**Computer science IA-K nearest neighbor**

Dawei Xiong

RDFZ S2C6

**Criterion A: Planning**

**Defining the problem**

Every day we may encounter some problems that are related to classification which can be minor or important. For example, how can we classify the type of the movie based on the scene presented in the movie. In fact, we can judge whether a movie is action or romantics based on the number of kiss scene. It can be something really important such as European refugees. Based on the number refugees that went into European continent, we can judge which country they get in.

**The proposed solution should take care of the issues such as:**

1. The K parameter in the algorithm is really crucial, for it can have a significant impact on the accuracy on the prediction. Thus, choosing a proper K is critical.
2. This algorithm can help people classify things when it is hard to classify. But this algorithm can give you the possibility that whether one thing belong

**Justification for the proposed product:**

I decided to use python to write the program, because this is the only programming language that I’m familiar so far.

**Criterion B: Design**

|  |  |  |
| --- | --- | --- |
| The default value is none, so you have to have input to get the correct answer | | |
|  |  |  |
|  |  |  |

**Criterion C: Development**

The entire algorithm is a Python program which is the most basic machine learning algorithm. KNN intend to classify the type of a set of data based on the data the algorithm has after training. The input is the data you want to classify and the output is the most possible category of the data.

**List of Techniques:**

* Evaluating user input
* Arrays & Matrix
* Data processing
* Reading a file
* Parameter passing
* Data visualization
* Result output

Program Structure:

The entire program is the KNN() class that contain three different function

* Classify() which is the core algorithm
* Vslz() which is the visualization function
* Open\_file() which can input excel data into python to process
* AutoNorm() which is used to evaluate the accuracy of the result of the algorithm

This structure is used to make the functionality of the program clear and easy for people to use (it would be better if I have time to design GUI ) both for developer and users.

Algorit

**Criterion D: Development**

**Criterion E: Evaluation**

**Recommendations for further development**

**GUI**: add graphical user interface to this algorithm to help ordinary user use it easily since not everyone would use IDE.

**Three dimensional**: until till now my algorithm can only handle data with two attributes, and the calculations largely rely on matrix. What if we need to classify an object with more than two attributes, the algorithm may become harder but the essence is the same.