Data-Intensive Programming: Project

Jesper Granat, Peter Todorov, Jiaqi Guo

October 19, 2018

1 Special notes

The name of the Spark App is "Assignment" as described, but the name of the source code file is "spark_app.scala" and the project will be compiled as "spark_app".

2 Division of work

Each group member did an equal amount of work. Mostly the project was done at the same time by all members of the group and the rest was divided equally.

3 Completed tasks

Each task was completed in this project.

4 Elbow search

4.1 Task 4

As seen in figure 1, there is a clear change in the slope of the curve at around $k = 60$. The data was acquired via running the training on several values of k and plotting the sum of the errors.

Figure 2 shows that the elbow point is at the same point as in the 2D case. This is because the weekday does not seem to have much of an effect on the clusters. The data was acquired similarly as in part 1.

4.2 Task 6

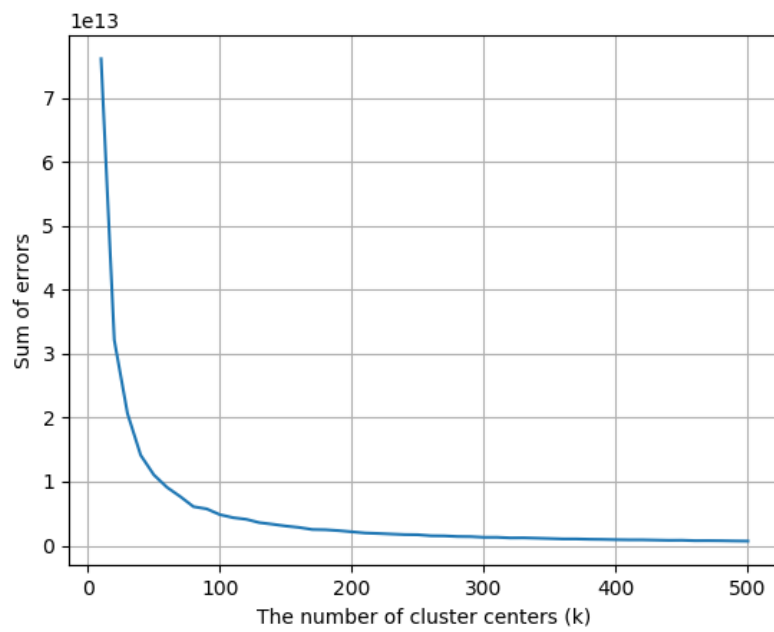Calculations show the elbow at $k = 70$, which seems consistent with figure 3.
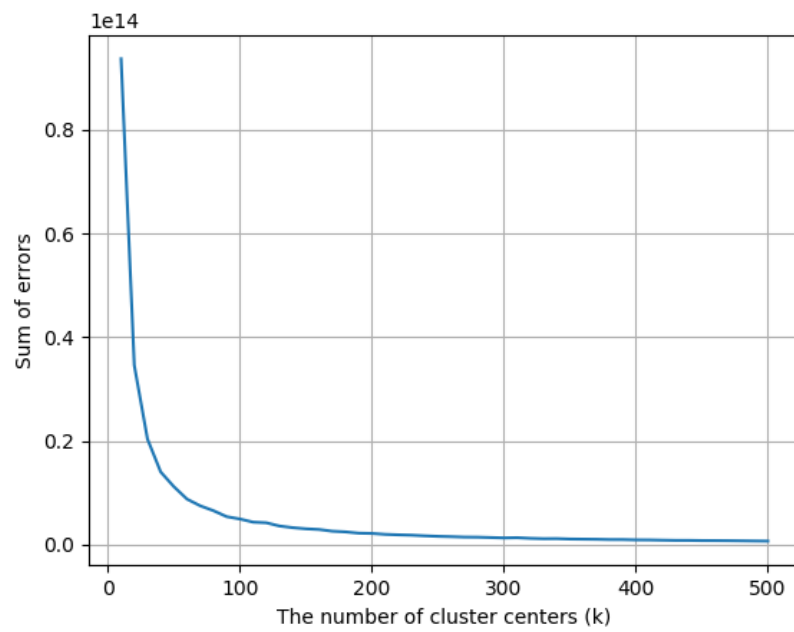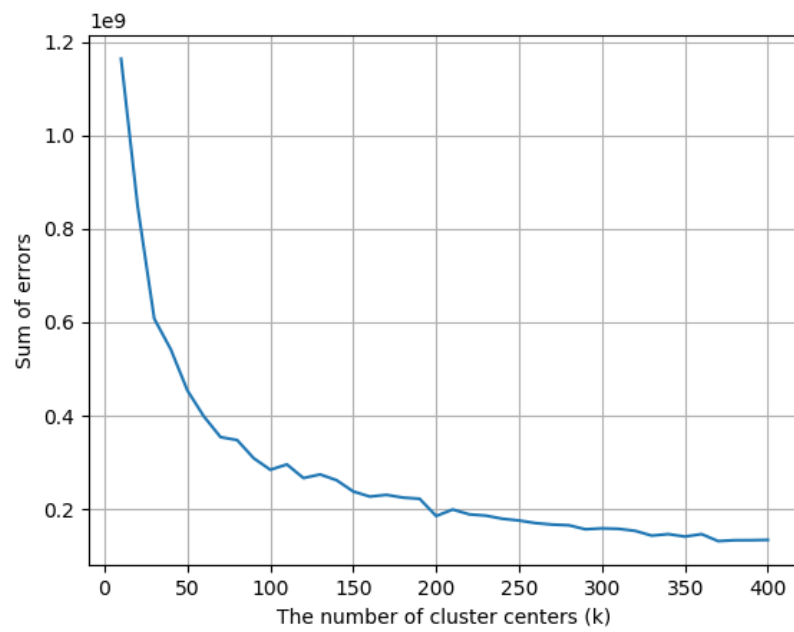
Figure 1: Elbow point of the 2D case

Figure 2: Elbow point of the 3D case

Figure 3: Elbow point of task 6