

Artificial Intelligence

Learning: k Nearest Neighbor

Name : Muhammad Husain Toding Bunga

Student ID : 1301153626

Class : IF39-INT

Lecturer SYM

Problem Description

In this third program assignment, we are given 1000 data where we have to define its class based on 4000 given training data. Each data will have 4 parameters, like, provocation, comments, and emotions ranged from 0 to 100.

Designed Method

For this problem, I will be using an algorithm called k Nearest Neighbor(kNN) to solve the problem. The algorithm will find the similarity between the data test and the training data and pick k most similar and will define the data's class based on the most dominant class of all picked data. But to determine how much most similar data to pick we need to do a cross validation to the train data.

By using brute force technique to generate some results using cross validation, I find that $k=101$ with fold=4 will give 70.12% accuracy. Even though this might not be the most optimum, I will use this value for the test data.

I will attach files for the cross validation results and the file for the knn results along with this document