Machine Learning
Spring 2024

Universität Hamburg
Sören Laue

Exercise Sheet 9

Exercise 1

1. Given the following dataset of 8 data points:

   $A = \{(1,1),(2,1),(4,3),(5,4),(6,5),(7,6),(9,8),(10,7)\}$

   Apply the k-means algorithm with $k = 2$, starting with the initial centroids $C_1 = (2,1)$ and $C_2 = (10,7)$. Perform two iterations of the algorithm, showing the assignment of data points to clusters and the updated positions of centroids after each iteration.

2. Now, suppose you want to apply the k-means algorithm with $k = 3$ to the same dataset. Choose the initial centroids as $C_1 = (1,1)$, $C_2 = (5,4)$, and $C_3 = (9,8)$. Perform two iterations of the algorithm, showing the assignment of data points to clusters and the updated positions of centroids after each iteration.

Exercise 2

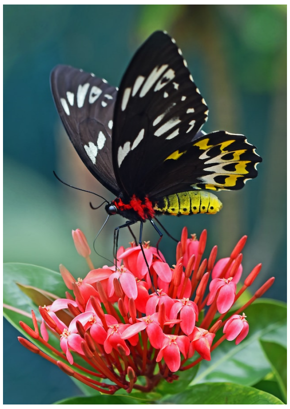
For the k-means algorithm, a cluster is given as C_j. Provide a mathematical proof that the cluster center μ_j for the cluster C_j minimizes the following equation

$$\min_{\mu_j} \sum_{x^{(i)} \in C_j} \left\| x^{(i)} - \mu_j \right\|^2$$

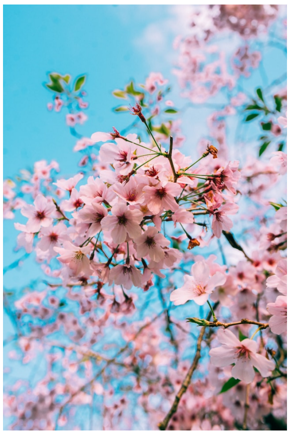
if and only if μ_j is the centroid of the cluster, i.e.,

$$\mu_j = \frac{1}{|C_j|} \sum_{x^{(i)} \in C_j} x^{(i)}.$$

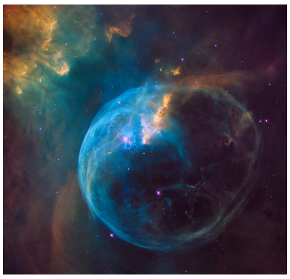
Exercise 3



(a) Photo by David Clode on Unsplash



(b) Photo by Y S on Unsplash



(c) Photo by NASA on Unsplash

You are given three RGB images. Compress these images using k-means. To do that run k-means on the data set where each data point resembles a pixel in 3-dim space (RGB) and replace each pixel with its nearest cluster center. Which number of clusters would you say is ok for compressing each image?

Exercise 4

Using PCA, compress the images from Exercise 2. For this, treat each $n \times m$ image as three different data matrices, where each color channel (Red, Green, and Blue) gives rise to exactly one data matrix $X \in \mathbb{R}^{n \times m}$. Plot the inverse transform of each image. Which number of principal components is enough to almost perfectly compress each image? Which number of principal components would you say is enough for a human to see what is on the image?

Please turn in your solutions by Thursday, June 20th.