

Practical Session 1

k -Nearest Neighbors & Decision Trees

1 k -Nearest Neighbors

Problem 1: Consider a dataset with 3 classes $\mathcal{C} = \{A, B, C\}$, with the following class distribution $N_A = 42, N_B = 67, N_C = 27$. We use unweighted k -NN classifier, and set k to be equal to the number of data points, i.e. $k = N_A + N_B + N_C =: N$. What can we say about the prediction for a new point x_{new} ?

- A) x_{new} will be classified as class A
- B) x_{new} will be classified as class B
- C) x_{new} will be classified as class C
- D) We don't have enough information to answer the question

How about if we use the weighted (by distance) version of k -Nearest Neighbors?

2 Decision Trees

Problem 2: The plot below shows data of two classes that can easily be separated by a single (diagonal) line. Does there exist a decision tree of depth 1 that classifies this dataset with 100% accuracy?

