

# Homework 5: kMeans

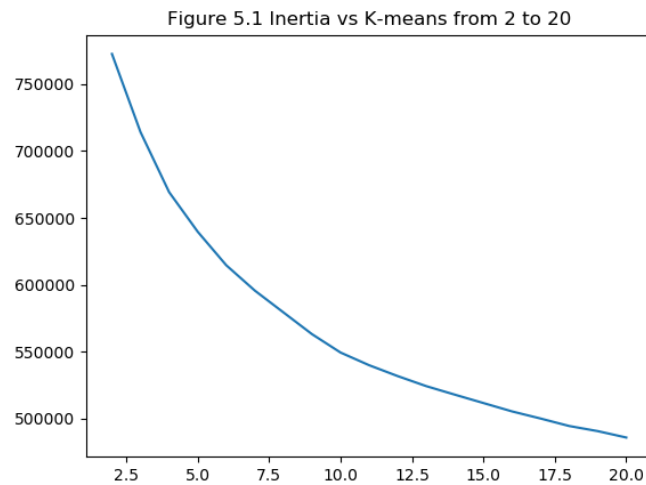
CS412

Released: April 23rd

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

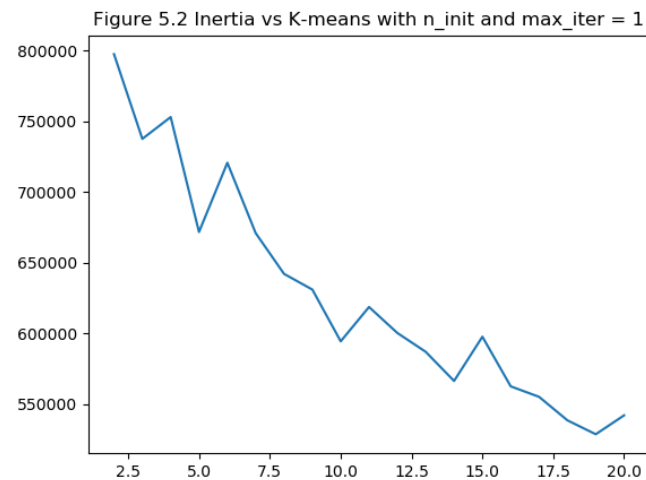
## 1 kMeans

(a) Figure 5.1



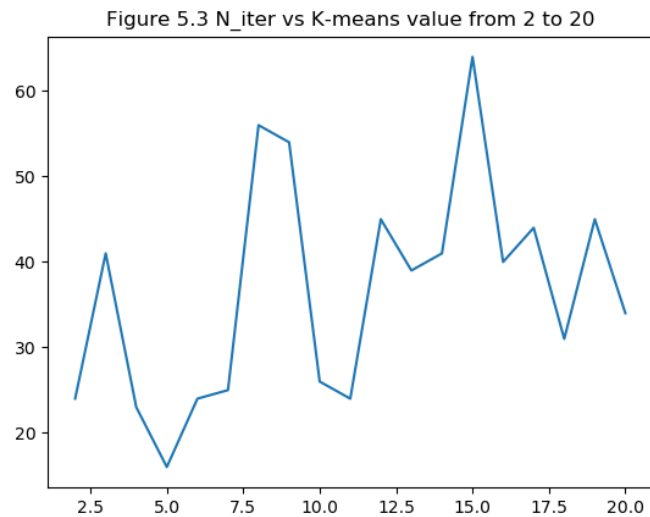
(b) According to the experimental data  $n\_clusters = 7.5$  is best application for data where the graph forms an elbow since  $n\_clusters = 20$  is going to be overfit.

(c)



(d) 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

(e) Figure 5.3



(f) 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`.