

10/27/2019

CS226 : Programming Assignment 2: Report

Submitted By: Abhishek Ayachit (862188073)

1. Which InputFormat did you use in the MapReduce program?

Input Format Used: TextInputFormat (job.setInputFormatClass(TextInputFormat.class))

2. What is the input and output format of the map function?

The input is a text file with type Text and the output is key value pairs which are of type DoubleWritable and Text respectively.

3. What is the logic of the map function?

The map function takes a text file as an input. The file is read and the euclidean distance between each point and the query point is calculated and written as the output of the map function.

4. If a combiner function is used, what is the signature of the combiner function (input and output) and what is its logic?

Combiner class summarizes the output of the map function and provide it to the reducer. It reduces the number of records sent to the reducer.

The signature: public static class KNNReducer extends Reducer<DoubleWritable, Text, DoubleWritable, Text>

It would be same as reduce class.

5. If a reduce function is used, what is the signature of the reduce function (input and output) and what is its logic?

The reducer class takes an input as key value pairs which are of type DoubleWritable and Text respectively and writes top k values to the context which is also key value pairs with DoubleWritable and Text respectively.

Signature is: public static class KNNReducer extends Reducer<DoubleWritable, Text, DoubleWritable, Text>

6. How many mappers and reducers are needed for your program?

Number of mappers: 4 (4 input splits)

Number of reducers: 1 (job.setNumReduceTasks(1);)

7. How many records are shuffled between the mappers and reducers?

Number of records shuffled after map task, i.e., input of the combiner is: 10507403

Number of records shuffled after combine task, i.e., input of the reducer is: 86