Machine Learning
Exercise Sheet 11

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

January 18th, 2018(noon)
Submission until January 25th, 2018(noon) by dropping at MACHINE LEARNING postbox
(please indicate in which tutorial are you participating!)
The postboxes is located inside the Samelsonplatz building to the right.

Exercise 21: K-Means (10 Points)

Given is the following unlabeled data set:

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

Assume the first cluster center is given through: $\mu_1 = (1 \quad 1)^\top$

- a) 3 pts For K = 2 and K = 3, compute the remaining cluster centers μ_2 and μ_3(for $K = 3$) from the data set.
- b) 4 pts Perform one iteration (i.e. assigning instances to clusters using the given centers, then re-estimating the cluster centers) of K-Means for the given data and K = 2.
- c) 3 pts Perform a second iteration of K-Means for K = 2. Which instance is now differently clustered?

Exercise 22: Gaussian Mixture Models (10 Points)

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

$$\mu_1 = -1 \quad \mu_2 = 1 \quad \mu_3 = 4$$

$$\sigma_1 = 1 \quad \sigma_2 = 2 \quad \sigma_3 = 0.4$$

Additionally, the probabilities for the individual clusters are:

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

- a) 4pts Compute the responsibilities for a point $x \in \mathbb{R}$ to belong to a cluster i as:

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

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

- b) 3 pts 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.

- c) 3 pts K-Means can be understood as a special form of GMM. Discuss, how this is possible, and describe the downsides of K-Means compared to GMMs.