Pseudocode

Calculate Lift

Lift-spark.py

--------------------------

Read the input file u.data by taking it in SparkContext

Split the line on ‘\t’ for userID, movieID and rating

If rating>=4

Make pairs of userID and movieID and send it to map function

Reduce the output by calling reduce function

Make the movie tuple(movie stripes) and then reduce it by using reduceByKey function which will give the corresponding count

Emit the movie names with the counts and the lift

Function map1:

Takes the userID and movieID pairs and assign it to a list(array)

Return this array

Function reduce1:

For each user ID

Make the movie pairs groupBy userID

Sum1=Count the number of times the movie has occurred

Generate the movie stripes along with the count

Generate the co-occurrence dictionary (associative array) and sum up the counts

Sum2=Divide the counts obtained by the total sum of all the counts of the movies listed in the dictionary for the movie stripes

Calculate condprob=sum2/sum1 and then calculate lift=condprob/sum1

For every key in dictionary, split the pairs and print the movie names along with their counts and append the lift calculated to this dictionary

Return this dictionary

Function loadMovieNames:

Load the file ‘u.item’ and split each line on ‘|’

Map the movieID’s with the corresponding names

Return movie name dictionary