

## Math 5364 Homework 19

1. Test whether the  $k$ -means clusters for the `wdbc` data set are statistically independent of Diagnosis using a chi-square test with a simulated  $p$ -value.
2. In this problem, you will apply the DBSCAN clustering method to the `wdbc` data.
  - (a) Create a plot of sorted  $k$ -dist values, where  $k = 5$ , and determine the optimal value of Eps.
  - (b) Perform DBSCAN using this value of Eps and  $k = 5$ .
  - (c) How many clusters were identified?
  - (d) What percentage of the points in the data were classified as noise?
  - (e) Test whether the DBSCAN clusters are statistically independent of Diagnosis using a chi-square test with a simulated  $p$ -value.