

8. Problem sheet for Statistical Data Analysis

Exercise 1 (6 points)

Given then following data points

$$x^{(1)} = (2, 8), \quad x^{(2)} = (2, 5), \quad x^{(3)} = (1, 2), \quad x^{(4)} = (5, 8) \quad (1)$$

$$x^{(5)} = (7, 3), \quad x^{(6)} = (6, 4), \quad x^{(7)} = (8, 4), \quad x^{(8)} = (4, 7) \quad (2)$$

and compute 2 iterations of the K-Means algorithm by hand using the Forgy's initialisation choosing $x^{(3)}$, $x^{(4)}$ and $x^{(6)}$. Calculate the loss function in each iteration.

Exercise 2 (10 Points)

1. Load the jpg image given in Moodle and get its RGB representation (then each pixel is given as an individual 3-dimensional data point).
2. Implement the k-means algorithm. Using k-means cluster all the pixels of an image into k clusters and assign each pixel the color represented by its nearest cluster center.
3. Assign each pixel the color value of its nearest cluster center.
4. Visualize the result.