Machine Learning Laboratory
Dept. of Computer Science, ETH Zürich
Prof. Dr. Joachim M. Buhmann
Web http://ml2.inf.ethz.ch/courses/cil

Series 3, Mar 17th, 2010 (The K-means Algorithm)

Solution 1 (K-means Theory):

1. (a) (Convergence of the K-Means Algorithm) The K-means algorithm converges since at each iteration it either reduces or keeps the same the value of the objective function J, where

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \|\mathbf{x}_n - \mathbf{u}_k\|_2^2 \quad \left(\|\mathbf{x}_n - \mathbf{u}_k\|_2^2 = (x_{1,n} - u_{1,k})^2 + \cdots + (x_{d,n} - u_{d,k})^2 \right)$$

with the constraint

$$\sum_{k=1}^{K} z_{k,n} = 1 \quad \text{and} \quad z_{k,n} \in \{0, 1\}.$$

When initializing the algorithm, at step 2 of the K-means algorithm we set

$$z_{k^*(\mathbf{x}_n),n} = 1 \quad \text{and} \quad z_{k',n} = 0,$$

where

$$k^*(\mathbf{x}_n) = \operatorname*{argmin}_k \left\{ \|\mathbf{x}_n - \mathbf{u}_1\|_2^2, \ldots, \|\mathbf{x}_n - \mathbf{u}_k\|_2^2, \ldots, \|\mathbf{x}_n - \mathbf{u}_K\|_2^2 \right\}.$$

This makes the value of J minimal considering that we have to assign the value 1 to one and only one $z_{k,n}$, and 0 to all others.
At step 3, the centroid update term you are familiar with:

$$\mathbf{u}_k = \frac{\sum_{n=1}^{N} z_{k,n} \mathbf{x}_n}{\sum_{n=1}^{N} z_{k,n}} \; \forall k, \; k = 1, \ldots, K \tag{1}$$

means that

$$0 = \sum_{n=1}^{N} z_{k,n} (\mathbf{x}_n - \mathbf{u}_k) \; \forall k, \; k = 1, \ldots, K$$

Note that this equals setting the derivative of J with respect to $\mathbf{u}_k$ to zero for all k, $k = 1, \ldots, K$, as a partiuclar derivative is given by:

$$\frac{\partial J}{\partial \mathbf{u}_k} = \frac{\partial \sum_{n=1}^{N} z_{k,n} \|\mathbf{x}_n - \mathbf{u}_k\|_2^2}{\partial \mathbf{u}_k} = \sum_{n=1}^{N} z_{k,n} \begin{bmatrix} \frac{\partial (x_{1,n} - u_{1,k})^2}{\partial u_{1,k}} \\ \vdots \\ \frac{\partial (x_{d,n} - u_{d,k})^2}{\partial u_{d,k}} \end{bmatrix} = -2 \sum_{n=1}^{N} z_{k,n} (\mathbf{x}_n - \mathbf{u}_k)$$

Note that $\frac{\partial^2 J}{\partial \mathbf{u}_k^2} \geq 0$, or in other words, the gradient of J with respect to $\mathbf{u}_k$ is pointing downwards (or is flat). Thus, the value of J does not increase after the centroid update. Considering all the above, it follows that repeating steps 2 and 3 in iterations means that the value of J will converge.

(b) (The K-Means Algorithm and Matrix Factorization) At step 2 of each iteration the K-means algorithm also minimises

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \|\mathbf{x}_n - z_{k,n} \mathbf{u}_k\|_2^2 \tag{2}$$

This follows from the constraints that $\sum_{n=1}^{N} z_{k,n} = 1$ and either $z_{k,n} = 0$ or $z_{k,n} = 1$, for all k, n, since they lead to the following equality:

$$\min_{\mathbf{Z}} \sum_{n=1}^{N} \sum_{k=1}^{K} \|\mathbf{x}_n - z_{k,n} \mathbf{u}_k\|_2^2 = \sum_{n=1}^{N} \left((K-1) \|\mathbf{x}_n\|_2^2 + \min\{ \|\mathbf{x}_n - \mathbf{u}_1\|_2^2, \ldots, \|\mathbf{x}_n - \mathbf{u}_K\|_2^2 \} \right)$$

Similarly, at step 3, for a given $\mathbf{Z}$ we minimize (2) since, for all k, the minimum of (2) with respect to $\mathbf{u}_k$ is given by

$$\frac{\partial \sum_{n=1}^{N} \|\mathbf{x}_n - z_{k,n}\mathbf{u}_k\|_2^2}{\partial \mathbf{u}_k} = 0$$

which leads to (1) since $z_{k,n} = z_{k,n}^2$ for all k, n. It follows that the K-means algorithm minimizes the objective function given by

$$J = \|\mathbf{X} - \mathbf{U}\mathbf{Z}\|_2^2$$

where $\mathbf{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_K]$, $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_N]$, $\mathbf{X} \in \mathbb{R}^{D\times N}$, $\mathbf{U} \in \mathbb{R}^{D\times K}$, and $\mathbf{Z} \in \mathbb{R}^{K\times N}$.