BRAC University: ML Lab - K-Nearest Neighbors

Zadid Hasan

August 19th, 2023

Here are the tasks that you should complete for today's lab:

1. Use the same titanic dataset we used in the first lab and do the same kind of pre-processing we did previously.
2. There is no training in the K-nearest neighbors algorithm.
3. In the testing section, take any test data a.
4. Let's assume that we have n features.
5. We have to figure out whether a survives or not.
6. Let's say, we have m data points in the training dataset. For each $i = \{1, 2, \cdots, m\}$, we calculate the distances:

   $$d_i = \sqrt{\sum_{j=1}^{n}(x_{i,j} - a_j)^2}$$

   This is the distance from a to all the m training data points.
7. From these, choose the $K = 5$ data points that are closest to a and notice where they survived or not. If the majority of them survived, then your output for a should be "survived" and if the majority of them didn't survive, you output should be "not survived".
8. In this way, calculate the accuracy of your algorithm. Since, we are choosing the k-nearest neighbors of a test data a, it's called K-nearest neighbor algorithm.