

TDA 231 Machine Learning: Homework 5 Solution Sketch

1 Practical problems

Problem 1.1 [*k*-Means Implementation, 20 points]

- (a) (7 points) • Make sure the centroid are updated correctly and the points updated to their closest centroid. A correct implementation grants half the grade. Further grading points can be gained by handling the notes below.
- The initial centroid should be generated randomly from the points already present in the dataset. Special care must be taken such as to avoid duplicate centroids. It means the random samples should be generated without replacement. This could be done for example in Python using the function `numpy.random.permutation` or `numpy.random.choice` with the parameter `replace` set to `False`.
 - The stopping condition of the main loop should avoid comparing exactly a *double* type. In particular, the students are expected to compare if the points assignment (of *integer* type) changed. If the centroids location are directly compared, a small epsilon difference should be allowed.
 - The code should be commented and variables name intuitive and easy to understand. Care should be made so as to avoid having an array variable changes its size at each iteration.
- (b) (3 points) • The plot should have a legend or a short explanation.
- The assignment change from the second iteration should be generated from the same run of the implemented kmeans. It means, it is not allowed to run two different instances of the kmeans, one with maximum iteration to 2 and the other till convergence
- (c) (5 points) • Special care must be taken so as to implement correctly the gaussian kernel as well as the distance of each point to the centroids.
- The code should be commented and variables name intuitive and easy to understand. Care should be made so as to avoid having an array variable change its size at each iteration.
- (d) (5 points) • Here it is expected that the plot of the RBF kernel kmeans should correctly clustered the points. Two co-centric circles are expected. The inner circle should have points in the same color A. The outer circle points should have all the same color B (color A different than color B)