# Assignment #4

**Elements of Machine Learning**

**Saarland University – Winter Semester 2024/25**


**Rabin Adhikari**
7072310
raad00002@stud.uni-saarland.de

**Dhimitrios Duka**
7059153
dhdu00001@stud.uni-saarland.de


## 1 Problem 1 (K-means)

### 1.1

First, we start by calculating the distance of point $x_2$ and $x_4$ to the initial centroids $\bar{x}_1$ and $\bar{x}_2$.

$$\begin{aligned} d_{x_2\bar{x}_1} &= \sqrt{0^2 + 3^2} = 3 \\ d_{x_4\bar{x}_1} &= \sqrt{3^2 + 3^2} = 3\sqrt{2} \\ d_{x_2\bar{x}_3} &= \sqrt{3^2 + 3^2} = 3\sqrt{2} \\ d_{x_4\bar{x}_3} &= \sqrt{0^2 + 3^2} = 3 \end{aligned} \tag{1}$$

Therefore, we can conclude that the first cluster would contain the data points $x_1$ and $x_2$ and the second cluster would contain the remaining data points $x_3$ and $x_4$.

Now for the second iteration, first we have to calculate the new centroids.

$$\begin{aligned} \bar{x}_1 &= (1, \frac{5}{2}) \\ \bar{x}_2 &= (4, \frac{5}{2}) \end{aligned} \tag{2}$$

Now, we calculate the distances of every point to the new centroids.

$$\begin{aligned}
d_{x_1\bar{x}_1} &= \sqrt{0^2 + (\frac{5}{2} - 1)^2} = \frac{3}{2} \\
d_{x_2\bar{x}_1} &= \sqrt{0^2 + (4 - \frac{5}{2})^2} = \frac{3}{2} \\
d_{x_3\bar{x}_1} &= \sqrt{(4-1)^2 + (1 - \frac{5}{2})^2} = \frac{3}{2}\sqrt{5} \\
d_{x_4\bar{x}_1} &= \sqrt{(4-1)^2 + (4 - \frac{5}{2})^2} = \frac{3}{2}\sqrt{5} \\
d_{x_1\bar{x}_2} &= \sqrt{(4-1)^2 + (\frac{5}{2} - 1)^2} = \frac{3}{2}\sqrt{5} \\
d_{x_2\bar{x}_2} &= \sqrt{(4-1)^2 + (\frac{5}{2} - 4)^2} = \frac{3}{2}\sqrt{5} \\
d_{x_3\bar{x}_2} &= \sqrt{0^2 + (\frac{5}{2} - 1)^2} = \frac{3}{2} \\
d_{x_4\bar{x}_2} &= \sqrt{0^2 + (\frac{5}{2} - 4)^2} = \frac{3}{2}
\end{aligned} \tag{3}$$

From the above calculations, we can see that there is no reassignment of any of the datapoints to a new cluster. Therefore, the algorithm has converged.

## 1.2

### 1.2.1

Based on the provided graph and the intuition behind the elbow huristic, we would choose a value of $k = 3$. The reason behind this choice is that for values smaller than 3, there is a large decrease in WCSS. However, for values larger than 3, there is a much slower decrease of the WCSS, suggesting diminishing returns.

### 1.2.2

As the number of clusters $k$ icreases, each cluster becomes smaller and more specific, thus containing fewer data samples. As a result, the data samples within a cluster are closer together, thus reducing the within-cluster variation. However, based on the above exercise, we can see that the within-cluster variation follows a elbow curve. This means that after a certain point, the improvement of the within-cluster variabtion beacomes smaller beacomes clusters start splitting data points that are already well-grouped.

### 1.2.3

If we suppose that $k = N$, where $N$ is the number of data samples, the within-cluster variation would be 0 since each cluster would contain only one sample and the distance of that sample from itself, which is the clusters center, would be 0.

### 1.2.4

From the given plots, we can conlude that Plot 2 is the plot that corresponds to the $k$-medoids clustering algorithm. This is due to that fact that in $k$-medoids, the center of the cluster is one of the data samples itself, while in the $k$-means clustering algorithm, the center of the cluster is not necessearly a sample point of the cluster.