**Unsupervised Image Segmentation**

Predictive Modeling Report

3/1/2021

1. Define the Predictive Modeling Problem

1. Input:

The data we are using comes from two separate datasets. The first is the Pets images dataset. There are 7,378 RGB images of cats and dogs in a variety of settings. The images range in sizes but are all resized to 240 x 160 x 3.

The second dataset is the Drone Images dataset. This dataset consists of RGB images taken from a drone anywhere from five to thirty meters in the air and are images of a variety of urban scenes. The dataset has 25 labeled classes such as tree, car, dirt, rocks, or pool. All images in this dataset are resized to 576 x 384 x 3. Both datasets are normalized to have the pixels in the range from zero to one.

Each image is the only input into the K-means algorithm. K-means only accepts a two dimensional input so the input is the height\*width x number of channels. We attempted multiple arrangements of features to feed into the K-means model for both datasets. The first was an original image, a conversion of the original image to hsv, and a conversion of the original image to lab. These conversions are concatenated as channels of the original image taking it from three channels to nine. The next set of features that we used as input to our K-means algorithm was the output of the autoencoder as well as the original image. The last set of features that we used as input was the output of the autoencoder, the output of our self-organizing maps, and the original image.

1. Output:

The Pets dataset the output we want to achieve is to have the dog or cat segmented from the background. We attempt to have the background labeled into one or two colors while the cat or dog is labeled as a third.

The Drone Images dataset we want to achieve something slightly different than with the Pets dataset. We give the K-means algorithm the number of labels in the ground truth mask image for the number of clusters to segment the image. The output we want to achieve is to label all matching sections of the image as the same color.

2. Predictive Models

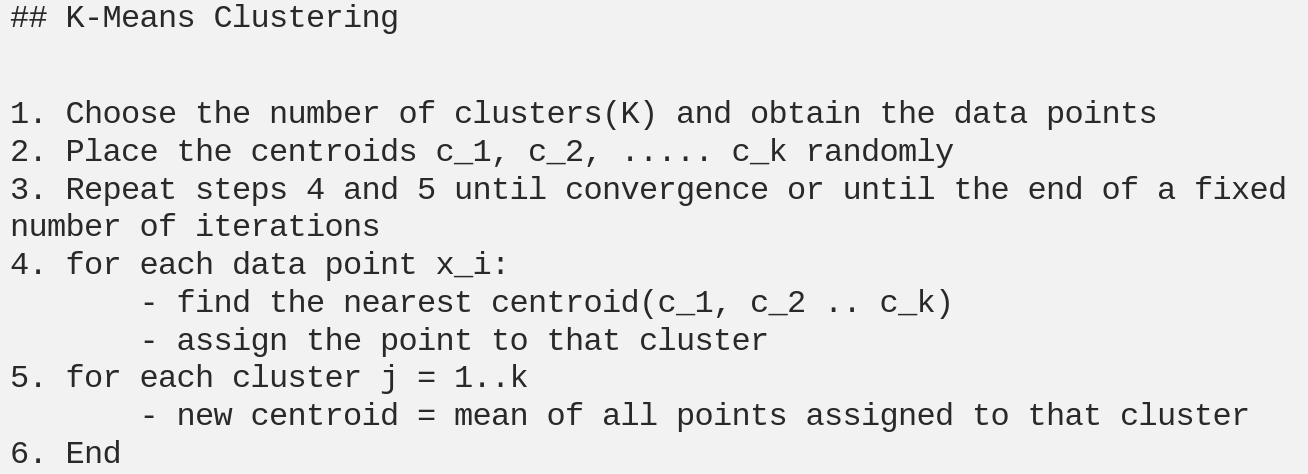
1. What are the methods? Give a general introduction of the methods with references

The predictive model that we used was the K-means algorithm to segment the images. K-means clusters data points based on the similarity or closeness between the data points. For image segmentation K-means does this for every pixel and clusters the data points based on how many clusters we tell the algorithm to use. The way K-means finds the distance between data points is by using the Euclidean distance. [1,2]

1. Describe the methods with a pseudo code using the definitions in Section 1.

The steps for the pseudo code and pseudo code for K-means is as follows:

* 1. First choose a value for the number of clusters
  2. Initialize the centroids randomly
  3. Calculate the Euclidean distance from the centroids and form clusters
  4. Find the centroid of each cluster and update centroids
  5. Repeat from step c



1. Justify the choice of the method.

K-means is an appropriate choice for unsupervised semantic segmentation. Given an image and the number of clusters we want the algorithm to segment the image into the way that the K-means algorithm works makes it appropriate to use. We assume that the pixels of the segments of the image will be closer in distance to each other than other segments in the image which makes this an appropriate choice of method.

3. Evaluation

1. What metrics do you use for evaluation?

At this point in time the only metrics we are using for evaluation is visual inspection of a few random examples to inspect how well our models do the segmentation. We compare the segmented image to the original image and can get a good understanding as to how well the the model does in segmenting the original image. We can attempt to use Mean Intersection Over Union (mIOU) to get a better understanding of exactly how well the model’s segmentation compares to the ground truth images.

1. What is your ground truth?

The ground truth images for the Pets dataset is a segmented version of the original image. The ground truth image consists of one of three class labels for each pixel in the image. A pixel can be labeled as background, outline, or pet.

The ground truth images for the Drones Images dataset is a segmented version of the original image. The ground truth image consists of one of 25 class labels for each pixel in the image. A pixel can be labeled as paved, dirt, grass, gravel, water, rocks, pool, vegetation, roof, wall, window, door, fence, fence-pole, person, dog, car, bike, tree, bald-tree, ar-marker, or obstacle.

1. Discuss the performance and the limitation of the method.

Based upon visual inspection of the models output it appears that the K-means algorithm does a fairly decent job on segmenting most images in both datasets. It appears that K-means does the best job at segmentation when it is given the features from the original image, the autoencoder, and the self-organizing maps. Visual inspection of a segmented image compared to the ground truth masked image is not the most accurate metric. Visual inspection can be biased by the individual that is doing the inspection. Work still needs to be done to implement mIOU on the datasets to get a more concrete value on how well the model is actually performing compared to the ground truth masks.

**Appendix**

**1. https://sigitwidiyanto.staff.gunadarma.ac.id/Downloads/files/38034/M8-Note-kMeans.pdf**

**2. https://doi.org/10.1348/000711005X48266**