Assignment 2

Topic: **K Nearest Neighbors**
Submission Date & Time: **5th Nov 2019 before start of class**

Provided Files:
Partial maps of Italy with 10%, 20%, 30% 40% & 50% pixels filled as either Land or Water. Remaining pixels are blank/white. Furthermore complete map of Italy is also provided for reference.

Tasks:

1. Implement generic versions of Manhattan & Euclidean distance so that distance can be calculated between two n-dimensional vectors.
2. Implement KNN algorithm that can handle n-dimensional hyper-space with option for varying value of k.
3. Implement Confusion Matrix and calculate Accuracy, Sensitivity & Specificity etc.
4. Implement Jaccard Similarity Index to calculate similarity between two sets.
5. Combine the above parts to run the KNN algorithm on the provided partial maps to predict each blank pixel as either Land or Water.
    i. Use the values 1, 3, 5, 7, & 9 for K to see which value of K works best given the different fill levels.
    ii. Also you should compare the two distance measures (Manhattan & Euclidean) for calculating the distance between points for each K.
    iii. Generate Confusion Matrix for each case to calculate the Accuracy of the algorithm for the given parameters.
    iv. Calculate Jaccard Similarity Index for each combination of parameters for evaluation.

You can use the complete map for reference in your calculations.

Submission:
Submit a zip file containing the following

1. Complete Python (or language of your choice) code.
2. Map images generated by each combination of k and distance measures. Theses files should be properly named (naming convention: Italy_xx-yyyy-k_z.png where xx = percentage of source file, yyyy = name of distance algorithm & z = value of k used)
3. Detailed report of your findings in Word or PDF format with Confusion matrix, Accuracy & Jaccard Similarity for each case along with final analysis of which combination of parameters perform better for which input case. Your report should also include images & tables as appropriate. Use 2 column standard IEEE conference proceeding format for your report. The report should include at least the following sections
    a) Title
    b) Author Name, Roll Number, Email & Affiliation
    c) Abstract
    d) Introduction
    e) Dataset Description
    f) Methodology & Implementation Details
    g) Results
    h) Discussion
    i) Conclusion
    j) References (IEEE standard format)

Sample Python Code

```
import numpy as np
import skimage.io as skio
import matplotlib.pyplot as plt

# read an image file in variable from current directory
img = skio.imread('Italy30.png')

# display an image on screen
skio.imshow(img)

# save an array as an image file in current directory
skio.imsave(arr, 'filename.png')

# get the dimensions of the 3-d array
rows, cols, chans = np.shape(img)

# create a new array of unsigned 8 bit integers of size 10 x 15 x 5
arr = np.array(shape = (10, 15, 5), dtype = np.uint8)

# create a new array filled with zeros of unsigned 8 bit integers of size 10 x 15 x 5
arr = np.zeros(shape = (10, 15, 5), dtype = np.uint8)

# create a new array filled with 100 of unsigned 8 bit integers of size 10 x 15 x 5
arr = 100 * np.ones(shape = (10, 15, 5), dtype = np.uint8)

# create a new array of booleans filled with false of size 10 x 15 x 5
arr = 100 * np.zeros(shape = (10, 15, 5), dtype = bool)
```