Course: INT 93S
Quarter: Summer 2017
Assigned: 11 July 2017
Due: 9:00 13 July 2017

Lab 3: Analysis and Visualization of Data

Turn-in:

1. Create a readme file titled “lastname_firstname_readme.txt” which includes:
    a. First name and last name of you and your partner
    b. How much of the lab you finished in class
    c. Any references you used
    d. Anything you would like the professor and TA to know.
2. Put your code and report in a folder named “lastname_firstname_lab4”
3. Zip the folder and upload it to Gauchospace

Instructions:

1. Import the data from the file: flea.csv *(1pt)*
2. Hypothesis: *(1pt)*
    a. Plot the dataset as a scatter plot with the labels.
    b. Which clustering algorithm do you think will work the best for this dataset? Why?
3. Analysis: *(6 pts total)*
    K-Means, Agglomerative, DBSCAN and Gaussian Mixture model clustering algorithms
    a. Normalize the data range for each column to be between 0 and 1. *(1 pt)*
        temp = column - np.min(column); normalized = temp / np.max(temp)
    b. Plot the K-means and Agglomerative results as 1 figure with 2 subplots *(2 pts)*
    c. Plot DBSCAN results in figure with 4X4 subplots *(1 pt)*
        Try eps values [.11,.12,.13,.14] and min_samples values [2,3,4,5]
    d. Plot the Gaussian Mixture model results. *(1 pt)*
        Color points accordingly and include ellipse for distribution
    e. Answer these questions: *(1 pt)*
        i. What parameters did you choose for each of the four algorithms? Why?
        ii. How accurate was each algorithm?
            (# of correctly assigned points) / (# of points)
4. Conclusion: *(1 pt)*
    a. Which clustering algorithm gives the best results for this data? Why?
5. Format: *(1 pt)*
    a. In a document file (word, google docs, etc), write up a lab report (1-2 pages).
    b. Your report should have 3 sections, a hypothesis, analysis and conclusion.
    c. Include the plots from the hypothesis and analysis sections.
    d. All plots in this lab must include a proper title, labels, and a legend.
    e. Export Lab report as a pdf saved as “lastname_firstname_lab4_report.pdf”
    f. Follow turn-in instructions, including the readme

Extra credit: *(1 pt)*

1. Choose your own dataset that has 2-3 numeric dimensions.
2. Follow steps 1-5 from the lab, producing a second lab report.