

CSE343: Machine Learning

Assignment-4

Shubham Sharma (2021099)

November 28, 2024

1 Section A (Theoretical)

Solution (a)

(a) Dimensions of the Resulting Feature Map

For a convolutional layer with stride $S = 1$ and no padding:

$$\text{Output height} = M - K + 1$$

$$\text{Output width} = N - K + 1$$

Thus, the dimensions of the resulting feature map are:

$$(M - K + 1) \times (N - K + 1)$$

(b) Number of Elementary Operations per Output Pixel

For a single output pixel, the kernel of size $K \times K$ is applied to the input channels.

- **Multiplications:** For P channels, there are $P \cdot K^2$ multiplications.
- **Additions:** Summing the results of the $P \cdot K^2$ multiplications involves $P \cdot K^2 - 1$ additions.

$$\text{Total operations} = P \cdot K^2 + (P \cdot K^2 - 1) = 2 \cdot P \cdot K^2 - 1 \approx 2 \cdot P \cdot K^2$$

(c) Computational Time Complexity

1. **Number of Output Pixels:** For a feature map of size $(M - K + 1) \times (N - K + 1)$, the total number of output pixels is:

$$(M - K + 1) \cdot (N - K + 1)$$

2. **Operations for One Kernel:** For each kernel, computing the entire feature map requires:

$$(M - K + 1) \cdot (N - K + 1) \cdot (2 \cdot P \cdot K^2)$$

3. **Operations for Q Kernels:** With Q kernels, the total computational cost is:

$$Q \cdot (M - K + 1) \cdot (N - K + 1) \cdot (2 \cdot P \cdot K^2)$$

The time complexity is therefore:

$$O(Q \cdot (M - K) \cdot (N - K) \cdot P \cdot K^2)$$

4. **For $\min(M, N) \gg K$:** When K is much smaller than $\min(M, N)$. The time complexity simplifies to:

$$O(Q \cdot M \cdot N \cdot P \cdot K^2)$$

Solution (b)

K-Means Algorithm: Assignment and Update Steps

K-Means is an iterative clustering algorithm that partitions a dataset into K clusters.

Assignment Step

Each data point is assigned to the cluster whose centroid is closest to it. The distance is computed using the Euclidean distance formula:

$$\text{distance} = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

where x is the data point, c is the centroid, and n is the dimension of the data.

Update Step

After assigning all points to clusters, the centroids of the clusters are recalculated as the mean of all the data points assigned to that cluster:

$$c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

where c_j is the updated centroid of cluster j , n_j is the number of points in cluster j , and x_i are the points in the cluster.

Both of the above steps are repeated until convergence, i.e., when the cluster assignments no longer change or the centroid movements fall below a threshold.

Determining the Optimal Number of Clusters

Elbow Method

This method involves plotting the *within-cluster sum of squares* (WCSS) for different values of K , where:

$$\text{WCSS} = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i - c_j\|^2$$

Here, $\|x_i - c_j\|^2$ is the squared distance between a point and its cluster centroid.

As K increases, WCSS decreases because more clusters mean that each cluster has fewer points and is more compact. The optimal K is chosen at the "elbow point" in the plot, where the rate of decrease in WCSS slows significantly.

Random Initialization of Centroids and Global Minima

Random Initialization

In K-Means, initial centroids can be chosen randomly. However, the quality of the final clusters depends on the initialization of centroids. Random initialization can lead to suboptimal solutions (local minima) because the algorithm may converge to a local minimum.

Global Minima

K-Means is not guaranteed to find the global minimum due to its sensitivity to initialization. To tackle this, there are other versions of KMeans, such as *K-Means++*, which select initial centroids in a more informed manner to increase the likelihood of convergence to a global minimum.

2 Section B (Scratch Implementation)

Part B: Final Centroids and Visualization of Clusters

a) Final Centroids

$$\text{Final Centroids} = \{(5.8, 2.125), (4.19, -0.05)\}$$

b) Clusters Visualization

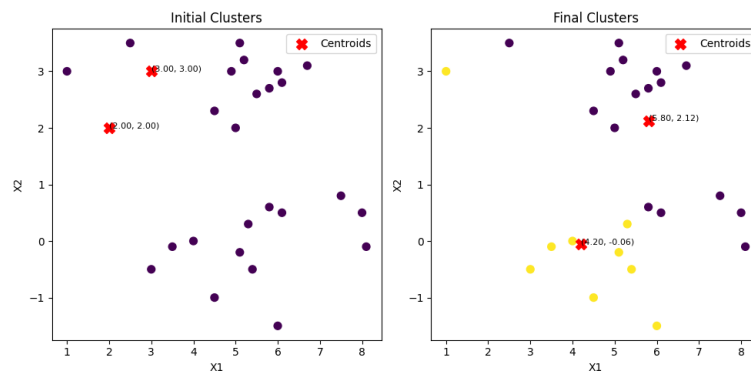


Figure 1: Comparison of Centroids Before and After the Algorithm Converges

Part C: Random Centroids and Visualization of Clusters

a) Final Centroids after Random Initialization of Centroids

$$\text{Final Centroids} = \{(4.85, 2.89), (5.56, -0.09)\}$$

b) Clusters Visualization

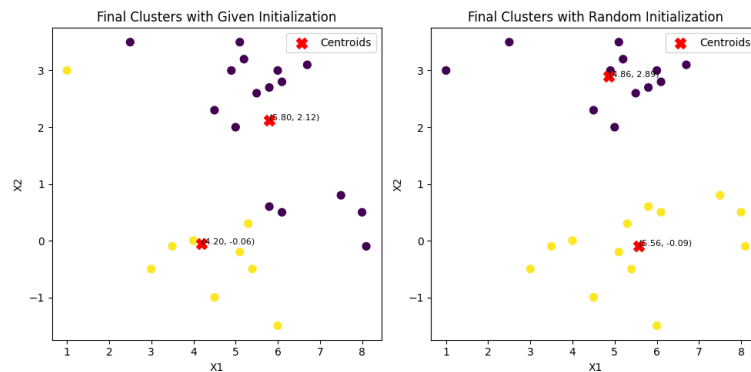


Figure 2: Comparison of Centroids After Given Initialization and After the Random Initialization

From the above graphs, I observed that the final cluster formed by randomly initialized centroids is much more compact and better than the initially given centroids.

Part D: Determining Optimal Number of Clusters using Elbow Method

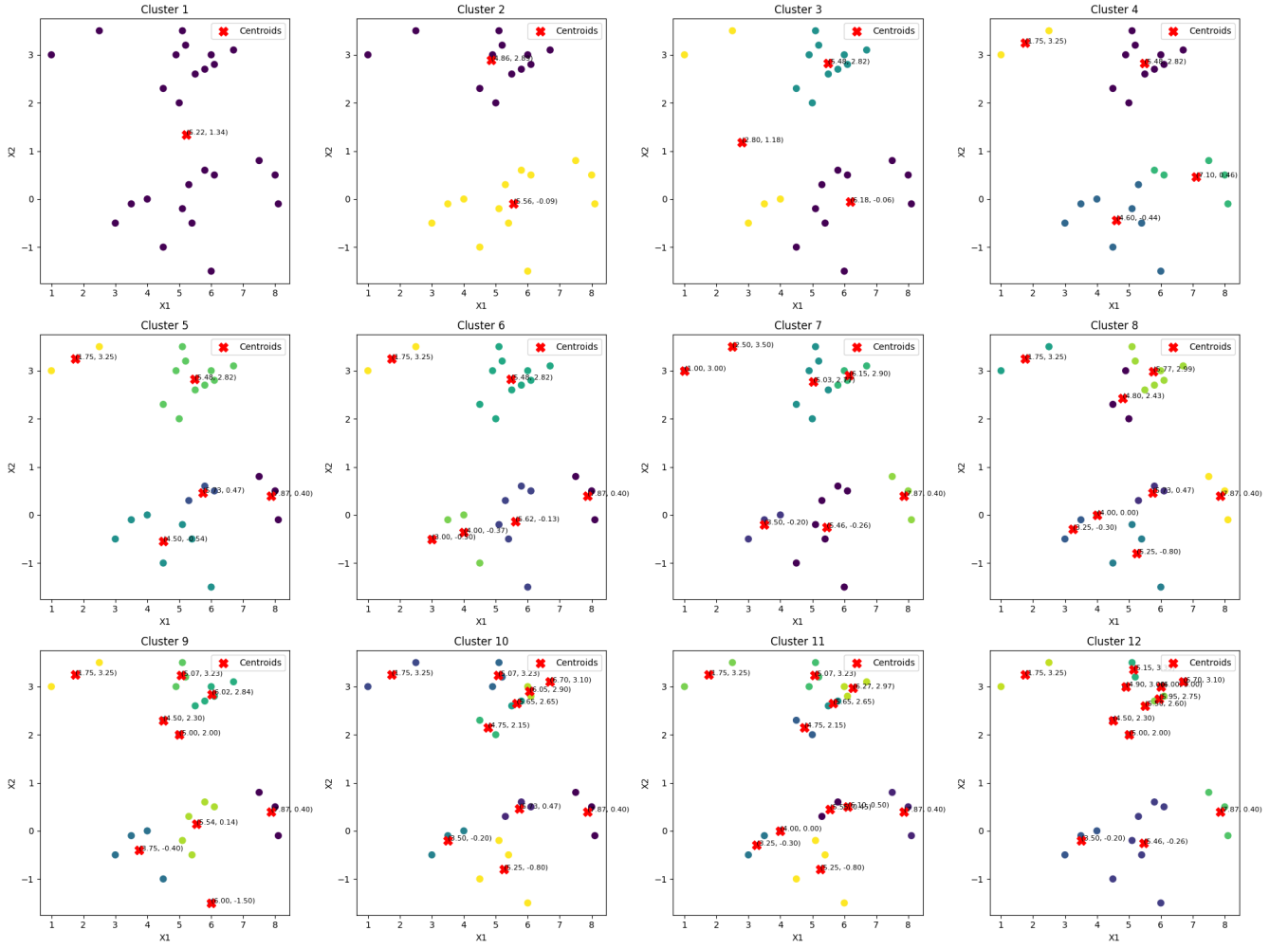


Figure 3: Final Centroids against different values of k

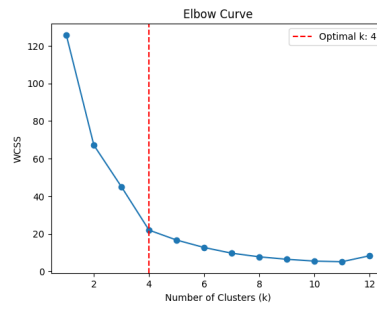


Figure 4: Within-Cluster Sum of Squares (WCSS) against different values of k

From the above graph, it is clearly visible that as the value of k is increasing, the WCSS score is decreasing, and the point where the rate of decrease in WCSS slows significantly is chosen as the Elbow point, and we get the optimal number of k value.

Optimal Value of K = 4