Walkthrough of the Project:

1. At first put the input.txt in Hadoop filesystem

$ bin/Hadoop fs -put ./input.txt /user/sen/logistic\_regression/input/

1. Sample training text-files as input:

$ bin/hadoop fs -ls /user/sen/logistic\_regression/input/

/user/sen/logistic\_regression/input/ input.txt

$ bin/hadoop fs -cat /user/sen/logistic\_regression/input/ input.txt

9,171,110,24,240,45.4,0.721,54,1

1. Run the training application:

$ bin/hadoop jar dr.jar Driver /user/sen/logistic\_regression/input

/user/sen/logistic\_regression/output

1. Output from training model

$ bin/hadoop fs -cat /user/sen/logistic\_regression /output/part-r-00000

theta0 -0.29324508

theta1 1.3288932

theta2 1.9172496

theta3 -3.579825

theta4 -2.189732

theta5 1.6235878

theta6 -0.87215364

theta7 0.069883265

theta8 0.017198116

1. Sample testing text-files as test input:

$ bin/hadoop fs -cat /user/sen/logistic\_regression/input/test/test\_input.txt

1. Run the test application:

hadoop jar test.jar Test 8 /user/sen/logistic\_regression /output/part-r-00000 /user/sen/logistic\_regression/input/test/file1

result18

1. Output from test model

real\_value\_for\_test\_1 : 1.0 1.0

real\_value\_for\_test\_2 : 1.0 1.0

real\_value\_for\_test\_3 : 0.0 0.0

real\_value\_for\_test\_4 : 1.0 0.0