Update 7: Reflection

Dataset: Heartbeat sounds from <https://www.kaggle.com/kinguistics/heartbeat-sounds>

Goal: To classify unlabeled sounds into murmurs, extrahls, artifacts, or normal heartbeat sounds

Original Project Idea

* Classify heartbeat sounds into one of the four categories: murmurs, extrahls artifacts, or normal

What I have accomplished

* Exploratory data analysis on characteristics of the data set (means, min, max, median, mode, etc).
* Chi-squared tests on dependency of the characteristics

What I’ve tried that doesn’t work

* Principal component analysis – variances were way too high for 2-3 components (magnitude on the order of 10^8), because there were close to 400,000 elements for each data point.

What is left to do

* Use the following classification methods to train and classify the data and compare accuracies using confusion matrix:
  + K means clustering
  + K nearest neighbors
  + Linear Discriminant Analysis (LDA)
  + Classification and Regression Trees (CART)
  + Random Forest
  + Support Vector Machines (SVM)
  + Bayes classifier (bnclassify, naivebayes)
  + Neural Network