CS335(AIML) - Lab04 - Report

170050053 - Sudhir Kumar

October 30, 2021

CS 335: KMeans Implementation

1.1 - (ii) LOAD DATA 1

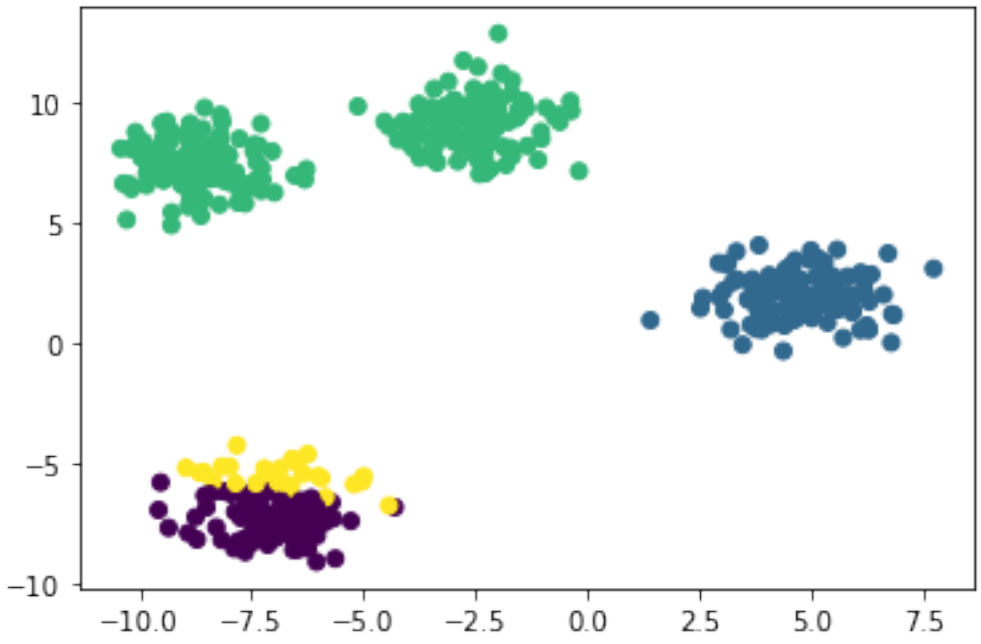

Figure 1: randomSeed == 10

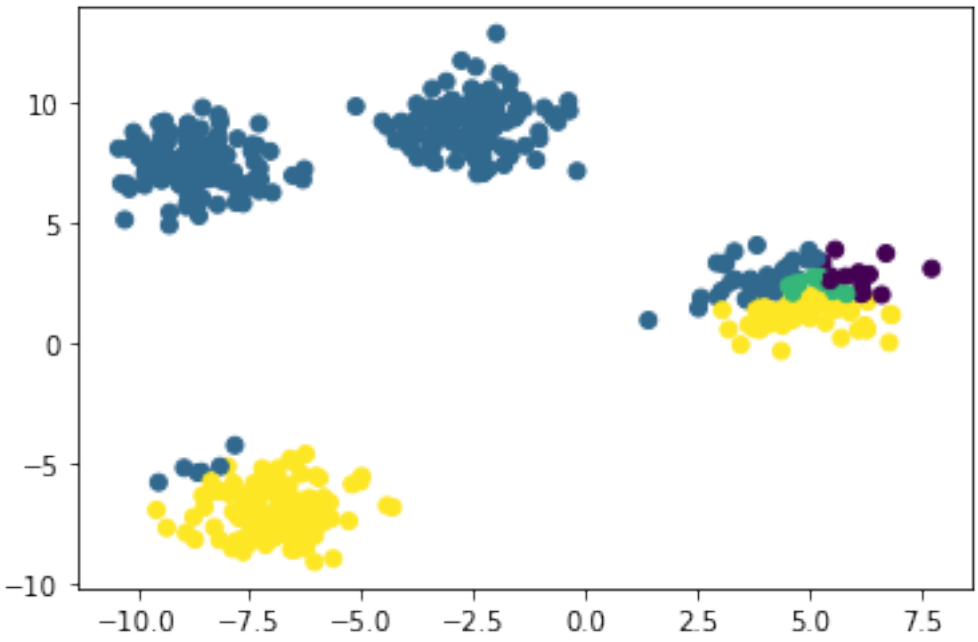

Figure 2: randomSeed == 1000

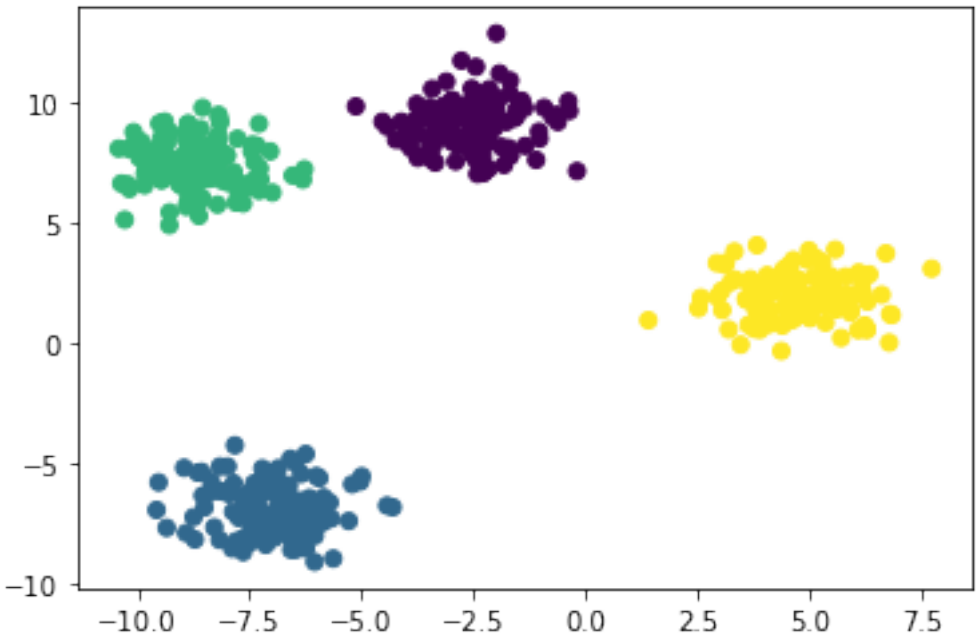

Figure 3: randomSeed == 120

1.1 - (ii) LOAD DATA 2

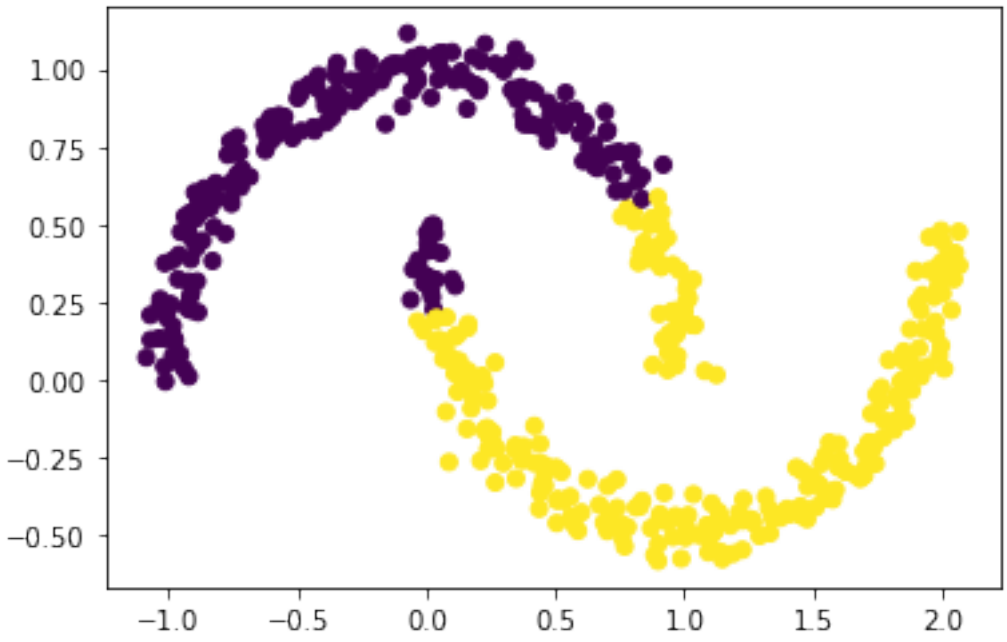

Figure 4: randomSeed ==10

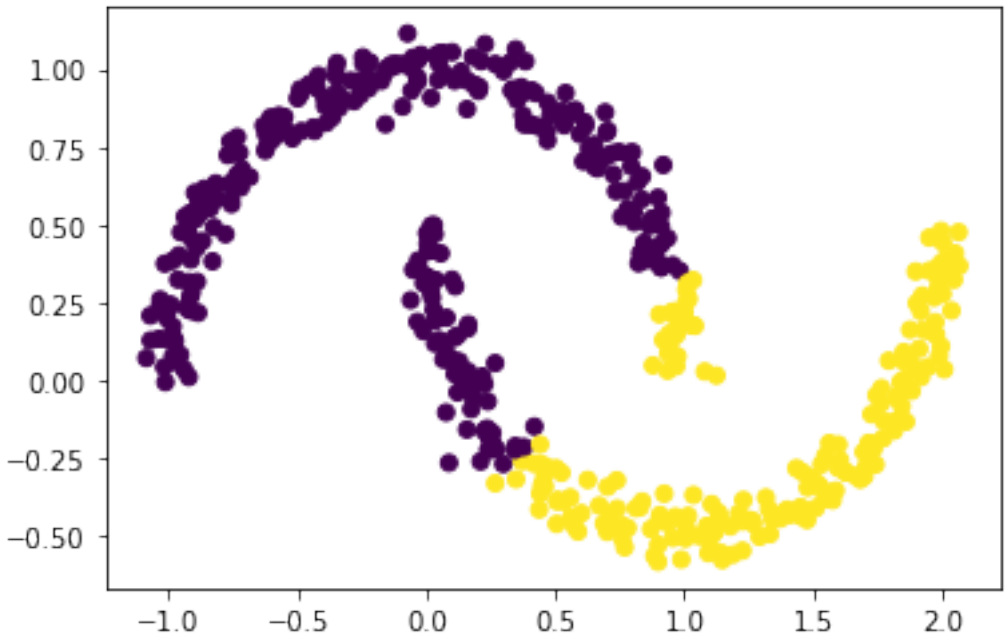

Figure 5: randomSeed == 1000

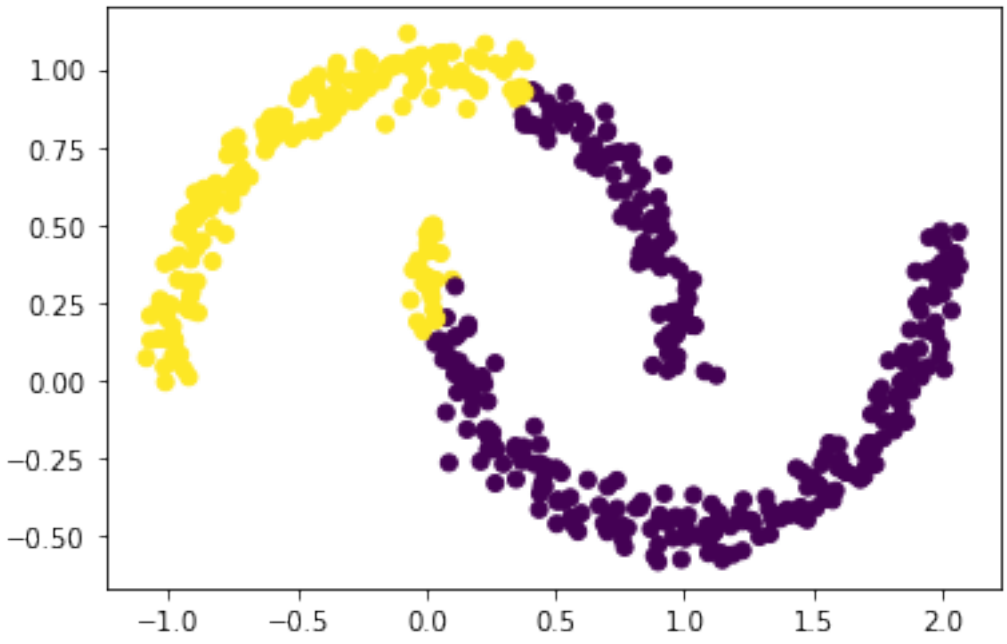

Figure 6: randomSeed == 120

1.1 - (ii) LOAD DATA 3

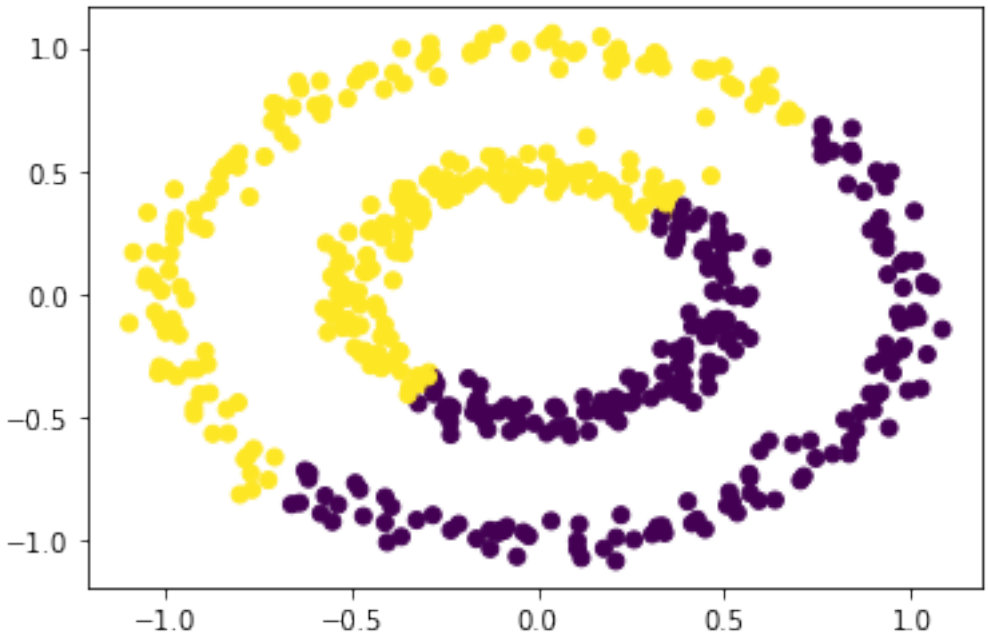

Figure 7: randomSeed == 10

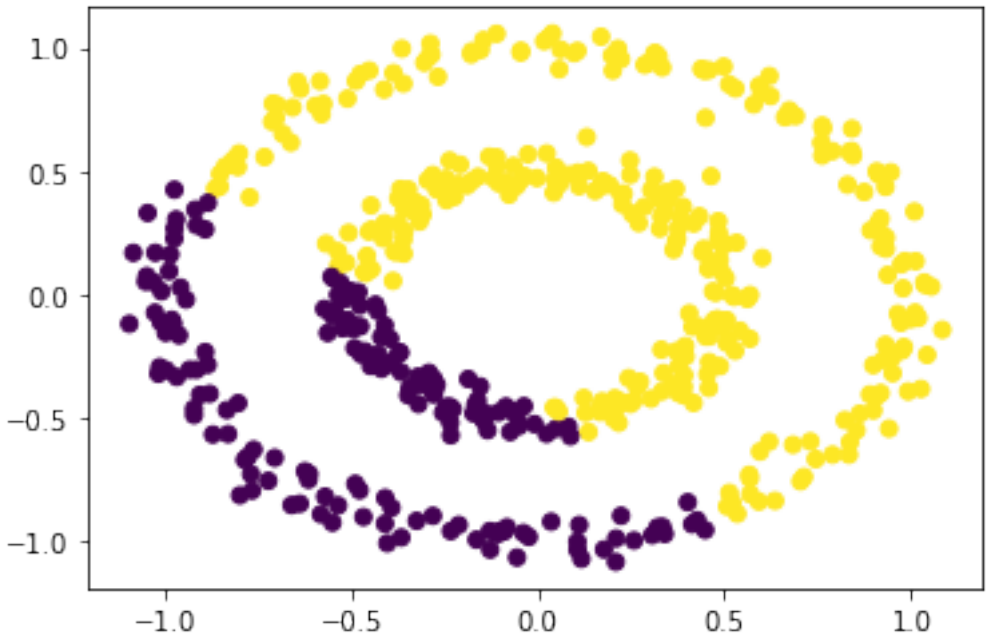

Figure 8: randomSeed == 1000

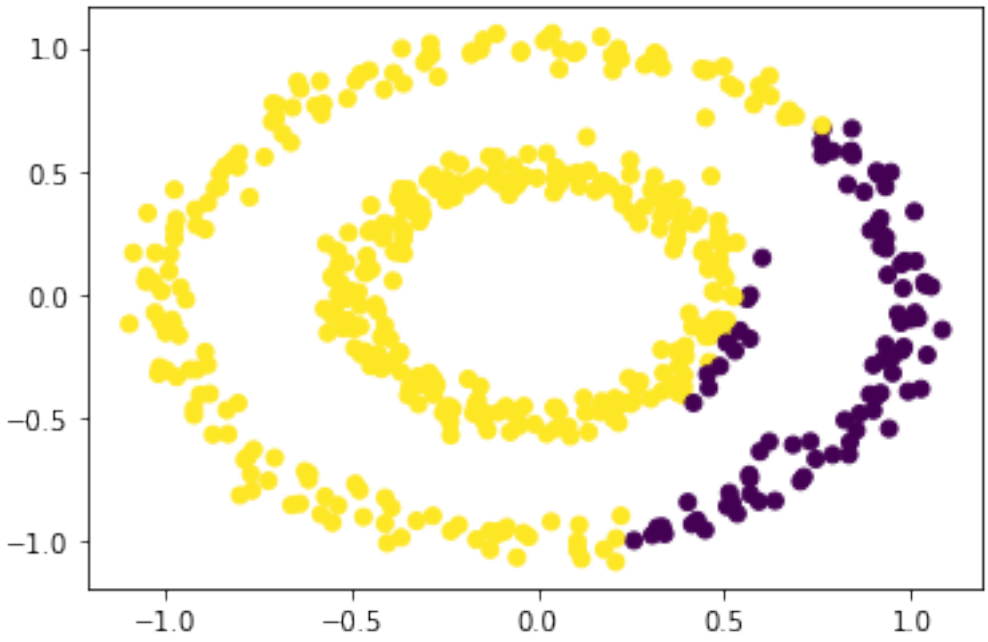

Figure 9: randomSeed == 120

Checked all three data with random seed = 10, 100, 120
first data is separable for random seed 120
rest are not separable for any above cases.

1.1 - (iii)

Q - What would be a good initialisation for the K-means algorithm? Briefly justify your choice.

Answer - For initialisation,
in clustering method of the k-means algorithm,
I think **centroid initialization** and iteration upon until it converge, is good choice.
But still it may be not good choice for some data-set.
We can find better initialization but those methods may be more complex, time consuming and intensive.
Still centroid initialization is simple and deterministic approach and most of the cases, It requires less iterations to reach convergence, so overall lesser run-time and a enough-efficient algorithm.

Simple Kernel Design

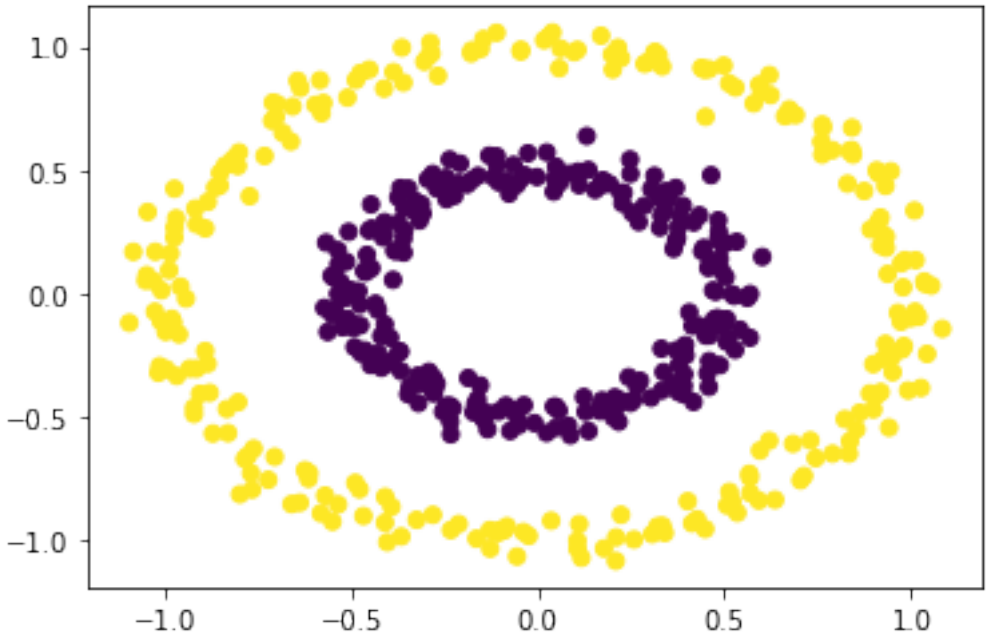

Figure 10: Simple Kernel Design