Unsupervised Learning (K-means)

Yu Chang

Arizona State University

ASU ID: 1231962201

ychan175@asu.edu

1. INTRODUCTION

The goal of this project is to apply the k-means algorithm to the provided dataset, which consists of a collection of two-dimensional points. There are two methods for selecting the first cluster centers.

First strategy: choose the initial centers at random from the provided samples.

Second strategy: choose the initial centers at random; for the i-th center (i>1), select a sample (from all available samples) so that its average distance to all of the preceding (i-1) centers is maximum.

2. Strategy 1

1. Choose the initial centers at random from the provided samples by changing the class parameters in scikit-learn to "random".

```
kmeansRandom = KMeans(
    # initialization technique
    init="random",
```

2. Determine the objective function as a function of k (where k ranges from 2 to 10).
3. Proceed in the same manner with an additional initialization.

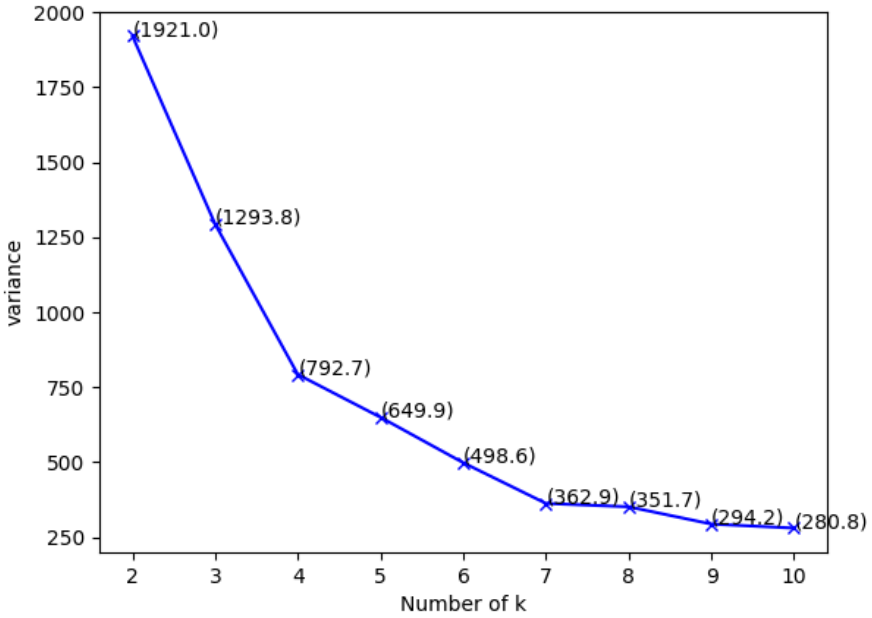
3. Strategy 2

1. Similar to Strategy 1. The only difference is Changing the class parameters in scikit-learn to "k-means++".

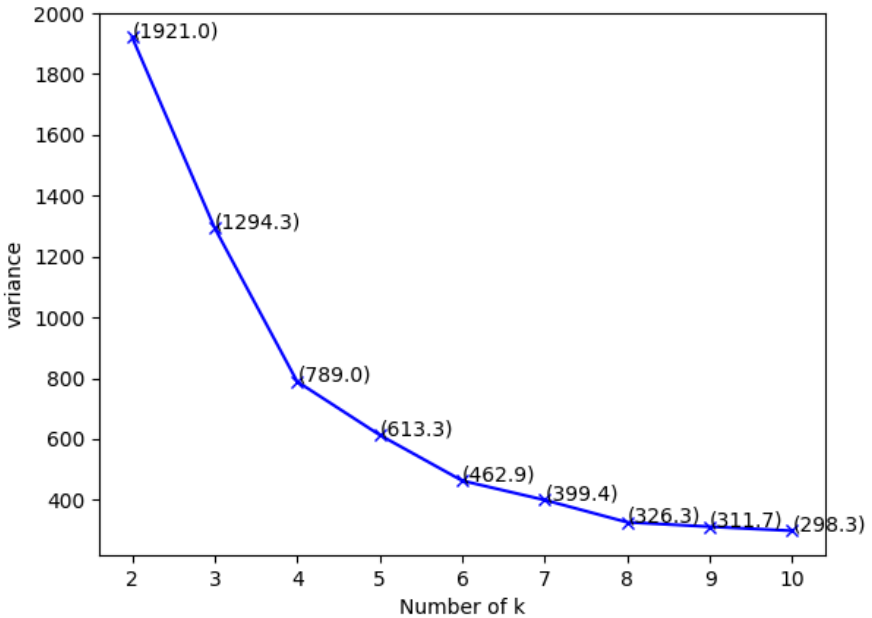
```
kmeansPP = KMeans(
    # initialization technique
    init="k-means++",
```

4. RESULTS

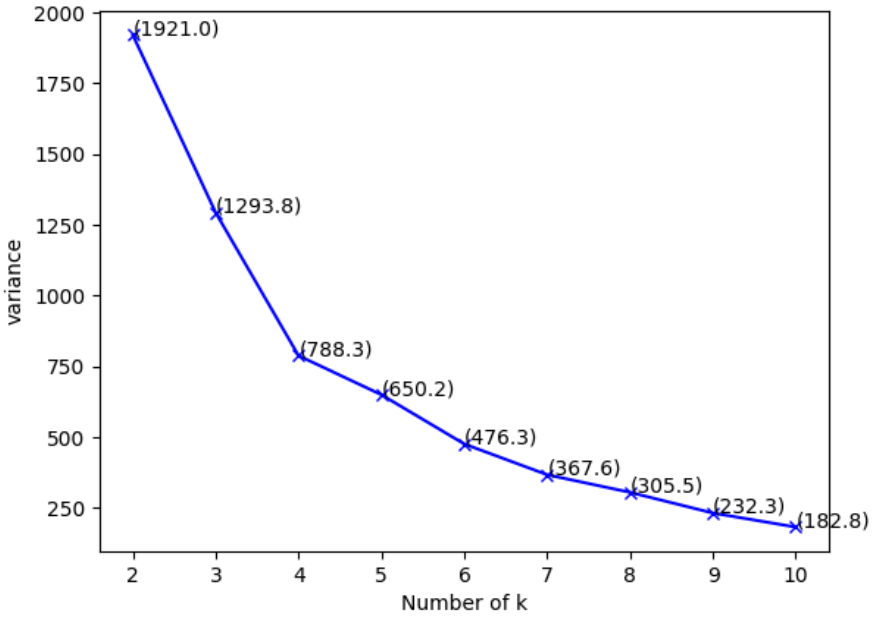
1. Strategy 1 initialization 1:



2. Strategy 1 initialization 2:



3. Strategy 2 initialization 1:



4. Strategy 2 initialization 2: