**Student’s Name:** Sachin Mahawar

**Roll Number:** B20129

**Mobile No:** 916643951

**Branch:** CSE

Chart, line chart

Description automatically generated

Figure 1 Eigenvalue vs. components

**Inferences:**

1. The eigenvalue decrease corresponding to each component increasing successively.
2. Because attributes are more dependent on first eigen vector so it has more spread around 1st eigen vector.

# a.

Chart, scatter chart

Description automatically generated

Figure 2 K-means (K=3) clustering on Iris flower dataset

# Inferences:

1. The clustering looks quite accurate forming good clusters as K-Means is an unsupervised algorithm.
2. K-means algorithm assumes cluster boundaries to be circular in 2D. From the output, the boundary doesn’t seem to be clearly circular.

**b.** The value for distortion measure is 63.8738

**c.** The purity score after examples are assigned to the clusters is 0.887

Chart, line chart

Description automatically generated

Figure 3 Number of clusters(K) vs. distortion measure

**Inferences:**

1. The distortion measure decreases with an increase in K.
2. Distortion measure decreases drastically for k=2 to k=3 then very gradually because our data has 3 species which also indicate optimal K value.
3. Number of species in data set is 3, So intuitively k=3 should be the number of optimum clusters. The elbow and distortion measure plot closely follow the intuition.

Table 1 Purity score for K value = 2,3,4,5,6 & 7

|  |  |
| --- | --- |
| **K value** | **Purity score** |
| 2 | 0.667 |
| 3 | 0.887 |
| 4 | 0.687 |
| 5 | 0.673 |
| 6 | 0.513 |
| 7 | 0.513 |

**Inferences**:

1. The highest purity score is obtained with K = 3.
2. Purity score increases from k=2 to k=3 and then decreases for further K values.
3. Because the number of species in our data is 3 so purity score is highest for k=3.
4. Except k=3, distortion measures decreases with increase in purity score.

# a.

Chart, scatter chart

Description automatically generated

Figure 4 GMM (K=3) clustering on Iris flower dataset

# Inferences:

1. Predicted results are very close to actual ones that’s why GMM looks quite accurate.
2. GMM algorithm assumes cluster boundaries to be elliptical in 2D. From the output, the boundary seems little bit elliptical.
3. Yes, from the graphs we can see that the boundaries in K-means were circular while in GMM they are elliptical.

**b.** The value for distortion measure is -280.87

**c.** The purity score after examples are assigned to the clusters is 0.98

Chart, line chart

Description automatically generated

Figure 5 Number of clusters(K) vs. distortion measure

**Inferences:**

1. The distortion measure increase with an increase in K.
2. Because there are three species, the distortion measure has a steeper slope between k=2 and k=3, then progressively grows until k=6, then abruptly increases beyond that.
3. From the number of species in the given dataset, intuitively k=3 should be the optimal value of k.

The elbow method also follow the intuition.

Table 2 Purity score for K value = 2,3,4,5,6 & 7

|  |  |
| --- | --- |
| **K value** | **Purity score** |
| 2 | 0.667 |
| 3 | 0.98 |
| 4 | 0.833 |
| 5 | 0.767 |
| 6 | 0.64 |
| 7 | 0.627 |

**Inferences**:

1. The highest purity score is obtained with K = 3.
2. Purity score increases for k=2 to k=3 then decreases for k=2 to k=7.
3. Because the number of species in our data is 3 so purity score is highest for k=3.
4. Except k=3, distortion measure decreases with increase in purity score.
5. By inferences, we can say that GMM is more accurate than K-means.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Figure 6 DBSCAN clustering on Iris flower dataset

**Inferences:**

1. Due to our eps values, accuracy is not that good.
2. The number of clusters are less than both K-means and GMM and boundaries are neither circular nor elliptical in DBSCAN.

**b.**

|  |  |  |
| --- | --- | --- |
| **Eps** | **Min\_samples** | **Purity Score** |
| 1 | 4 | 0.667 |
| 10 | 0.667 |
| 5 | 4 | 0.333 |
| 10 | 0.333 |

**Inferences:** **0.333**

1. For the same eps value, increasing min\_samples doesn’t change purity score.
2. For the same min\_samples, increasing eps value decrease purity score.