Homework 3

The version of Spark is version 2.1.0

The version of Scala is version 2.11.8

The version of Python is version 2.7.10

- **Task 1**

  How to run the program: the first parameter is the class of the scala, and the second input is the file path of source code of testing data, and the third input is the file path of data with ratings. To run the .jar file you need to input file like following:

  Small data:

  ```
  ➜  spark-2.1.0-bin-hadoop2.7 bin/spark-submit --class modelBasedCF --master "lo
  cal[*]" /Users/Brian/Desktop/ChiWeiLiu_CF_ItemBased.jar /Users/Brian/Desktop/in
  f553/HW3/testing_small.csv /Users/Brian/Desktop/inf553/HW3/ratings_small.csv
  ```

  Big data:

  ```
  ➜  spark-2.1.0-bin-hadoop2.7 bin/spark-submit --class modelBasedCF --master "lo
  cal[*]" /Users/Brian/Desktop/ChiWeiLiu_CF_ItemBased.jar /Users/Brian/Desktop/in
  f553/HW3/testing_20m.csv /Users/Brian/Desktop/inf553/HW3/ratings_20m.csv
  ```

  The file will be generated under the Spark folder, and the file name will be ChiWeiLiu Scala.

- **Task 2**

  How to run the program: the first parameter is the file path of source code. The second input is the file path of data for prediction, and the third input is the file path of data with ratings

  ```
  ➜  spark-2.1.0-bin-hadoop2.7 bin/spark-submit --master "local[*]" /Users/Brian/
  Desktop/inf553/HW3/testCF.py /Users/Brian/Desktop/inf553/HW3/ratings_small.csv
  /Users/Brian/Desktop/inf553/HW3/testing_small.csv
  ```

  The file will be generated under the Spark folder, and the file name will be ChiWei_Liu_result_task2.

- **Accuracy sheet**

	Task 1		Task 2
	Small	Large	Small
>=0 and <1	14872	3232480	15821
>=1 and <2	4499	731707	3798
>=2 and <3	769	82514	568
>=3 and <4	112	7542	67
>=4	4	208	2
RMSE	0.9519	0.8193	0.8632