Homework 5: kMeans

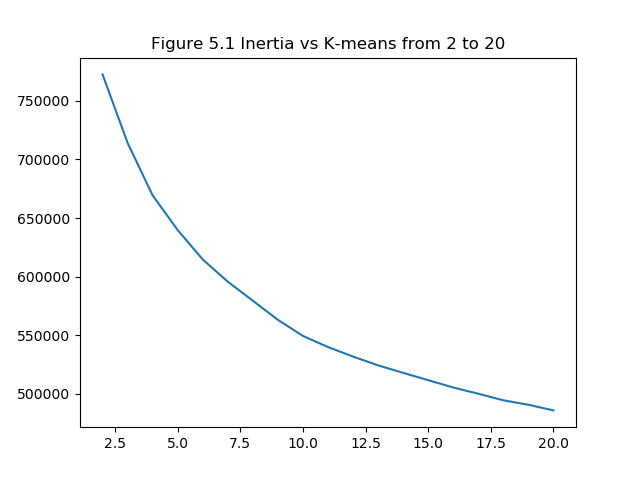
CS412

Released: April 23rd

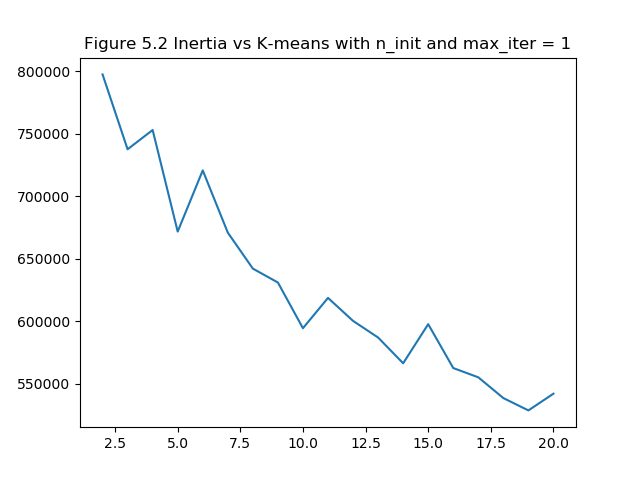
Due: April 30th, 11:30pm on Gradescope

1 kMeans

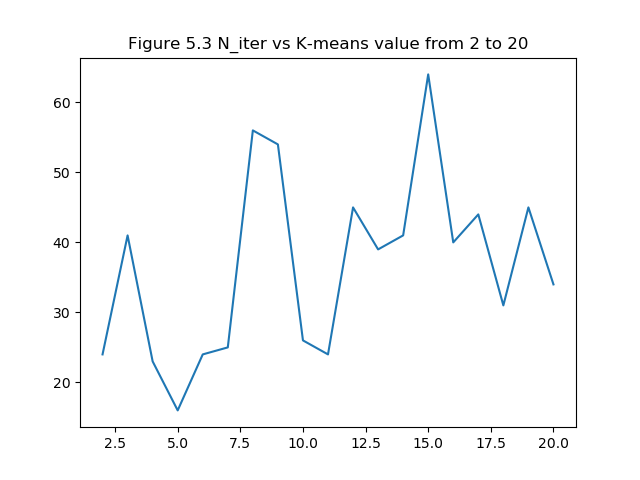
1. Figure 5.1



1. According to the experimental data n\_clusters = 7.5 is best application for data where the graph forms an elbow since n\_clusters = 20 is going to be overfit.



1. The main difference between the two graph is that when using default values for n\_init and max\_iter the value polynomial decreases smoothly whereas when changing the default values to max\_iter and n\_init =1 the figure no more decreases constantly. Since the K-mean algorithm is run just once which is not enough to decide the best k Mean model and hence the second figure does not decrease as smoothly
2. Figure 5.3



1. I was expecting number of iterations to be less somewhere around the value of n\_clusters = 5 and now after looking at the figure 5.3 I could say that it is close to what I expected, since in part (b) I found best application of data to be for the value of n\_clusters = 7.5. Now if we carefully observe the figure 5.3 the lowest value of n\_iter is between n\_clusters = 5 and 7.5.