Mia Rodgers: <https://github.com/miamrodgers/4310-ML/blob/main/ME8_clustering/clustering_basic.ipynb>

Alex Larsen: <https://github.com/alarsen123/ML-HW/blob/main/ME8_clustering/ME8_clustering/clustering_basic.ipynb>

ME8

For this exercise, we practiced using clustering algorithms on a synthetically generated moons dataset and the sklearn breast cancer dataset. On each dataset, we tested KMeans and DBSCAN algorithms. Based on the graphs for the moons dataset, the DBSCAN model seemed to work much better than the KMeans model with 2 clusters.

We then tried to find the best parameters for each model using the breast cancer dataset. For the KMeans data, we found that models with 2 clusters performed the best, having the highest silhouette score out of all the different numbers of clusters ranging from 2 to 9.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

For the max\_iter, we tested 1-9, with 5 max iters having the best accuracy and 1 having the best silhouette score. For n\_init, 1 had both the highest accuracy and silhouette score. Our final model ended up predicting 93.15% of samples correctly.

Then for DBSCAN, we found that eps = 0.2 and smaller and min\_samples being two or greater had correctly classified all of the labels.

From this ME we concluded that DBSCAN performed better than KMeans overall.