**Computer Assignment on K-NN**

The Cancer dataset is given in the csv format. The dataset has 9 attributes and class label for each instance as shown below: [In the class column 2 is for benign 4 for malignant cells]

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| clump  thickness | unif\_cell\_size | unif\_cell\_shape | marg\_adhesion | single\_epith\_cell\_size | bare\_nuclei | bland\_chrom | norm\_nucleoi | mitoses | class |

Divide the dataset into training set and test set.

[Hint: It can be divided by randomizing the indices and then splitting the dataframe according to the indices.]

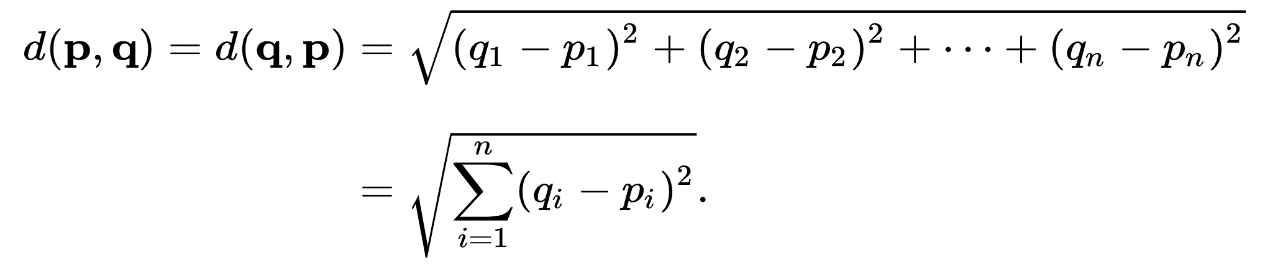
Define functions to compute the value of the distance metrics: Euclidean, Normalized Euclidean and Cosine Similarity.

Define and implement the function to return k-Nearest Neighbours with k=1, 3, 5 & 7 and predict the class of the Test data-set for each k value and each distance metric.

Compute the accuracy and Plot a bar chart to compare the performance of hyperparameters.

NOTE:

1. **Euclidean Distance**Euclidean Distance between two points p and q in the Euclidean space is computed as follows:



2. **Normalized Euclidean Distance**  
Normalized Euclidean distance is the Euclidean distance between points after the points have been normalized.

3. **Cosine Similarity**Cosine Similarity is the similarity measure between two non-zero vectors. Cosine Similarity between two vectors A and B is computed as follows:

