

Module 8: Scikit and Introduction to Hadoop

Assignment

edureka!

edureka!

1. How to get the datasets that come with scikit-learn package?
2. Give example of two datasets that is shipped with scikit-learn package. Show example how to use one.
3. List at least 5 modules from scikit-learn?
4. Take input as few random points in a two dimensional space and divide them into 4 clusters.
5. Use K-means algorithm by using Scikit-learn to find the centroids of each cluster.
6. Plot the points along with the centroids. To distinguish the centroids, use some special symbols like '+' or '#' etc.
7. How many daemon processes run on a Hadoop cluster? Explain.
8. If Hadoop spawned 200 tasks for a job and one of the task failed. What will Hadoop do in this case?
9. Where does mapper output store the intermediate results?
10. When does reducers start their execution in a map reduce process?