**Machine Learning Week 2 Progress Report**

Group Members:

Aarya Parekh AU2120075

Diya Patel AU2120193

Kavyan Savani AU2120231

Vatsal Shah AU2040082

**04-02-2024**

In the previous week, we saw as to how Barmpounakis et al [1] implemented the Jenks Clustering algorithm, we wanted to familiarize ourselves with the clustering technique itself, so we researched what all different clustering algorithms are there and made an extensive list of which ones to explore in-depth. In the end, we reached the conclusion that we should go through these four algorithms:

1. Jenks Optimization method
2. DBSCAN method
3. K-means clustering
4. Spectral clustering

Even though the methods other than [1] that we analysed, do not seem like a proper fit for the task, they gave us a better insight as to why the authors chose Jenks Optimization method.

**05-02-2024**

**Jenks Optimization method:**

We went through the different implementations of Jenks algorithm in python. Jenks optimization method is the machine learning equivalent of min-max algorithm. The difference being that Jenks optimization tries to minimize the variance within the classes, and maximize the variance between the classes.

**06-02-2024**

**DBSCAN method:**

DBSCAN essentially works by identifying core points, which have at least a minimum number of point neighbours within a distance of ε, and expanding clusters by connecting core points to their neighbouring points. Points that are not core points themselves but lie within the ε-neighbourhood of a core point are considered border points and are assigned to the cluster of the core point. Points that are neither core points nor border points are considered noise. This might not be the best approach because of how the algorithm clusters based on proximity.

**08-02-2024**

**K-means clustering:**

K-means clustering, firstly selects K cluster centroids which are the basis of the K-clusters found by the algorithm. It is important to note that the accuracy of the algorithm is dependent on the selection of these initial K centroids. We think even this method might not be suitable for the task of lane detection.

**Future Goals/Aims:**

Our target is to understand spectral clustering, maybe Gaussian Mixture Model as well, by the end of the next week and select the clustering method we would like to opt for. Within the timeframe of this and the next submission we also aim to have one working implementation of the algorithm we choose on a smaller dataset (*not* pNEUMA) with known ground-truth so we can evaluate it better.

**References**

1. Barmpounakis, Emmanouil, Guillaume M. Sauvin, and Nikolaos Geroliminis. "Lane detection and lane-changing identification with high-resolution data from a swarm of drones." Transportation research record 2674.7 (2020): 1-15.