k-means

The k-means problem is finding the least-squares assignment to centroids.

There are multiple algorithms for finding a solution.

There is an obvious approach to find the global optimum: **enumerating all k^n possible assignments** - that will yield a global minimum, but in exponential runtime.

Much more attention was put to finding an approximate solution in faster time.

The Lloyd/Forgy algorithm is an EM-style iterative model refinement approach, that is guaranteed to converge to a local minimum simply because there is a finite number of states, and the objective function must decrease in every step. This algorithm runs in O(n\*k\*i) where i << n usually, but it may find a local minimum only.

The MacQueens method is technically not iterative. It's a single-pass, one-element-at-a-time algorithm that will not even find a local minimum in the Lloyd sense. (You can however run it multiple passes over the data set, until convergence, to get a local minimum too!) If you do a single pass, its in O(n\*k), for multiple passes add i. It may or may not take more passes than Lloyd.

Then there is Hartigan and Wong. I don't remember the details, IIRC it was a clever, more lazy, variant of Lloyd/Forgy, so probably in O(n\*k\*i), too (although probably not recomputing all n\*k distances for later iterations?)

You could also do a randomized alogrithm that just tests l random assignments. It probably won't find a minimum at all, but run in "linear" time O(n\*l).

Oh, and you can try different random initializations, to improve your chances of finding the global minimum. Add a factor t for the number of trials...