# CO553 - Questions k-NN

Antoine Cully

October 2019

## 1 Implementation

You can use numpy in python (among other tools) to implement a k-NN algorithm. The IRIS dataset (provided in the dataset document) can be used to evaluate the classification capabilities of your k-NN implementation and the Robot dataset can be used as a regression task.

## 2 Questions

Here is a set of various questions to improve your understanding of k-NN.

1. In k-NN, the query time is longer than the training time.
   - A TRUE
   - B FALSE
2. Which of the following option is true about k-NN algorithm?
   - A It can be used for classification
   - B It can be used for regression
   - C It can be used in both classification and regression
3. Which of the following is true about Manhattan distance?
   - A It can be used for continuous variables
   - B It can be used for categorical variables
   - C It can be used for categorical as well as continuous
   - D None of these
4. What are the appropriate things to do with k-NN when you have noisy dataset?
   - A I will increase the value of k
   - B I will decrease the value of k
   - C k does not dependent on the noise
   - D None of these
5. In k-NN, what is the effect of increasing/decreasing the value of k?
   - A The boundary becomes smoother with increasing value of K
   - B The boundary becomes smoother with decreasing value of K
   - C Smoothness of boundary does not dependent on value of K
   - D None of these
6. For embedded applications (i.e., running on a smartphone), what is the most appropriate family of algorithm?
   - A Eager learners
   - B Lazy learners

7. Given $d$ the distance between a point of the dataset and the query point. Which of the following weight function is appropriate for Distance-Weighted k-NN?

   A $w = \exp(-d)$
   B $w = -\log(\min(0.25 * d, 1.0))$
   C $w = -d$

8. Which of the following statements is true for k-NN classifiers?

   A The classification accuracy is better with larger values of k
   B The decision boundary is smoother with smaller values of k
   C The decision boundary is linear
   D k-NN does not require an explicit training step

9. The curse of dimensionality only affects k-NN.

   A TRUE
   B FALSE

10. In Figure 1, sketch the 1-nearest neighbour decision boundary for this dataset.

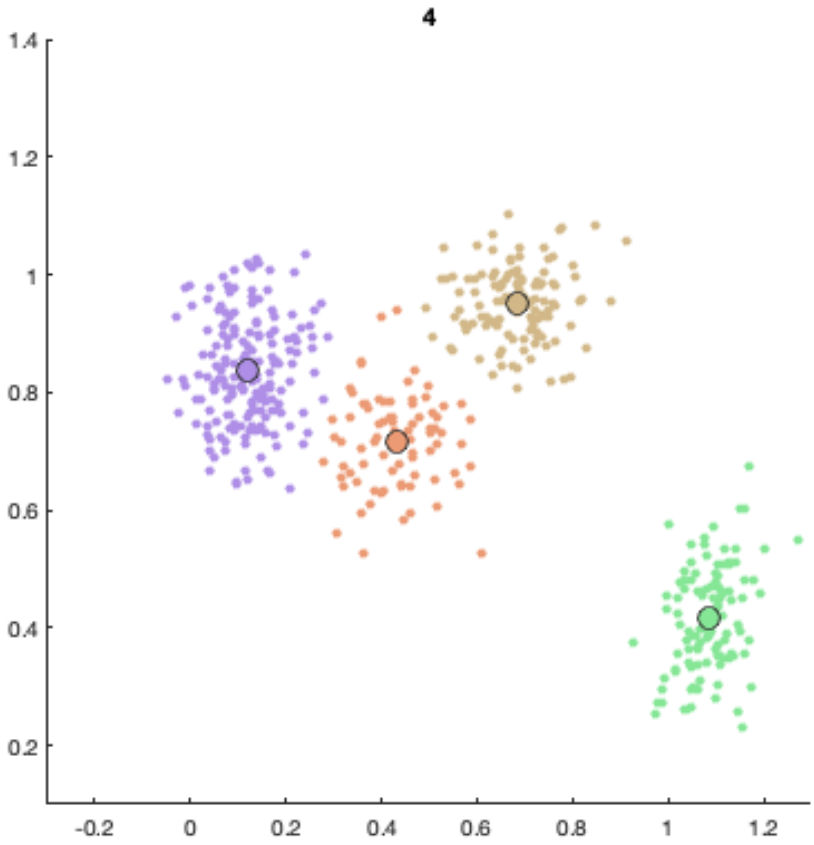


Figure 1: 4classes dataset.