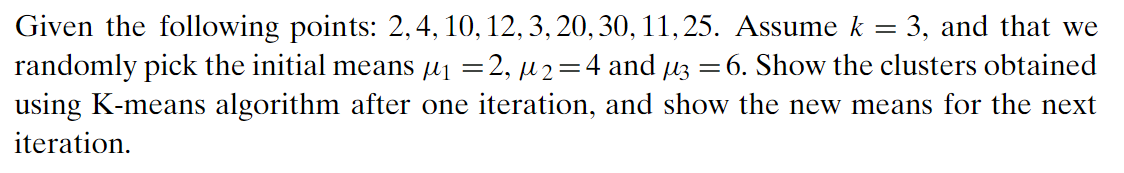
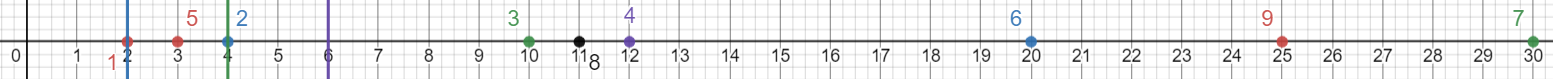
**Data Mining (CSE542)**

**Homework 04**

**ID: \_\_ Name: \_\_조원석\_ Date: \_\_2023-05-08\_\_**

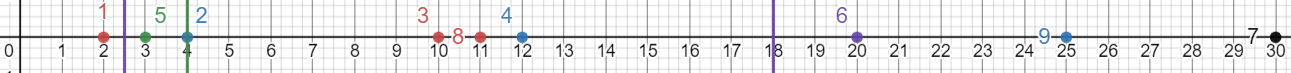
**Task-1**

****

****

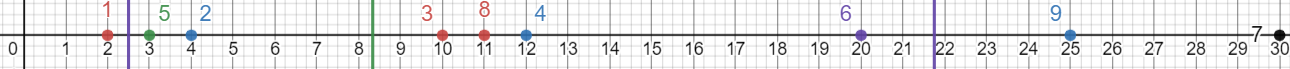
**Initial means =>**

**Until assume k=3, I do that.**

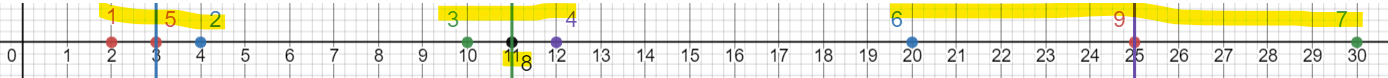
****

**Closest mean\_1 : C1 = {2, 3} C2 = {4}, C3{10, 11, 12, 20, 25, 30}**

**New means\_1 => 18**

****

**2nd iteration : C1 = {2, 3} C2 = {4, 10, 11}, C3{ 12, 20, 25, 30 }**

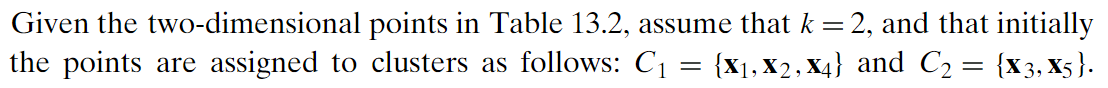
**New means\_2 =>   
**

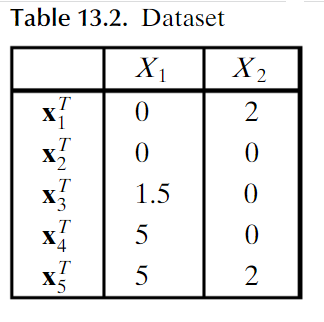
**3rd iteration : C1 = {2, 3, 4} C2 = {10, 11, 12}, C3{ 20, 25, 30 }**

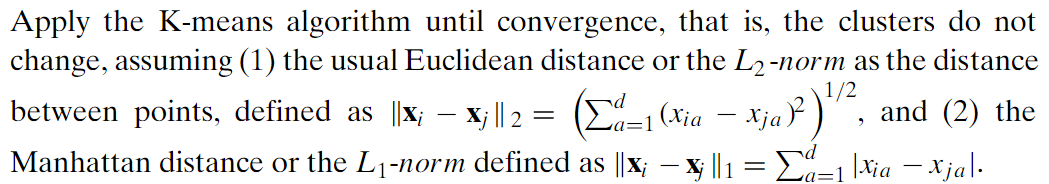
**New means\_3 =>**

**So answer is**

**Task-2**

****

****

****

* **Eucildean distance**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **,** | **,** | **Cluster** |
|  | **2.1** | **3.4** |  |
|  | **1.8** | **3.4** |  |
|  | **0.7** | **2.0** |  |
|  | **3.4** | **2.0** |  |
|  | **3.6** | **2.0** |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **,** | **,** | **Cluster** |
|  | **1.42** | **5.1** |  |
|  | **0.83** | **5.1** |  |
|  | **1.2** | **3.6** |  |
|  | **4.5** | **1.0** |  |
|  | **4.7** | **1.0** |  |

* **Manhattan distance**

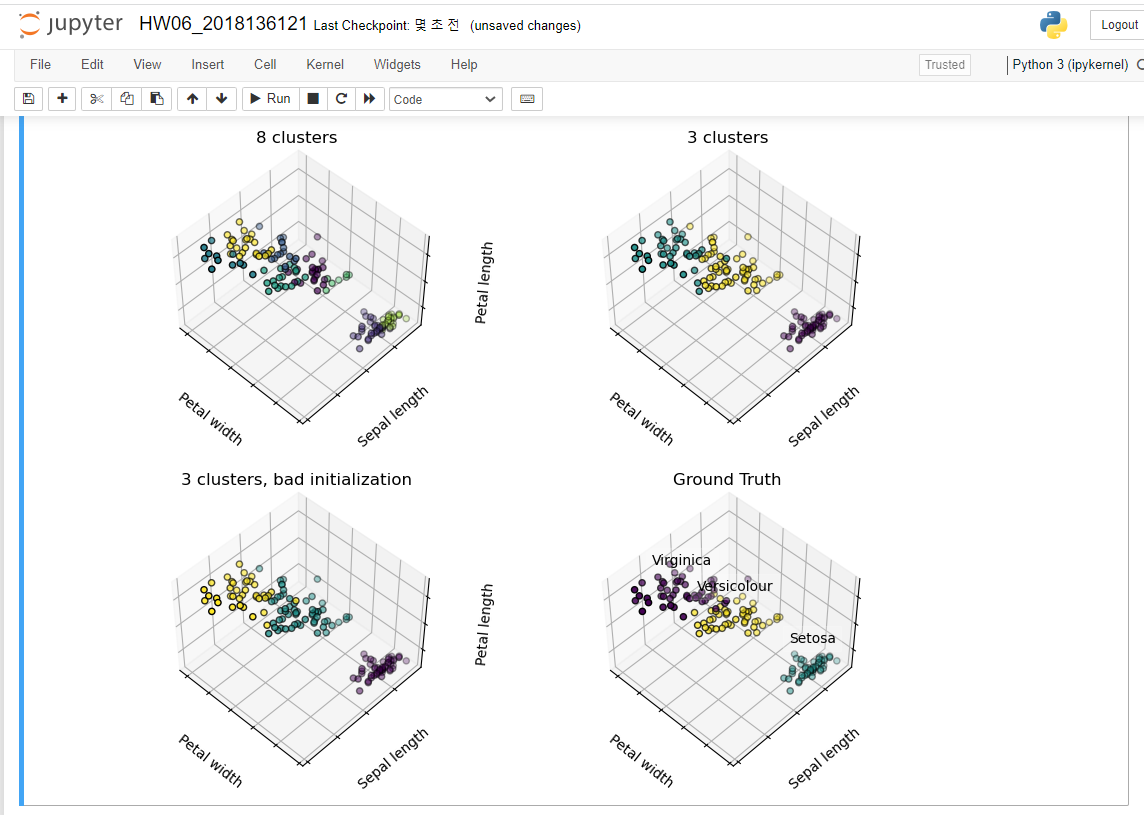
|  |  |  |  |
| --- | --- | --- | --- |
|  | **,** | **,** | **Cluster** |
|  | **3** | **4.25** |  |
|  | **2.34** | **4.25** |  |
|  | **0.84** | **2.75** |  |
|  | **4** | **2.75** |  |
|  | **4.66** | **2.75** |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **,** | **,** | **Cluster** |
|  | **1.83** | **6** |  |
|  | **1.17** | **6** |  |
|  | **1.67** | **4.5** |  |
|  | **5.17** | **1.0** |  |
|  | **5.83** | **1.0** |  |

**Task-3.** Run the code and record the results for the following examples from the site

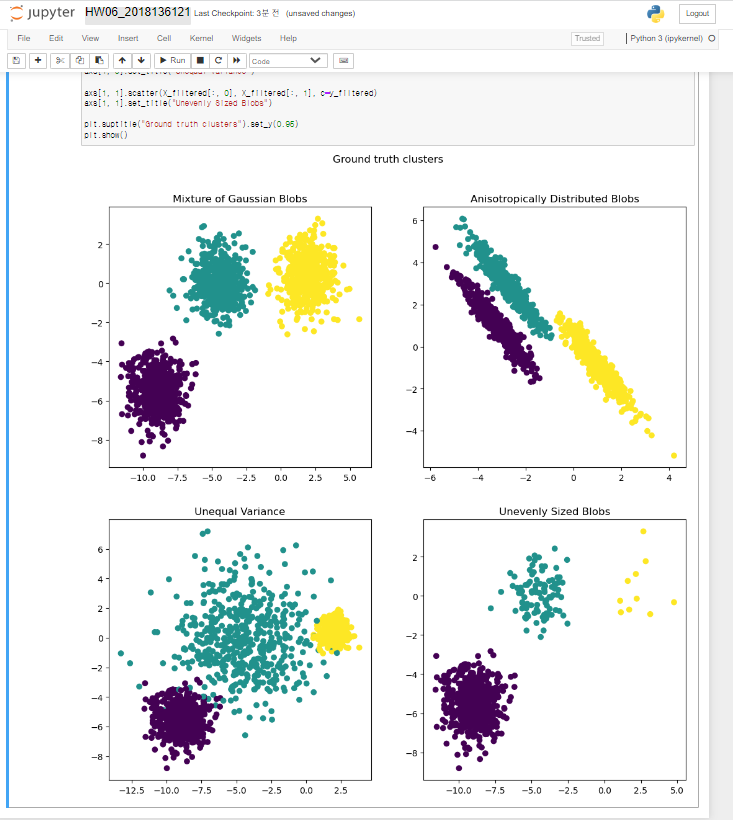
<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

1. K-means Clustering



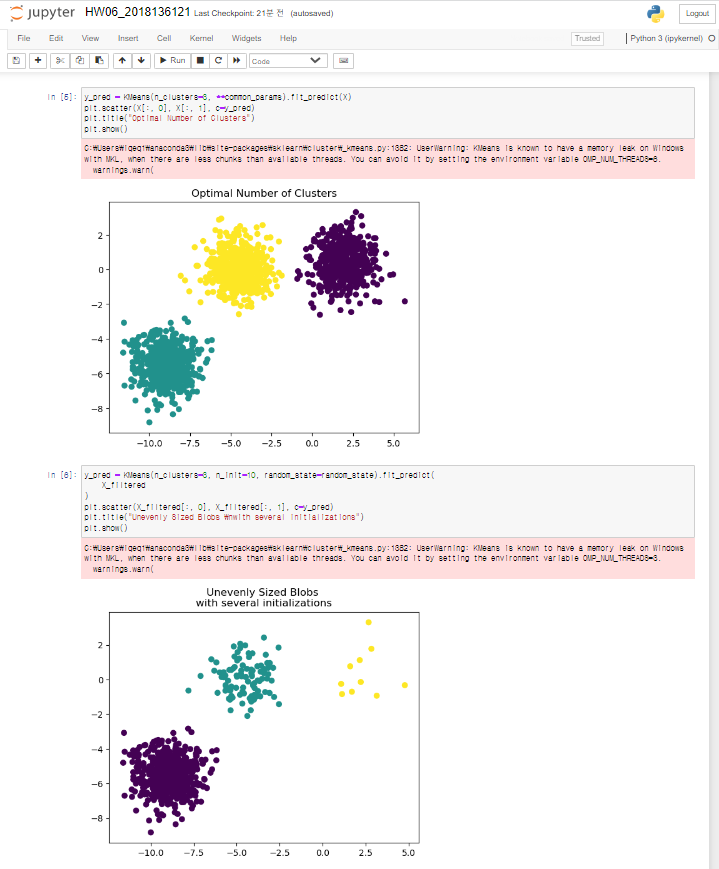
1. Demonstration of k-means assumptions

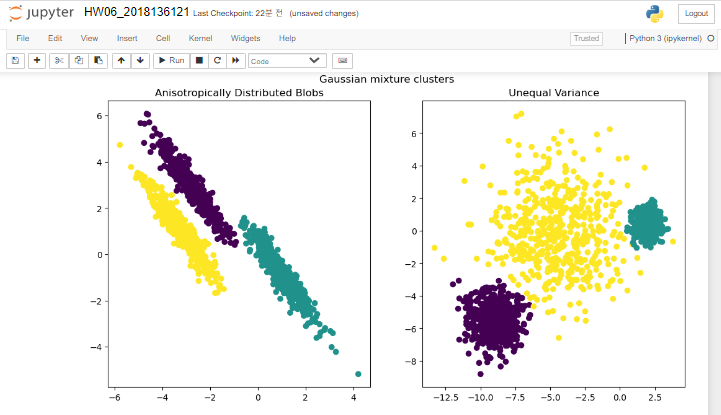
* Data generation



* Fit model and flot results



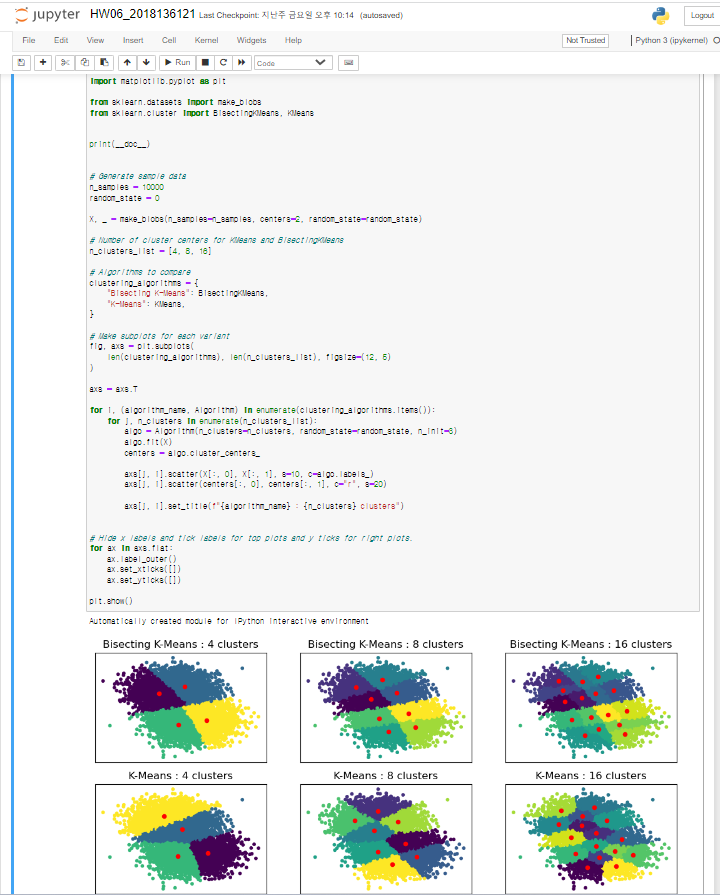




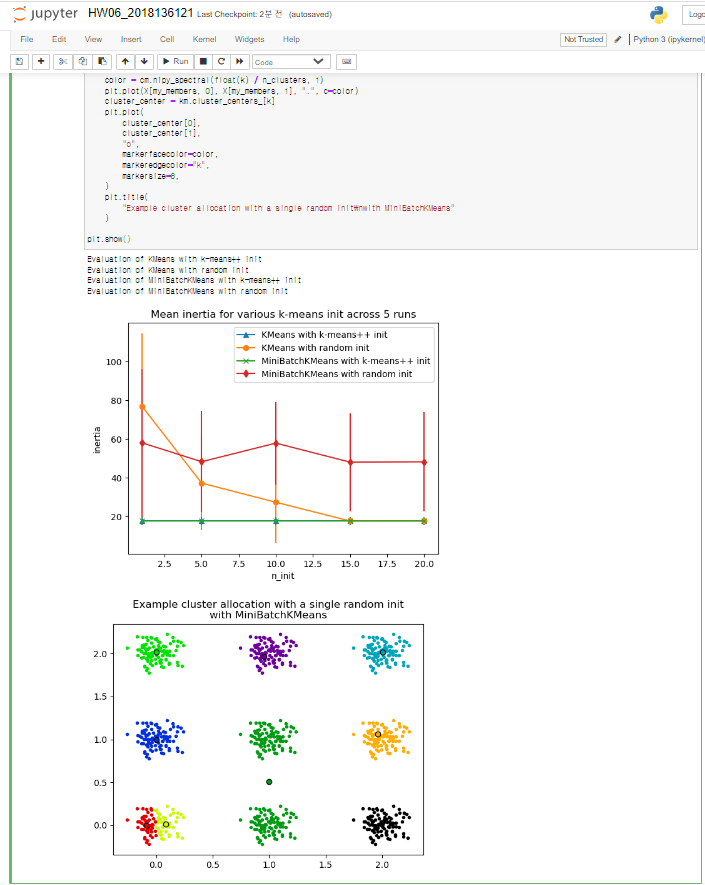
1. Comparison of the K-Means and MiniBatchKMeans clustering algorithms



1. Bisecting K-Means and Regular K-Means Performance Comparison



1. Empirical evaluation of the impact of k-means initialization



1. Clustering text documents using k-means

