

Clustering Homework

All submissions:

- **QUESTIONS:** Edit readme with answers to the discussion questions
- **CODE:** Share clustering code:
 - Option 1: Link to online editor (Google Colab or Replit)
 - Option 2: Add **clusteringHW.py** file
 - Option 3: Add **clusteringHW.ipynb** file

Mild

- Complete the clustering algorithm for 1D data.
 - (Put the code from section 5 into a loop that repeats a certain number of times - Section 5 Option 5B)
 - (Feel free to comment out 5A and disregard 5C)
- Plot each of the 5 data sets.
- **Answer: “Q1: Which sets best show clustering?”**
- Choose 1 of the sets that show good clustering.
- In Section 4, update the line `centroids = np.random.choice(data, size=k, replace=False)` to use the `np.random.uniform` function instead of `choice`, using “low” as your lowest data value and “high” as the highest data value.
- Choose different k-values to best fit the data
- **Answer: “Q2: What were the centroids? How best did the k-value fit the data?”**
- **Answer: “Q3: What meaning can you extract from the clustering? If you could not, what does that tell us about the data?”**
- Submission:
 - Python code that displays final clustered graph with centroids
 - Readme file with Answers

Spicy

- Find a .CSV online and upload it into your project
 - (You can find lots of good data from [Kaggle](#))
 - (If using Colab, place it in the ‘content’ folder)
- **Answer: “Q1: What dataset .csv did you choose?”**
- Complete the clustering algorithm for 1D data.
 - (Put the code from section 5 into a loop that repeats until the centroids stop changing)
 - (Feel free to comment out 5A & 5B when running code in 5C.)
- Print out number of iterations until convergence
- **Answer: “Q2: How many iterations did your program need?”**
- **Answer: “Q3: What were the centroids? How best did the k-value fit the data?”**
- **Answer: “Q4: What meaning can you extract from the clustering? If you could not, what does that tell us about the data?”**
- Submission:

- Python code that displays final clustered graph with centroids
- Readme file with Answers

Caliente

- Create a 2-dimensional clustering algorithm based on the 1D code along.
- Run the clustering algorithm on the iris data (in scikitlearn, provided code and k).
- Optional: Create an algorithm to find the optimal k value by graphing the sum of squared differences.
- **Answer: “Q1: How many iterations did your program need?”**
- **Answer: “Q2: What were the centroids? How best did the k-value fit the data?”**
- **Answer: “Q3: What meaning can you extract from the clustering? If you could not, what does that tell us about the data?”**
- Submission:
 - Python code that displays final clustered graph with centroids
 - Readme file with Answers