

Exercises – Topic1 – Classification using kNN

1. Apply knn to the iris dataset to classify to the three types of iris – setosa, versicolor, virginica. How does the accuracy compare with that of knn?

Suggestions: i) remember that the dataset is arranged in the three classes in sequence, you need to shuffle and randomize before creating training and test datasets; ii) consider a 80/20 split of the data to training/ test dataset.

Solution: a solution is shown here: <https://www.youtube.com/watch?v=GtgJEVxl7DY> (check after you have worked on the exercise on your own).

2. Apply kNN to the following heart disease dataset to predict the presence (or not) of heart disease: <https://archive.ics.uci.edu/ml/datasets/heart+Disease>

A description of the dataset is provided. From the data folder, use the processed Cleveland dataset. The last column #14 is the outcome – angiographic disease status: 0 (no disease) or 1, 2, 3, 4 (yes). Look for the optimum k.