2413, Machine Learning, Tutorial 8
Universität Bern

Simon Jenni (jenni@inf.unibe.ch)

K-means and Mixtures of Gaussians

1. Consider the following data points:
   $x^{(1)} = (1, 1)^T$,
   $x^{(2)} = (1, 3)^T$,
   $x^{(3)} = (7, 1)^T$,
   $x^{(4)} = (7, 3)^T$.

   - Apply the k-means clustering algorithm, when $k = 2$, and the initial centres are $c_1 = (10, 4)^T$ and $c_2 = (0, 2)^T$.

   - Apply the k-means clustering algorithm with a different initialisation. The number of clusters is $k = 2$, and the initial centres are $c_1 = (4, 4)^T$ and $c_2 = (4, 0)^T$.

   - Compare the results of the two runs of the k-means algorithm above.

2. Consider the following data points:
$x^{(1)} = (1, 1)^T$,
$x^{(2)} = (1, 3)^T$,
$x^{(3)} = (2, 1)^T$,
$x^{(4)} = (2, 3)^T$,
$x^{(1)} = (7, 1)^T$,
$x^{(2)} = (7, 3)^T$,
$x^{(3)} = (8, 1)^T$,
$x^{(4)} = (8, 3)^T$.
We start with a hard cluster assignment of the data points, where $p(z^{(i)} = 1 \mid x^{(i)}; \phi, \mu, \Sigma) = 1$ for $i \in \{1, 2, 3, 4\}$ and $p(z^{(i)} = 2 \mid x^{(i)}; \phi, \mu, \Sigma) = 1$ for $i \in \{5, 6, 7, 8\}$. Apply the Expectation Maximisation (EM) algorithm to estimate the parameters of the Mixtures of Gaussians model.

3. There is a connection between K-means and the Mixtures of Gaussians model. You can modify the Maximization step of the later by setting $\Sigma = \epsilon \cdot I$, where I is the identity matrix. Prove that when $\epsilon \to 0$, the Expectation Maximization algorithm reduces to the K-means algorithm.

4. Let us denote the dimension of the training data with m, and let n be the number of data points. What happens to the Expectation Maximisation algorithm, when $m > n$?