

Homework

You are provided with the file 11-Homework.mlx, and the data file, *breast_cancer.csv* which is study conducted at the University of Wisconsin to classify mammograms based on features observed. You can learn more details of this study and the features from the website: [Wisconsin Breast Cancer Data Set](#). The MLX file includes a function, *breast_cancer_datamatrix()* that will read this data file and return the datamatix **X** and the response variable **y**.

You will evaluate k-Means and Agglomerative Clustering Techniques for this dataset.

Specifically, for the **k-Means method**:

- Evaluate the *purity of clusters* formed with $k = 2$ and $k = 3$
- Perform silhouette analysis
- Plot the Elbow Graph with SSEs

And for the **Agglomerative or Hierarchical Method**

- Display the clustering tree using Euclidean Distance with Linkage Analysis performed using *Average* Distance between clusters
- Perform *cophenet* similarity and average link *inconsistency* analysis
- Compare the *purity values* for clipping the hierarchy at 2, 3 and 4 maxclusters.

Provide a brief summary of your results.