Richard Johnson

L20455045

Assignment 3

Dataset used:

Airfoil Self-Noise Data Set

<https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>

Source:

Provide the names, email addresses, institutions, and other contact information of the donors and creators of the data set.

Donor:

Dr Roberto Lopez

robertolopez '@' intelnics.com

Intelnics

Creators:

Thomas F. Brooks, D. Stuart Pope and Michael A. Marcolini

NASA

Data Set Information:

The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments.

Attribute Information:

This problem has the following inputs:

1. Frequency, in Hertzs.

2. Angle of attack, in degrees.

3. Chord length, in meters.

4. Free-stream velocity, in meters per second.

5. Suction side displacement thickness, in meters.

The only output is:

6. Scaled sound pressure level, in decibels.

Preprocessing:

1. I added the column names to match the attributes described in the source documentation.
2. I checked data for any null values, there were none
3. Since the dataset had no missing values and all attributes were numerical, no additional preprocessing was required

K-Means:

I used several K values ranging from 3-4 and with a distance value of +1 to +3 of K value for each K value.

DBSCAN:

I was not able to get the DBSCAN to plot for visualization.

Parameters chosen:

eps=0.05, min\_samples=5

Estimated number of clusters: 30

Estimated number of noise points: 1322

I was not able to get any different information from changing the epsilon or the minimum number of samples. I am not sure where the problem in the code lies. I tried different versions and was not able to get the model to change its values.