Set Spark home environment variable.

Scala – 2.11
Spark – 2.3.1

Running Task 1:

$ $SPARK_HOME/bin/spark-submit --class SaiTeja_Chava_ModelBasedCF

/Users/sai/IdeaProjects/untitled/out/artifacts/untitled_jar/SaiTeja_Chava_hw2.jar

"/Users/sai/Documents/Data Mining INF 553/Assignments/hw2/Data/train_review.csv"

"/Users/sai/Documents/Data Mining INF 553/Assignments/hw2/Data/test_review.csv"

I mentioned the paths to test and train in the command above. Replace them with appropriate paths accordingly.

Task 1

>= 0 and < 1: 28485
>= 1 and < 2: 14106
>= 2 and < 3: 2370
>= 3 and < 4: 237
>= 4: 38

RMSE = 1.0855222036834633

Running Task 2

$ $SPARK_HOME/bin/spark-submit --class SaiTeja_Chava_UserBasedCF

/Users/sai/IdeaProjects/untitled/out/artifacts/untitled_jar/SaiTeja_Chava_hw2.jar

"/Users/sai/Documents/Data Mining INF 553/Assignments/hw2/Data/train_review.csv"

"/Users/sai/Documents/Data Mining INF 553/Assignments/hw2/Data/test_review.csv"

I mentioned the paths to test and train in the command above. Replace them with appropriate paths accordingly.

Task 2

>= 0 and < 1: 29951
>= 1 and < 2: 11041
>= 2 and < 3: 2996

>= 3 and < 4: 890
>= 4: 358
RMSE = 1.1074213674692792
The total execution time taken is 365 sec.

Description:

In Intellij I edited the configurations and gave in the VM options section the following:

-Xms23000M -Xmx25000M

Basically, increase the executor memory and driver memory to 12 – 13GB. Increase accordingly. Else lot of spilling happens and you won't be able to run the code.