MiningLab WEI Chen

1) how many documents you have for each class. Moreover, provide some info about the source of the documents (e.g. Wikipedia, which website etc.)

16 text documents have been collected dealing with the fruit 'apple', and other 16 dealing with Apple comp

any.

All documents are collected from online news or online business website.

(Google News)(http://www.naturesbasket.co.in)

2) show the accuracy for at least 3 different values of k. Does the accuracy improve as a function of k? Do you expect it to increase or decrease as a function of k?

k	Accuracy
1	1.0
2	1.0
3	0.909
4	0.909
5	0.727
6	0.818
7	0.727
8	0.727
9	0.727
10	0.727
11	0.727
12	0.727
13	0.727
14	0.727
15	0.727
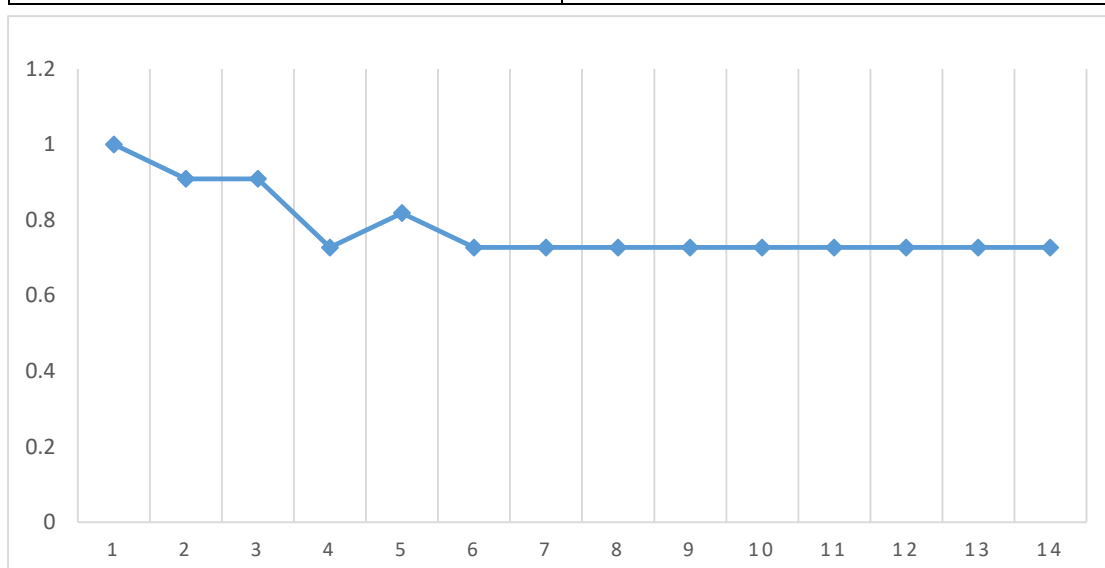
According to KNN algorithm, the parameter k should be properly chosen. If k is too small, the algorithm could be sensitive to noise points; if k is too large, neighborhood may include points from other classes.

Thus, the performance of the algorithm is expected to increase with k, then reach a summit with certain value of k, after that decrease with k till convergence.