In traditional clustering, there are $K$ clusters $C_1, C_2, \ldots, C_K$ with means $m_1, m_2, \ldots, m_K$. A *least squares error measure* can be defined as

$$D = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - m_k\|^2.$$

which measures how close the data are to their assigned clusters. A least-squares clustering procedure could consider all possible partitions into $K$ clusters and select the one that minimizes $D$. Since this is computationally infeasible, the popular methods are approximations. One important issue is whether or not $K$ is known in advance. Many algorithms expect $K$ as a parameter from the user. Others attempt to find the best $K$ according to some criterion, such as keeping the variance of each cluster less than a specified value.

**Iterative K-Means Clustering** The *K-means* algorithm is a simple, iterative hill-climbing method. It can be expressed as:

**Form K-means clusters from a set of $n$-dimensional vectors.**

1. Set $ic$ (iteration count) to 1.
2. Choose randomly a set of $K$ means $m_1(1), m_2(1), \ldots, m_K(1)$.
3. For each vector $x_i$ compute $D(x_i, m_k(ic))$ for each $k = 1, \ldots, K$ and assign $x_i$ to the cluster $C_j$ with the nearest mean.
4. Increment $ic$ by 1 and update the means to get a new set $m_1(ic), m_2(ic), \ldots, m_K(ic)$.
5. Repeat steps 3 and 4 until $C_k(ic) = C_k(ic + 1)$ for all $k$.

**Algorithm 10.1** K-Means Clustering.

This algorithm is guaranteed to terminate, but it may not find the global optimum in the least squares sense. Step 2 may be modified to partition the set of vectors into $K$ random clusters and then compute their means. Step 5 may be modified to stop after the percentage of vectors that change clusters in a given iteration is small. Figure 10.4 illustrates the application of the K-means clustering algorithm in RGB space to the original football image of Figure 10.1.

**Isodata Clustering** *Isodata clustering* is another iterative algorithm that uses a split-and-merge technique. Again assume that there are $K$ clusters $C_1, C_2, \ldots, C_K$ with means $m_1, m_2, \ldots, m_K$, and let $\Sigma_k$ be the covariance matrix of cluster $k$ (as defined next). If the $x_i$'s are vectors of the form

$$x_i = [v_1, v_2, \ldots, v_n]$$