Deadline: Friday January 29th, 10:00 Upload a `.pdf` file via LearnWeb. (e.g. exported Jupyter notebook)

1. K Means Clustering (8 points)

x_1	x_2
0	0
-1	0
-1	-1
0	-1
1	1
1	2

Table 1

A. [2p] Compute the (squared) **distance matrix** $D_{ij} = \text{dist}_{\text{eucl.}}(x_i, x_j)^2$, given the data from Table 1.

B. [2p] Name two possible ways to choose the updated mean in the K-means clustering.

C. [4p] Perform K-means clustering on the dataset from Table 1. Use the first and last datapoints as initial centers ($K = 2$). Given the final parameters, which cluster would $x^* = \binom{1}{1}$ belong to?

2. Hierarchical Clustering (6 points)

A. [2p] Compute the **distance matrix** $D_{ij} = \text{dist}(x_i, x_j)$, using the **Manhatten distance** (i.e. L^1), given the data from Table 1.

B. [4p] Perform **agglomerative Hierarchical Clustering** using **single linkage** as the cluster distance measure. Draw the associated tree.

3. Gaussian Mixture Models (6 points)

A Gaussian mixture model containing $K = 3$ components has been learned for some one-dimensional training data. The individual Gaussians are given by

$$\mu_1 = -1 \qquad \mu_2 = 1 \qquad \mu_3 = 4$$
$$\sigma_1 = 1 \qquad \sigma_2 = 2 \qquad \sigma_3 = 0.4$$

Additionally, the probabilities for the individual clusters are:

$$\pi_1 = 0.4 \qquad \pi_2 = 0.4 \qquad \pi_3 = 0.2$$

A. [4p] Compute the responsibilities for a point $x \in \mathbb{R}$ to belong to a cluster i as:

$$r_i(x) = \frac{\pi_i \mathcal{N}(\mu_i, \sigma_i)}{\sum_{i'} \pi_{i'} \mathcal{N}(\mu_{i'}, \sigma_{i'})}$$

for all three clusters for the points $x \in \{-2, 2, 4.5, 6\}$ and assign the instances to clusters.

B. [2p] What happens if we extrapolate from the data, i.e. go to regions where we had no training data? Which Gaussian will be the dominant one? Explain why.