**Homework 1:**

* Using the OFFLINE data and two different clustering methods predict the location of the ONLINE data set.
* Describe how you prepared the data.
* Describe how you estimated your error and found the best fit ASSUMING you CANNOT USE THE ONLINE DATA.
  + If you cannot use Online data, then use KNN output and metrics;

Look at the metrics related to Clustering: Silhouette Score – used to evaluate clustering models; Silhouette Score – Per scikit learn documentation, “The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b). To clarify, b is the distance between a sample and the nearest cluster that the sample is not a part of. Note that Silhouette Coefficient is only defined if number of labels is 2 <= n\_labels <= n\_samples - 1.

* + Used KNN and different clustering method – Random Forest – or – DBSCAN – there are 4-5 measures to use to evaluate.
  + DBSCAN – BIC metric,
  + KNN – recommended because it is better at physical distance. Uses proximity
  + Gaussian mixture model – could be an option?
  + ROOT MEAN SQUARE ERROR –
  + \*\*READ THE KNN METRICS around 600ish-700ish\*\*
  + Could you use the Offline set only and cross validate, create data frame with x, y; use KNN to predict the x, y and compare; that would give you an accuracy score.

<https://www.rdocumentation.org/packages/cluster/versions/2.1.0/topics/silhouette>

* Describe the best fit for the data.
* There are 2 macIDs located at the same position. Does one give better performance than the other?
* What about using them both?
* What is the drawback (if any of using this method to real-time locate an object)?
* Describe a method that may be an improvement based on your perceived drawbacks.

Kevin got to 374 with comments. NW to keep commenting out.

Datasets

<http://rdatasciencecases.org/Data/offline.final.trace.txt>

http://rdatasciencecases.org/Data/online.final.trace.txt