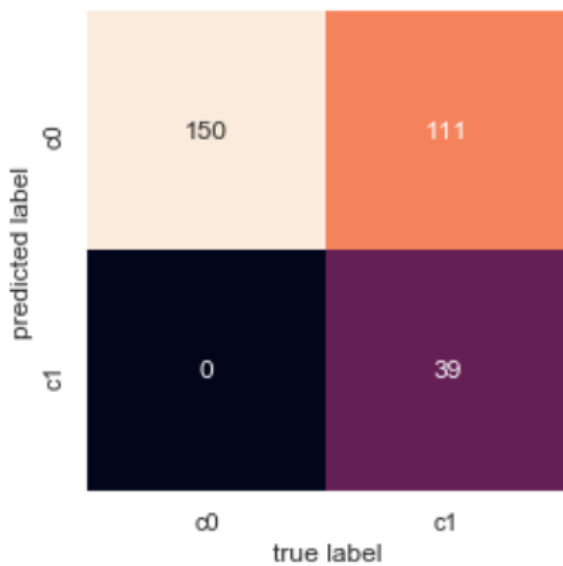
Accuracy= 0.6766666666666666

For n_init=10 and max_iter=100:



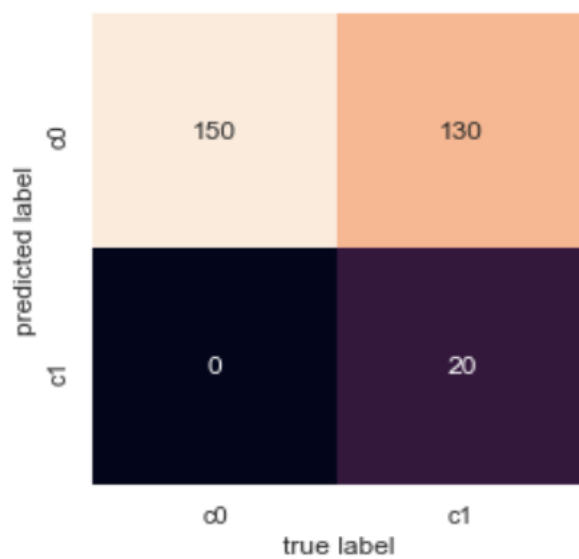
Accuracy= 0.6766666666666666

For n_init=10 and max_iter=100:



Accuracy= 0.63

For n_init=10 and max_iter=100:



Accuracy= 0.5666666666666667

The accuracy decreases as we increase either of the parameter n_init and max_iter.

Q.3) Multiclass classification.

Ans- In this section we are using handwritten digits data, to train the model.

Q .3.1) Do they actually correspond to the digits 0–9?

Ans- Yes, they correspond to the digits 0-9 except for the digit 8. The digit 8 does not look like it should.



Q.3.2) If you label each cluster with the digit that occurs most frequently within it, then what is your classification accuracy with this unsupervised method?

Ans- The accuracy of the model is 0.7952142459654981.

The accuracy for all the “digits” are-

Digit 0- 0.99438202247

Digit 1- 0.3021978021

Digit 2- 0.8505747126

Digit 3- 0.8516483516

Digit 4- 0.90055248618

Digit 5- 0.747252747252

Digit 6 - 0.9725274725

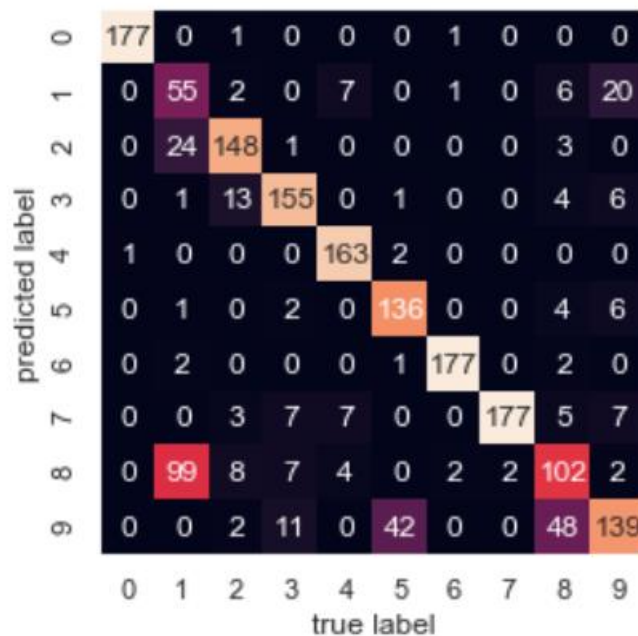
Digit 7 - 0.9725274725

Digit 8 - 0.5604395604

Digit 9 - 0.763736263736

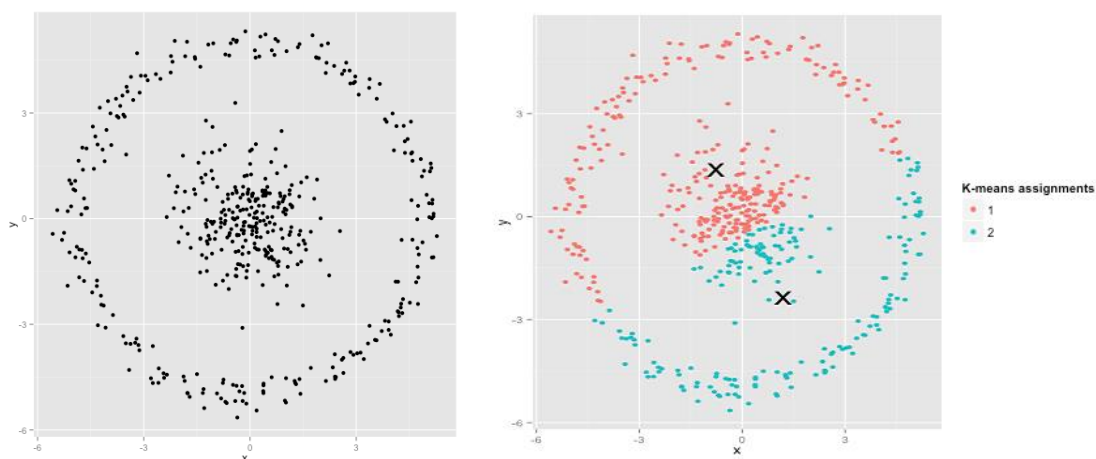
Q.3.3) What kinds of misclassifications are happening, and why?

Ans – If we look for digit 1 and 8 in the confusion matrix, the model misclassifies the digit 1 as digit 8.



It happens so, because k-means does not work for non-spherical data. Some non-spherical data are –

K-means classifies this as



Classifying the same digit dataset with k=5



Other than digit 0 (1st cluster) no other cluster clearly represent any digit.

The confusion matrix for k=5 –

predicted label	true label									
	0	1	2	3	4	5	6	7	8	9
0	176	0	1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	0	124	160	4	0	1	1	1	73	0
3	0	51	13	169	0	83	0	0	58	159
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	2	0	0	0	160	10	179	0	5	1
7	0	7	3	10	21	88	0	178	38	20
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

We get this result because the model is trying to group all the 10 classes into 5 classes. For example, digit 5 is predicted as digit 3 and digit 7. So, this result does not make any sense.