**Part 6 written answers: studying convergence properties by computing Euclidean distance of all words in a cluster to its cluster mean.**

**For small texts (easy\_sanity\_test.txt) using k = 3, 100 iterations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | For small texts, the algorithm tends to converge after 1-2 iterations. This test had initial points [cat, squirrel, pumpkin]. These are the final clusters.   |  | | --- | | Cluster 1 - wolf | | Cluster 1 - fox | | Cluster 1 - squirrel | | Cluster 1 - cat | | Cluster 1 - dog | | Cluster 1 - tiger | | Cluster 2 - ten | | Cluster 2 - nine | | Cluster 2 - eight | | Cluster 2 - seven | | Cluster 2 - six | | Cluster 2 - five | | Cluster 2 - four | | Cluster 2 - three | | Cluster 2 - two | | Cluster 2 - on | | Cluster 3 - tomato | | Cluster 3 - basil | | Cluster 3 - parslei | | Cluster 3 - appl | | Cluster 3 - banana | | Cluster 3 - squash | | Cluster 3 -pumpkin | |

**For large texts (using sleepy hollow), using input K = 30 clusters and k = 100 clusters**

|  |
| --- |
|  |
| For large texts, the k-clustering converges at about 10 iterations, for k values tested between 30-200. After 10 iterations, the means and the average of Euclidean distance of points to their cluster means no longer change significantly. |