* Results of Local Experiments:

Correct Label:

unknown08.txt:C7

unknown09.txt:C4(more)/C1

unknown04.txt:C1

unknown10.txt:C1(more)/C4

unknown05.txt:C4

unknown07.txt:C7

unknown06.txt:C4

unknown02.txt:C1

unknown03.txt:C1

unknown01.txt:C1

Local experiments:

* k=11 100% correct, but too much noise

unknown08.txt is labeled as follows

C4:9.1% C7:72.7% C1:18.2%

unknown09.txt is labeled as follows

C4:63.6% C1:36.4%

unknown04.txt is labeled as follows

C4:18.2% C7:9.1% C1:72.7%

unknown10.txt is labeled as follows

C4:36.4% C1:63.6%

unknown05.txt is labeled as follows

C4:72.7% C7:18.2% C1:9.1%

unknown07.txt is labeled as follows

C4:9.1% C7:72.7% C1:18.2%

unknown06.txt is labeled as follows

C4:72.7% C7:9.1% C1:18.2%

unknown02.txt is labeled as follows

C4:27.3% C1:72.7%

unknown03.txt is labeled as follows

C4:36.4% C7:18.2% C1:45.5%

unknown01.txt is labeled as follows

C4:27.3% C7:9.1% C1:63.6%

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

* k=8 100% correct, still some noise

unknown08.txt is labeled as follows

C4:12.5% C7:87.5%

unknown09.txt is labeled as follows

C4:75.0% C1:25.0%

unknown04.txt is labeled as follows

C1:100.0%

unknown10.txt is labeled as follows

C4:25.0% C1:75.0%

unknown05.txt is labeled as follows

C4:100.0%

unknown07.txt is labeled as follows

C7:100.0%

unknown06.txt is labeled as follows

C4:87.5% C1:12.5%

unknown02.txt is labeled as follows

C1:100.0%

unknown03.txt is labeled as follows

C4:37.5% C7:12.5% C1:50.0%

unknown01.txt is labeled as follows

C4:25.0% C1:75.0%

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

* k=6 100% correct, clearest (best k choice)

unknown08.txt is labeled as follows

C7:100.0%

unknown09.txt is labeled as follows

C4:83.3% C1:16.7%

unknown04.txt is labeled as follows

C1:100.0%

unknown10.txt is labeled as follows

C4:33.3% C1:66.7%

unknown05.txt is labeled as follows

C4:100.0%

unknown07.txt is labeled as follows

C7:100.0%

unknown06.txt is labeled as follows

C4:100.0%

unknown02.txt is labeled as follows

C1:100.0%

unknown03.txt is labeled as follows

C4:33.3% C7:16.7% C1:50.0%

unknown01.txt is labeled as follows

C1:100.0%

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

* k=4 100% correct, but unknown10.txt becomes unclear

unknown08.txt is labeled as follows

C7:100.0%

unknown09.txt is labeled as follows

C4:100.0%

unknown04.txt is labeled as follows

C1:100.0%

unknown10.txt is labeled as follows

C4:50.0% C1:50.0%

unknown05.txt is labeled as follows

C4:100.0%

unknown07.txt is labeled as follows

C7:100.0%

unknown06.txt is labeled as follows

C4:100.0%

unknown02.txt is labeled as follows

C1:100.0%

unknown03.txt is labeled as follows

C7:25.0% C1:75.0%

unknown01.txt is labeled as follows

C1:100.0%

In summary, I selected k=6 in the end. Accuracy on test set: 100%

* Confusion matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| File | KNN Prediction | FuzzyKNN Prediction | Real | True/False | Precision | K Choice/Similarity |
| unknown01.txt | C1 | C1:100.0% | C1 | TRUE | 100% | K=6 | Euclidean |
| unknown02.txt | C1 | C1:100.0% | C1 | TRUE |  |  |
| unknown03.txt | C1 | C4:33.3% C7:16.7% C1:50.0% | C1 | TRUE |  |  |
| unknown04.txt | C1 | C1:100.0% | C1 | TRUE |  |  |
| unknown05.txt | C4 | C4:100.0% | C4 | TRUE |  |  |
| unknown06.txt | C4 | C4:100.0% | C4 | TRUE |  |  |
| unknown07.txt | C7 | C7:100.0% | C7 | TRUE |  |  |
| unknown08.txt | C7 | C7:100.0% | C7 | TRUE |  |  |
| unknown09.txt | C4 | C4:83.3% C1:16.7% | C4(more)/C1 | TRUE |  |  |
| unknown10.txt | C1 | C4:33.3% C1:66.7% | C1(more)/C4 | TRUE |  |  |