Name: Shikhar Mohan
Roll No.: 18EC10054
Date: 3/11/2021

Observations:
Results on all hyperparameters can be found upon executing the main.py file as well.

For k = 1
Cosine similarity accuracy: 0.6857
Normalized euclidean distance accuracy: 0.6857
Euclidean distance accuracy: 0.9571

For k = 3
Cosine similarity accuracy: 0.6714
Normalized euclidean distance accuracy: 0.6857
Euclidean distance accuracy: 0.9429

For k = 5
Cosine similarity accuracy: 0.6714
Normalized euclidean distance accuracy: 0.6857
Euclidean distance accuracy: 0.9429

For k = 7
Cosine similarity accuracy: 0.6714
Normalized euclidean distance accuracy: 0.6857
Euclidean distance accuracy: 0.9429

Best hyperparameter:
The best performing hyperparameter here is clearly k = 1 and Euclidean distance metric with 95.714% accuracy.

Analysis:
1. Since kNN depends upon the location of the n-dimensional features in n-space, normalisation leads to loss of information. This loss harms the model's performance greatly, which is why un-normalized euclidean distance is the best performing distance metric (as well as default standard practise when using the kNN algorithm in practical situations).
2. Moreover, it can be seen that there is a class imbalance in the dataset, which is understandable since it is natural for there to be more benign tumors than malignant in any well sampled real world data regarding tumors.
3. The class imbalance is a reason for the 0.6957 accuracy being a common result. This result comes from predicting all tumors as benign, and since nearly 68.57% of the tumors in the test dataset are benign, we get this accuracy. Even though this is a decent accuracy metric on face value, from a realistic medical perspective it is meaningless since there is no point to a detector which labels every tumor benign. This is a flaw in the accuracy metric itself, which is why for problems like these other metrics such as f1-score, jaccard score, dice score, et cetera are used.