Image Classification with KNN

Asra Selen Özdemir
090200713
CRN: 11512

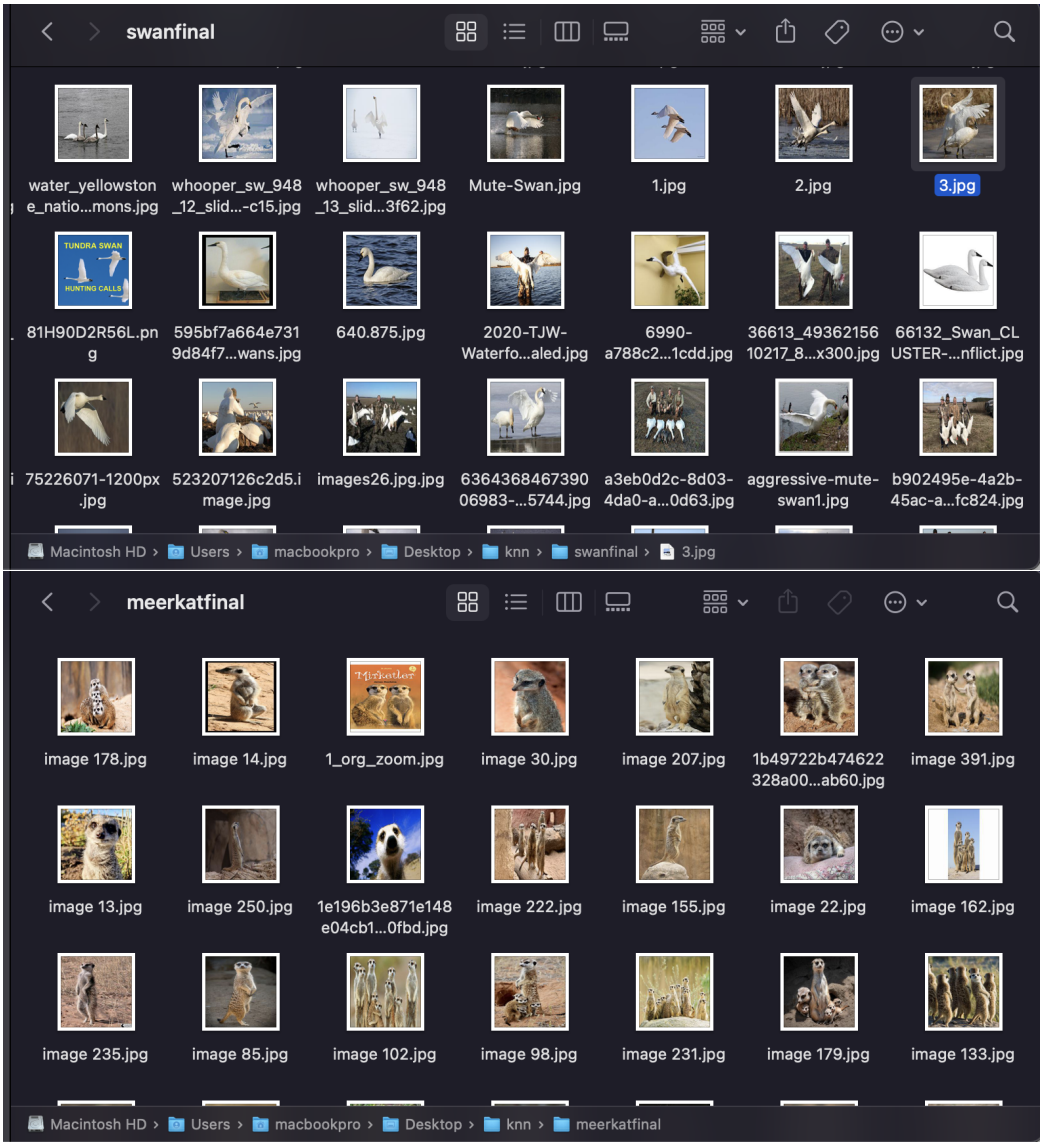
1. Aim

Aim of this homework is creating a dataset consisted of 2 types of images by collecting the data manually,classifying those images by using k-nearest neighbours method and observing the change of accuracy with respect to the chosen nearest neighbour number k.

2. KNN Algorithm

KNN is a supervised learning method which classifies the elements by labeling them according to labels of their closest k neighbours.Distance measurement can be done with different metrics but here, Euclidean distance($r=\sqrt{x^2+y^2}$)is used.

3. Data:

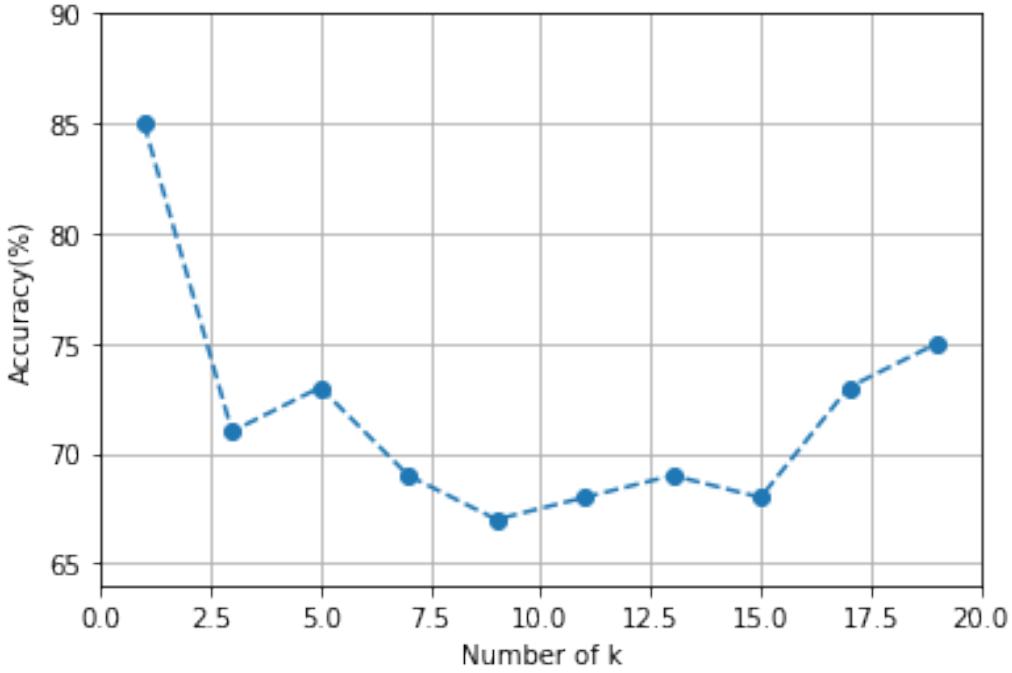


500 swan and 500 meerkat images are used in the dataset.

4. Analysis of Code

1:Upload the dataset,resize and turn every image to grayscale.
2:Flatten the image arrays.
3:Label the image arrays by adding a new element 0 for swan and 1 for meerkat.
4:Shuffle the rows and split the data into testing and training set.
5:Select 100 testing rows and calculate the Euclidean distance to find the resemblance between images.
6:Create a new matrix which contains distance and label information.Sort the matrix in an increasing fashion.
7:Determine the k and take the first k number of rows from the matrix.
8:Calculate the mean of the labels, if it is less than 0.5, label of the testing row is voted to become a "0" by other image labels nearby,otherwise label this testing row as "1".
9:Compare the calculated labels of all the test rows by their labels in original data matrix.
10:Calculate the accuracy by dividing the number of truly calculated labels to number of all testing rows.

5. Result



As it can be seen from the graph, accuracy is dramatically high around the interval of k=(2,2.5) due to the fact that there are actually 2 clusters,swan and meerkat.