Recommendation Systems – HW3 – INF 553 – Nilay Chheda (9053988992)

NOTE: You should have Spark 1.6.1 built for Hadoop 1 and library dependencies for Scala 2.11.x.
(link) https://archive.apache.org/dist/spark/spark-1.6.1/spark-1.6.1-bin-hadoop1-scala2.11.tgz

Task1: Used average rating of a user to fill out the missing predictions from ALS due to missing entries for movie IDs in entire train dataset. Apart from that used vanilla ALS with tweaked parameter to beat the baseline.

Command

```
$ ./spark-submit --class hw3_task1 --master local[*] --driver-memory 4g
Nilay_Chheda_task1.jar <csv entire data file path> <csv test data file path>
```

Accuracy Information

Task1

	Small	Large
≥0 and <1	15000	3239905
≥1 and <2	4315	718097
≥2 and <3	815	87091
≥3 and <4	120	9077
≥4	6	281
RMSE	0.9549435950591381	0.821956990633416

Task2: Used standard item based CF with co-rated items and scaling based on number of co-rated users, threshold of 0.5 and above for Pearson coefficient with 5 nearest neighbors. Used average rating of a user to fill out the missing predictions from item based CF.

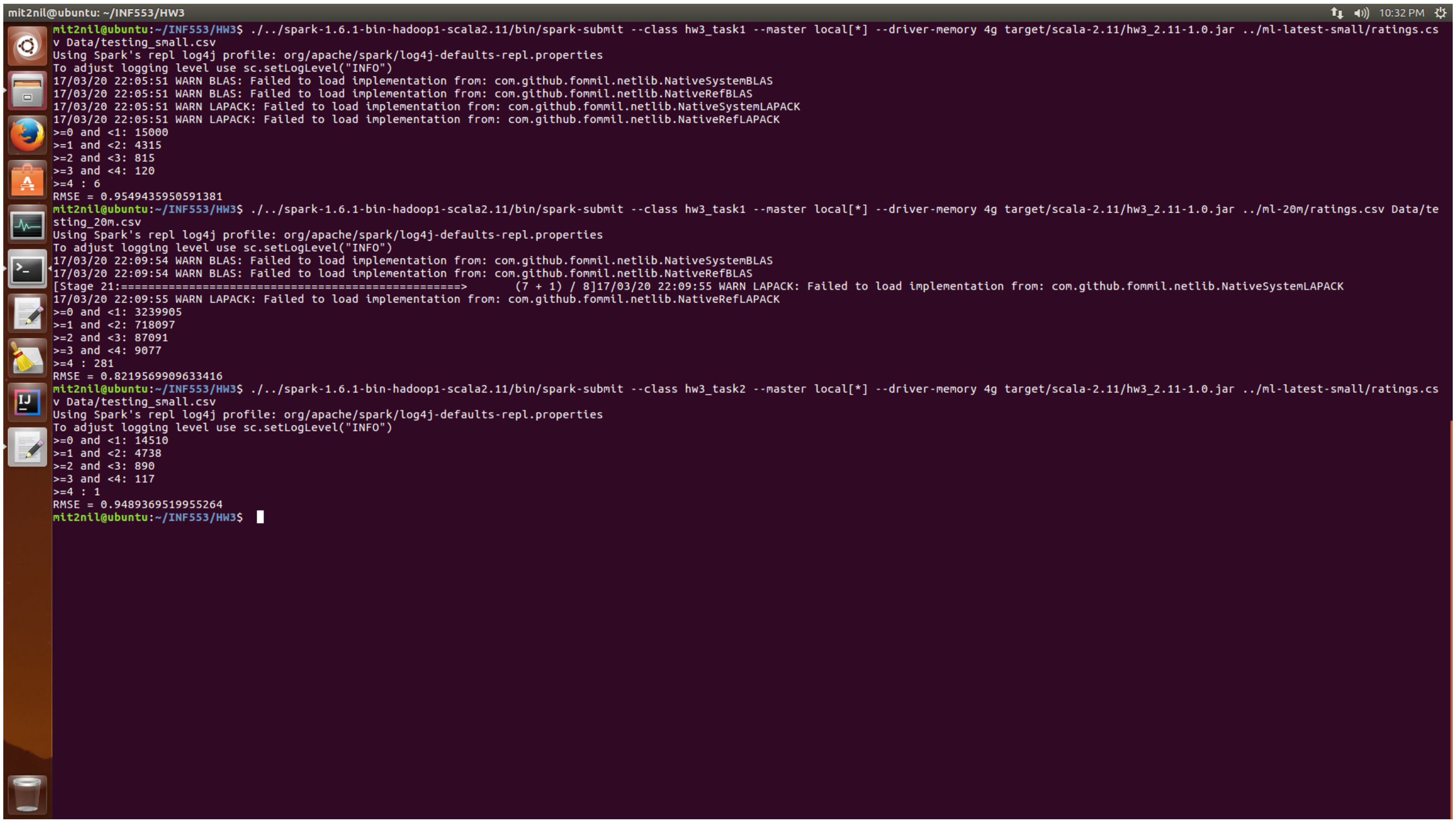
Command

```
$ ./spark-submit --class hw3_task2 --master local[*] --driver-memory 4g
Nilay_Chheda_task2.jar <csv entire data file path> <csv test data file path>
```

Accuracy Information

Task2

	Small
≥0 and <1	14510
≥1 and <2	4738
≥2 and <3	890
≥3 and <4	117
≥4	1
RMSE	0.9489369519955264



```
mit2nil@ubuntu:~/INF553/HW3$ ./../spark-1.6.1-bin-hadoop1-scala2.11/bin/spark-submit --class hw3_task1 --master local[*] --driver-memory 4g target/scala-2.11/hw3_2.11-1.0.jar ../ml-latest-small/ratings.cs
v Data/testing_small.csv
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties
To adjust logging level use sc.setLogLevel("INFO")
17/03/20 22:05:51 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
17/03/20 22:05:51 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
17/03/20 22:05:51 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeSystemLAPACK
17/03/20 22:05:51 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeRefLAPACK
>=0 and <1: 15000
>=1 and <2: 4315
>=2 and <3: 815
>=3 and <4: 120
>=4 : 6
RMSE = 0.9549435950591381
mit2nil@ubuntu:~/INF553/HW3$ ./../spark-1.6.1-bin-hadoop1-scala2.11/bin/spark-submit --class hw3_task1 --master local[*] --driver-memory 4g target/scala-2.11/hw3_2.11-1.0.jar ../ml-20m/ratings.csv Data/te
sting_20m.csv
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties
To adjust logging level use sc.setLogLevel("INFO")
17/03/20 22:09:54 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
17/03/20 22:09:54 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
[Stage 21:==================================================>       (7 + 1) / 8]17/03/20 22:09:55 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeSystemLAPACK
17/03/20 22:09:55 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeRefLAPACK
>=0 and <1: 3239905
>=1 and <2: 718097
>=2 and <3: 87091
>=3 and <4: 9077
>=4 : 281
RMSE = 0.8219569909633416
mit2nil@ubuntu:~/INF553/HW3$ ./../spark-1.6.1-bin-hadoop1-scala2.11/bin/spark-submit --class hw3_task2 --master local[*] --driver-memory 4g target/scala-2.11/hw3_2.11-1.0.jar ../ml-latest-small/ratings.cs
v Data/testing_small.csv
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties
To adjust logging level use sc.setLogLevel("INFO")
>=0 and <1: 14510
>=1 and <2: 4738
>=2 and <3: 890
>=3 and <4: 117
>=4 : 1
RMSE = 0.9489369519955264
mit2nil@ubuntu:~/INF553/HW3$ 
```