Misc Announcements/Quiz

Ready for some real ML work?

We are looking for programmers for the edigs project (edigs.org). This is a soon-to-be-deployed (mid-april deployment) application for prospective tenants to find better housing. We have an android app and web app and are looking for someone who could help us create an iphone app as well. In addition, we are doing applied machine learning work to predict the expected cost of the utilities, and an interested student could be involved in helping us to implement and extend that aspect of the system. Please email **jmankoff@cs.cmu.edu** if interested. - Jenn Mankoff

Ready for some real ML work?

I will have some slots for my group also for summer work (credit or $$). Send me a cv.

Some biology background is a big plus.

- William

Homework 7 - Hints

- How do you define the strata? first assign rows and columns to **blocks** then assign blocks to strata

1				
	1			
		1		
			1	
				1

	2			
		2		
			2	
				2
2				

1	2	3	4	5
5	1	2	3	4
4	5	1	2	3
3	4	5	1	2
2	3	4	5	1

Homework 7 - Hints

- How do you define the blocks?
- Any (non-random) one-one mapping from a row/column index to [1,…K] is fine

1				
	1			
		1		
			1	
				1

	2			
		2		
			2	
				2
2				

1	2	3	4	5
5	1	2	3	4
4	5	1	2	3
3	4	5	1	2
2	3	4	5	1

Homework 7 - Hints

- How do you read the data and put it into triple format? it's a nice warmup...
  - rdd.wholeTextFiles("/path/to/dir") returns key,value pairs where key=file name, value=file contents as a string
  - flatMap the result with a python function that takes a string and converts it to a *list* of triples

    movieId:

    userId1,rating1,date1

    userId2,rating2,date2

    ....

Homework 7 - Hints

- What's parallel and what's sequential?

- for epoch t=1,….,T in sequence
  - for stratum s=1,…,K in sequence
    - for block b=1,… in stratum s in parallel
      - for triple (i,j,rating) in block b in sequence
        » do SGD step for (i,j,rating)

  only needs part of W,H

- Hint 1: You need some loops around rdd operations
- Hint 2: Look at rdd.mapPartitions() to see how to control a parallel/sequential breakdown like the inner loop

April QUIZ

Which of the following programs (1) graduated William Cohen in 1984 (2) graduated Alex Beutel in 2011 (3) has been in 12 NCAA Final Fours, 9 NCAA Championship games, and won 5 NCAA Championship titles, since 1980, when William Cohen entered as a freshman?

A. Duke University
B. UNC/Chapel Hill
C. Stanford and MIT put together
D. Duke! go Duke! **DUKE!**